medRxiv preprint doi: https://doi.org/10.1101/2024.10.28.24314051; this version posted November 1, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Extremely rare CNVs contributing to Alzheimer disease risk: a case-control association analysis of exome sequencing data from 22,319 individuals

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], Anne-Claire Richard[1], Stéphane Rousseau[1], Shahzad Ahmad[9,10], Najaf Amin[9,11], Philippe Amouyel[3], Olivia Belbin[12,13], Céline Bellenguez[3], Claudine Berr[14], Anne Boland[15], Paola Bossù[16], Femke Bouwman[6,8], Jose Bras[17,18], Jordi Clarimon[12,13], Antonio Daniele[19], Jean-François Dartigues[20], Stéphanie Debette[20,21], Jean-François Deleuze[15], Nicola Denning[22], Oriol Dols-Icardo[12,13], Cornelia M. van Duijn[9,11], Victoria Fernandez[23], Wiesje M. van der Flier[5,6], Juan Fortea[12], Nick C. Fox[24], Ruth Frikke-Schmidt[25,26], Daniela Galimberti[27,28], Emmanuelle Genin[29], Roberta Ghidoni[30], Vilmantas Giedraitis[31], Johan J. P. Gille[32], Detelina Grozeva[33], Rita Guerreiro[17,18], Edna Grünblatt[34], John Hardy[35], Steffi G. Riedel-Heller[36], Mikko Hiltunen[37], Clive Holmes[38], Henne Holstege[4,5,6,8], Jakub Hort[39], Holger Hummerich[40], M. Arfan Ikram[9], M. Kamran Ikram[9], Martin Ingelsson[31], Iris E. Jansen[5,6,41], Amit Kawalia[42], Robert Kraaij[43], Patrick G. Kehoe[44], Marc Lathrop[45], Morgane Lacour[46], Afina W. Lemstra[5,6], Alberto Lleó[12,13], Lauren Luckcuck[33], Marcel M. A. M. Mannens[47], Rachel Marshall[33], Carlo Masullo[48], Simon Mead[40], Patrizia Mecocci[49,50], Alexandre de Mendonça[51], Alun Meggy[22], Shima Mehrabian[52], Merel O. Mol[53], Kevin Morgan[54], Alexandre Morin[46], Benedetta Nacmias[55,56], Penny J. Norsworthy[40], Florence Pasquier[57], Pau Pastor[58,59], Fabrizio Piras[60], Julius Popp[61], Alfredo Ramirez[42,62,63,64,65], Rachel Raybould[22], Richard Redon[66], Marcel J. T. Reinders[8], Fernando Rivadeneira[43], Jeroen G. J. van Rooij[43,53], Natalie S. Ryan[24], Salha Saad[33], Pascual Sanchez-Juan[13,67], Nikolaos Scarmeas[68,69], Philip Scheltens[5,6], Jonathan M. Schott[24], Davide Seripa[70], Daoud Sie[32], Rebecca Sims[33], Erik A. Sistermans[32], Sandro Sorbi[55,56], Kristel Sleegers[71,72,73], Resie van Spaendonk[32], John C. van Swieten[53], Niccolo' Tesi[4,5,6,8], Betty M Tijms[5], André G. Uitterlinden[43], Magda Tsolaki[74,75], Pieter Jelle Visser[5], Michael Wagner[63,64], Julie Williams[33], Aline Zarea[46], EADB Consortium, David Wallon[46], Sébastien Jacquemont[7,76], Jean-Charles Lambert[3], Sven J. van der Lee[4,5,6], Camille Charbonnier[77], 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.Department of Epidemiology, Erasmus Medical Centre, Rotterdam, the Netherlands. 10.Leiden Academic Centre for Drug Research, Leiden, the Netherlands. 11.Nuffield Department of Population Health Oxford University, Oxford, UK. 12.Department of Neurology, II B Sant Pau, Hospital de la Santa Creu i Sant Pau, Universitat Autònoma de Barcelona, Barcelona, Spain. 13.Biomedical Research Networking Center on Neurodegenerative Diseases, National Institute of Health Carlos III, Madrid, Spain. 14.Université Montpellier, INSERM, Institute for Neurosciences of Montpellier, Montpellier, France. 15.Université Paris-Saclay, CEA, Centre National de Recherche en Génomique Humaine(CNRGH), 91057, Evry, France 16.Experimental Neuro-psychobiology Laboratory, Department of Clinical and Behavioral Neurology, Istituto di Ricovero e Cura a Carattere Scientifico Santa Lucia Foundation,

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2024.10.28.24314051; this version posted November 1, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Rome, Italy. 17.Department of Neurodegenerative Science, Van Andel Institute, GrandRapids, MI, USA. 18.Division of Psychiatry and Behavioral Medicine, Michigan State University College of Human Medicine, Grand Rapids, MI, USA. 19.Department of Neuroscience, Catholic University of Sacred Heart, Fondazione Policlinico Universitario A. Gemelli Istituto di Ricovero e Cura a Carattere Scientifico, Rome, Italy. 20.Université Bordeaux, INSERM, Bordeaux Population Health Research Center, Bordeaux, France. 21.Department of Neurology, Bordeaux University Hospital, Bordeaux, France. 22.UKDRI Cardiff, School of Medicine, Cardiff University, Cardiff, UK. 23.Memory Clinic of Fundaciò ACE, Catalan Institute of Applied Neurosciences, 08028 Barcelona, Spain. 24.Dementia Research Centre, University College London Queen Square Institute of Neurology, London, UK. 25.Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark. 26.Department of Clinical Biochemistry, Rigshospitalet, Copenhagen, Denmark. 27.Fondazione Istituto di Ricovero e Cura a Carattere Scientifico C' Granda, Ospedale Policlinico, Milan, Italy. 28.University of Milan, Milan, Italy. 29.Université Brest, INSERM, Etablissement Français du Sang, Centre Hospitalier Universitaire Brest, Unité Mixte de Recherche 1078, GGB, Brest, France. 30.Molecular Markers Laboratory, IRCCS Istituto Centro San Giovanni di Dio Fatebenefratelli, Brescia, Italy. 31.Department of Public Health and Carins Sciences/Geriatrics, Uppsala University, Uppsala, Sweden. 32.Genome Diagnostics, Department of Human Genetics, VU University, AmsterdamUMC (locationVUmc), Amsterdam, the Netherlands. 33.Medical Research Council Centre for Neuropsychiatric Genetics and Genomics, Division of Psychological Medicine and Clinical Neuroscience, School of Medicine, Cardiff University, Cardiff, UK. 34.Department of Child and Adolescent Psychiatry and Psychotherapy, Psychiatric University Hospital Zurich, University of Zurich, Zurich, Switzerland; Neuroscience Center Zurich, University of Zurich and ETH Zurich, Zurich, Switzerland; Zurich Center for Integrative Human Physiology, University of Zurich, Zurich, Switzerland. 35.Reta Lila Weston Research Laboratories, Department of Molecular Neuroscience, University College London Institute of Neurology, London, UK. 36.Institute of Social Medicine, Occupational Health and Public Health, University of Leipzig, Leipzig, Germany. 37.Institute of Biomedicine, University of Eastern Finland, Kuopio, Finland. 38.Clinical and Experimental Science, Faculty of Medicine, University of Southampton, Southampton, UK. 39.Memory Clinic, Department of Neurology, Second Faculty of Medicine, Charles University, Motol University Hospital, Prague, Czech Republic. 40.Medical Research Council Prion Unit at University College London, University College London Institute of Prion Diseases,London, UK. 41.Department of Complex Trait Genetics, Center for Neurogenomics and Cognitive Research, Amsterdam Neuroscience, Vrije University, Amsterdam, the Netherlands. 42.Division ofNeurogenetics and Molecular Psychiatry, Department of Psychiatry and Psychotherapy, Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany. 43.Department of Internal Medicine, Erasmus Medical Centre, Rotterdam, the Netherlands. 44.Translational Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK. 45.McGill University and Genome Quebec Innovation Centre, Montreal, Quebec, Canada. 46.Univ Rouen Normandie, Normandie Univ, Inserm U1245 and CHU Rouen, Department of Neurology and CNRMAJ, F-76000 Rouen, France. 47.Department of Human Genetics, Amsterdam UMC, University of Amsterdam, Amsterdam Reproduction and Development Research Institute, Amsterdam, the Netherlands. 48.Institute of Neurology, Catholic University of the Sacred Heart, Rome, Italy. 49.Division of Gerontology and Geriatrics, Department of Medicine and Surgery, University of Perugia, Perugia, Italy 50.Division of Clinical Geriatrics, Department of Neurobiology, Care Sciences and Society, Karolinska Institutet, Stockholm, Sweden 51.Faculty of Medicine, University of Lisbon, Lisbon, Portugal. 52.Department of Neurology, University Hospital 'Alexandrovska', 1431 Sofia, Bulgaria. 53.Department of Neurology, Erasmus Medical Centre, Rotterdam, the Netherlands. 54.Human Genetics, School of Life Sciences, University of Nottingham, Nottingham, UK. 55.Department of Neuroscience, Psychology, Drug Research and Child Health University of Florence,

medRxiv preprint doi: https://doi.org/10.1101/2024.10.28.24314051; this version posted November 1, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Florence, Italy. 56.IRCCS Fondazione Don Carlo Gnocchi, Florence, Italy. 57.Université Lille, INSERM, Centre Hospitalier Universitaire Lille, UMR1172, Resources and Research Memory Center (MRRC) of Distalz, Licend, Lille, France. 58.Fundació Docència i Recerca MútuaTerrassa and Movement Disorders Unit, Department of Neurology, University Hospital MútuaTerrassa, Barcelona, Spain. 59.Memory Disorders Unit, Department of Neurology, Hospital Universitari Mutua de Terrassa, Barcelona, Spain. 60.Laboratory of Neuropsychiatry, Department of Clinical and Behavioral Neurology, Istituto di Ricovero e Cura a Carattere Scientifico Santa Lucia Foundation, Rome, Italy. 61.University Hospital of Psychiatry Zürich, University of Zürich, Zürich, Switzerland. 62.Department of Psychiatry and Glenn Biggs Institute for Alzheimer's and Neurodegenerative Diseases, San Antonio, TX, USA. 63.Department of Neurodegenerative Diseases and Geriatric Psychiatry, University Hospital Bonn, Medical Faculty, Bonn, Germany. 64.German Center for Neurodegenerative Diseases, Bonn, Germany. 65.Cluster of Excellence Cellular Stress Responses in Aging-Associated Diseases, University of Cologne, Cologne, Germany. 66.Université de Nantes, Centre Hospitalier Universitaire Nantes, Centre National de la Recherche Scientifique, INSERM, l'institut du Thorax, Nantes, France. 67.Neurology Service, Marqués de Valdecilla University Hospital (University of Cantabria and IDIVAL), Santander, Spain. 68.Department of Neurology, Aiginition Hospital, National and Kapodistrian University of Athens, 11528 Athens, Greece. 69.Department of Neurology, Columbia University, New York, NY 10032, USA. 70.Laboratory for Advanced Hematological Diagnostics, Department of Hematology and Stem Cell Transplant, Lecce, Italy. 71.Complex Genetics of Alzheimer's Disease Group, VIB Center for Molecular Neurology, VIB, Antwerp, Belgium. 72.Laboratory of Neurogenetics, Institute Born - Bunge, Antwerp, Belgium. 73.Department of Biomedical Sciences, University of Antwerp, Antwerp, Belgium. 74.First Department of Neurology, Medical School, Aristotle University of Thessaloniki, Thessaloniki, Greece. 75.Alzheimer Hellas, Thessaloniki, Greece. 76.Department of Pediatrics, Université de Montréal, Montreal, Quebec, Canada. 77.Univ Rouen Normandie, Normandie Univ, Inserm U1245 and CHU Rouen, Department of Biostatistics and CNRMAJ, F-76000 Rouen, France.

medRxiv preprint doi: https://doi.org/10.1101/2024.10.28.24314051; this version posted November 1, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Figures (legends below)

Figure 1: Summary of the study

A. Study workflow and data quality control
This study included whole exome sequenced samples from 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.

B. Illustration of the dosage analysis

C. Forest plot associated with the meta-analysis of results from both main analysis and replication
Figure displays OR with 95%Confidence Intervals.
QC: Quality Control; CNV: Copy Number Variation; EOAD: Early Onset Alzheimer Disease; LOAD: Late Onset Alzheimer Disease; OR: Odd-ratio
Orange: discovery, green: replication, black: meta-analysis

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 reparation (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. Percentage of deletions and duplications carriers by status and gene
In red: deletions; In blue: complete duplications. Data include both main analysis and replication datasets.

C. 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

Supplementary Tables

Table S1. Pathogenic variants in stage 2 from the Holstege et al. study

Table S2. Samples per contributing study

Table S3. List of large CNVs

Table S4. Results of all AD cases versus controls (dosage, deletions, duplications)

Table S5. Carriers of CNVs encompassing the chr 22q11.21 locus

Table S6. Deletion-only analysis (FDR ≤ 20%)

Table S7. Duplication-only analysis (FDR ≤ 20%)

Table S8. List of all CNVs in prioritized genes and validation status

Table S9. APOE genotypes in locus ERCC2/KLC3 duplication carriers and non-carriers

medRxiv preprint doi: https://doi.org/10.1101/2024.10.28.24314051; this version posted November 1, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Table S10. APOE4-adjusted dosage analysis

Table S11. Dosage analysis adjusted for population structure

Table S12. List of loss-of-function variants

Table S13. Loss-of-function analysis (CNV and SNV/indels)

Table S14. Cumulate CNV analysis

Table S15. Replication analysis

Supplementary Figures (legends are inserted as notes in the .ppt document)

Fig S1. Example of a large CNV on chromosome 9 suggesting a blood-specific event (probable clonal hematopoiesis)

Fig S2. Theoretical example of definition of CNV coordinates according to transcript and capture kit coordinates.

Fig S3. *ERCC2-KLC3* locus on chr19 with CNVs displayed on the UCSC genome browser

Fig S4. *MBL2* locus with CNVs displayed on the UCSC genome browser

Fig S5. *ADI1* locus with CNVs displayed on the UCSC genome browser

Fig S6. *FADS6* locus with CNVs displayed on the UCSC genome browser

Fig S7. *ABCA1* locus with CNVs displayed on the UCSC genome browser

Fig S8. *ABCA7* locus with CNVs displayed on the UCSC genome browser

Fig S9. *TYROBP* locus with CNVs displayed on the UCSC genome browser

Fig S10. CNVs affecting known AD-associated genes

Fig S11. *CTSB* locus with CNVs displayed on the UCSC genome browser

Fig S12. Time required for the calling step of our pipeline with two versions of CANOES

Fig S13. QQ-plot for dosage analysis before and after adjustment for ancestry in EOAD versus Controls analysis.

Fig S14. Projection of *MBL2* CNV carriers on the first two ancestry components.

Figure S15. 22q11.21 locus with CNVs from replication displayed on the UCSC genome browser

Figure S16. *ADI1* locus with CNVs from replication displayed on the UCSC genome browser

Figure S17. *FADS6* locus with CNVs from replication displayed on the UCSC genome browser

medRxiv preprint doi: https://doi.org/10.1101/2024.10.28.24314051; this version posted November 1, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Abstract

Rare coding single nucleotide variants (SNV) and short insertions or deletions (indels) contribute to Alzheimer disease (AD) genetic risk, from pathogenic variants in autosomal dominant genes to risk factors with diverse effects. In contrast, copy number variants (CNV) have been scarcely studied, with the exception of a few autosomal dominant examples, such as *APP* gene duplications.

We took advantage from a large case-control dataset of 22,319 exomes (4,150 early-onset AD (EOAD, onset ≤65 years), 8,519 late onset AD (LOAD) and 9,650 controls) to detect CNVs. We first identified 17 causative CNVs in autosomal dominant genes (9 novel: 7 *APP* and 2 *MAPT* duplications). After exclusion of carriers of these, we performed an original two-step analysis: (i) a protein-coding genome-wide analysis at the transcript level using a gene dosage strategy (EOAD versus controls) and then (ii) an integrated loss-of-function (LOF) analysis gathering short truncating variants with CNV-deletions in genes prioritized in (i) and in a list of known AD risk genes.

We identified AD association with dosage of 20 genes at 4 different *loci* with a false discovery rate (FDR) below 10%, including the chr22q11.21 central region (FDR=0.0386), a region in linkage disequilibrium with the *APOE* locus on chr19 (FDR=0.0271), and two single-gene loci, namely *FADS6* (FDR=0.0271), and *ADI1* (FDR=0.0916). Replication in an independent dataset made of genotyping array data from 2,780 EOAD cases, 15,222 LOAD cases and 273,979 controls was consistent with the results obtained in the discovery dataset. The integrated LOF analysis helped narrowing the region of interest to the *SCARF2-KLHL22-MED15* region at the 22q11.21 locus. In addition, the integrated LOF analysis highlighted rare deletions in the known AD-risk genes *ABCA1* and *ABCA7* that represented 10% (3/30) and 8.6% (10/115) of LOF alleles of these genes, respectively, as well as 4 *TYROBP* deletions. Finally, we identify *CTSB* LOF alleles as candidate rare AD risk factors (p=0.0089).

In conclusion, our results show that carriers of a deletion of *FADS6* or a 22q11.21 deletion encompassing the *SCARF2-KLHL22-MED15* region, including some patients with DiGeorge syndrome, may have a higher risk of developing AD. CNVs represent a source of genomic variation that can contribute to AD etiology in new genes but also in GWAS defined-genes.

medRxiv preprint doi: https://doi.org/10.1101/2024.10.28.24314051; this version posted November 1, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Introduction

The etiology of Alzheimer disease (AD) is heterogeneous. Some families exhibit autosomal dominant early-onset AD (EOAD, onset ≤65 years), explained by rare variants in the *PSEN1*, *PSEN2* or *APP* genes, but carriers of such highly penetrant variants represent less than 0.5% of all AD cases (Nicolas 2024). In non-Mendelian cases, AD is considered as a complex disorder with a high genetic component (Gatz et al. 2006).

In complex AD, a large diversity of risk factors have been reported. They are usually classified according to their respective frequencies and effect sizes. Among common variants (frequency >1%), the ε4 allele of the *APOE* gene is the main risk factor with odd ratios (OR) of 3-4 and 11-14 for heterozygous and homozygous carriers, respectively (Genin et al. 2011). A recent large genome-wide association study (GWAS) reported 75 loci, most of them being associated through common non-coding single nucleotide polymorphisms (SNPs) with a modest effect on AD risk (Bellenguez et al. 2022). On the other hand, a burden of rare (frequency <1%) to ultra-rare (up to singleton) variants in *SORL1*, *TREM2, ABCA7, ABCA1* and *ATP8B4* as well as a handful of rare recurrent single variants demonstrated a wider diversity of effects, ranging from modest odds ratios for the least rare variants (OR=[1.5; 3], e.g., in *ABI3*, *TREM2* (R62H), *NCK2*, *SORT1*) to strong effects for some ultra-rare variants (e.g., loss-of-function *SORL1* variants) (Bellenguez et al. 2022; Holstege et al. 2022; Jonsson et al. 2013; Nicolas et al. 2016; Schramm et al. 2022; Sims et al. 2017; Steinberg et al. 2015). Importantly, except for a few rare recurrent variants showing nominal association of their own, evidence for rare variant association with AD risk is generally obtained using burden tests gathering truncating variants with missense, predicted deleterious variants, at the gene level (Nicolas 2024). Burden tests performed on the category of truncating variants alone showed an exome-wide level of association for *SORL1* and *ABCA7* but only suggestive signals for *ABCA1* and *TREM2* (Holstege et al. 2022). For the latter genes, truncating variants remain extremely rare, explaining the difficulty to detect their association with AD risk due to insufficient statistical power in the current datasets. Importantly, genetic results are in line with known mechanisms, suggesting a deleterious effect of haploinsufficiency or loss of function of *SORL1*, *TREM2*, *ABCA7* and *ABCA1* on Amyloid β (Aβ) peptides aggregation. Likewise, extremely rare loss-of-function variants in the *TYROBP* gene encoding the main partner of TREM2, have recently been associated with AD risk (Stefansson et al. 2024).

While the most studied category of genetic variants in AD is represented by single nucleotide variants (SNV) or short insertions and deletions (indels), Copy Number Variants (CNV) have also been involved in AD. Duplications of the *APP* locus (Rovelet-Lecrux et al. 2006) and in-frame deletions of exon 9 or exon 9 and 10 of *PSEN1* (Crook et al. 1998; Karlstrom et al. 2008) cause autosomal-dominant AD. Some

medRxiv preprint doi: https://doi.org/10.1101/2024.10.28.24314051; this version posted November 1, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

common CNVs were identified by SNP arrays and associated with a modest risk of AD (Brouwers et al. 2012; Wang et al. 2022), but none of them was genome-wide significant. On the other hand, a handful of studies focusing on rare CNVs were performed using Comparative Genome Hybridization (CGH) arrays, on a limited number of EOAD patients (Hooli et al. 2014; Rovelet-Lecrux et al. 2012). Unique CNVs affecting genes of potential interest were prioritized, but segregation data in families were missing and such CNVs remained too rare to be assessed in case-control studies (Rovelet-Lecrux et al. 2012). Larger datasets should thus help detecting rare CNVs associated with AD risk.

Up to now, CNV studies in AD have mainly been performed using chip technologies, i.e. SNP arrays or CGH arrays, at a smaller scale. Such chips can detect CNVs with intermediate resolution, generally with a minimal size from 50 to 100 kb, depending on the array density. Thus, they classically miss smaller CNVs. Sequencing technologies can also be used for CNV detection and can detect CNVs of any size. Bioinformatics tools can detect CNVs from sequencing data based on four main approaches (Teo et al. 2012; Zhao et al. 2013): relative distribution of reads along the sequence or read depth approach, relative positions of paired reads from each other, multiple and partial alignments of reads and *de novo* assembly. When working from sequencing data obtained following capture, as for exome sequencing, only the read depth approach can be applied with a high level of accuracy at the exome level for rare CNVs (Gabrielaite et al. 2021; Hehir-Kwa, Pfundt, and Veltman 2015), while data obtained from whole genome sequencing require combinations of tools, also including read depth approaches although with distinct procedures and tools. We previously assessed the performances of a CNV-calling workflow based on the CANOES tool (Backenroth et al. 2014), which showed good sensitivity (87.25 %) and positive predictive value (85.2%) for rare CNVs from exome sequencing data (Quenez et al. 2021). One of the largest sequencing datasets available worldwide to perform a case-control study in AD is probably the combination of the European and American consortia into the ADES-ADSP dataset (Alzheimer Disease European Sequencing – Alzheimer Disease Sequencing Project), which recently unveiled burdens of rare variants in *ATP8B4* and *ABCA1* in addition to the previously known genes. This dataset has the advantage of being enriched in EOAD cases (35.7 %), among which we can expect a higher contribution of rare CNVs, as already observed for rare SNVs and indels (Holstege et al. 2022). Moreover, this dataset is made of a majority of exomes (68.6% after QC) and bioinformatic methods for CNV detection adapted to exome sequencing are mainly reliable for rare CNVs.

Here, we took advantage of the large number of exome sequencing data of EOAD cases and controls available in the ADES-ADSP dataset. We performed uniform CNV calling on 22,319 samples using a validated workflow and built up a QC pipeline to perform harmonized transcript-based analyses. We report CNVs in autosomal dominant genes, search for novel loci associated with AD, and study CNVs affecting genes known to harbor genetic factors.

medRxiv preprint doi: https://doi.org/10.1101/2024.10.28.24314051; this version posted November 1, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

METHODS

Exome sequencing dataset

We considered the exomes from ADES and ADSP datasets as described in ref. (Holstege et al. 2022) in a two-stage analysis. Only samples passing the quality control (QC) described in detail in ref. (Holstege et al. 2022) 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 (McKhann et al. 1984, 2011)), based on clinical examination 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 also excluded from stage-1 data in the Holstege et al. 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 (stages 1+2) fulfils the same criteria (Table S1). Detailed methods for individual QC are available as a supplementary information file in ref. (Holstege et al. 2022).

Because the ADES-ADSP dataset is built from multiple studies, and to ensure a good accuracy of CANOES (Backenroth et al. 2014), the CNV caller used here, we focused on samples as homogeneous as possible in batches of at least 50 individuals (Table S2). When library preparation and sequencing batches information were not available, we grouped samples belonging to a same study, prepared with a same capture kit, and sequenced in a same sequencing center, into CNV-calling batches. 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 (Holstege et al. 2022). 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.

medRxiv preprint doi: https://doi.org/10.1101/2024.10.28.24314051; this version posted November 1, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

CNV Calling

CNV calling was processed following a workflow centered on CANOES (Backenroth et al. 2014), a tool based on the distribution of the depth of coverage information across samples. It includes a correction based on GC content of each target to reduce the background variability often observed in NGS data (Benjamini and Speed 2012).

The workflow was applied from BAM files, as previously described (Quenez et al. 2021). BAM files were retrieved either from the pipeline described in Holstege et al. or directly downloaded from the dbGAP website (ADSP-stage 2, 1554 samples). All samples were processed on either the Cartesius supercomputer or the CEREBRO cluster from the sequencing facility of the University of Rouen Normandie (ASGARD platform). Of note, ADSP-stage 2 BAM files were aligned to the GRCh38 version of the human genome, whereas other BAM files were aligned on the GRCh37 version. As our analysis is based on transcripts affected by CNVs and because reference genome was not specific to cases or to controls, we expect that the different reference genomes do not affect our analysis. For each capture kit, targets covering the same exon and separated by less than 30 bp were merged. Then, for each sample, the number of reads covering each merged target was determined using BEDTools (Quinlan and Hall 2010). For each CNV-calling batch, we removed non-informative regions, namely regions where >90% of the samples showed less than 10 reads on the target. Finally, CANOES was run on each CNV-calling batch to generate CNV calls.

CNV and samples quality check

After CNV calling, we excluded all individuals with ≥50 calls, indicating an excess of variability in the sample's read distribution compared to other samples from the same CNV-calling batch and following CANOES user instructions (Backenroth et al. 2014). Then, to account for potential biases due to clonal hematopoiesis associated with large mosaic, blood-specific age-related CNVs (Bouzid et al. 2023; Machiela et al. 2017), we excluded carriers of CNVs that are large enough to be unlikely germline. As calling is performed from capture-based exomes and because of the presence of low complexity regions decreasing the calling accuracy, such large CNVs can be detected as multiple, smaller CNVs on a same chromosome. Thus, we computed the cumulative size of detected CNVs of the same type (deletion/duplication) for each sample and each chromosome, as well as the total chromosomal region covered from the first to last CNV per chromosome. For each cumulative size per chromosome greater than 2.5 Mb and encompassing more than 10 Mb for a given chromosome, we proceeded to a manual visualization (Figure S1) on the UCSC genome browser. Carriers of candidate large likely somatic CNVs were excluded from the analysis (Table S3).

medRxiv preprint doi: https://doi.org/10.1101/2024.10.28.24314051; this version posted November 1, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Annotation of CNVs

For CNV annotation, we first compared the frequency of CNV calls to two public databases: the Database of Genomic Variants (DGV), gold Standard section (MacDonald et al. 2014) and the non-neuro non-Finnish European section of the gnomAD database v2.1 (Chen et al. 2022; Karczewski et al. 2019), considering a mutual overlap of 70%. In addition, we marked CNV overlapping > 50% with segmental duplication regions, as extracted from the UCSC Table Browser (https://genome-euro.ucsc.edu/cgi-bin/hgTables). Second, in order to classify the potential effect of CNVs and to overcome the differences between capture kits that could impact the proper classification of some CNVs and thus the results of statistical analyses, we harmonized the dataset by working at the transcript level (see supplementary information, Figure S2). Indeed, classic CNV annotation tools may not be applied to heterogeneous datasets, as the definition of a complete versus partial gene duplication or deletion usually rely on the actual definition of a given gene, based on genome coordinates. As some capture kits vary on the regions actually targeted for a given gene (e.g. capture of untranslated regions or not, capture of all coding exons from all known transcripts or not), this may lead to heterogeneous definitions of partial or complete deletions/duplications (see for a theoretical example Figure S2). Thus, we built our own annotation pipeline that relied on each individual's capture kit to define whether a given CNV partially or completely affects a given transcript. Thus, we redefined the transcripts positions according to the capture kit used (see supplementary information, Figure S2). Consequently, a transcript was considered as completely deleted or duplicated if all targets linked to this transcript were affected by the event.

For all comparisons between capture kits, CNV calls and public databases, we used a combination of the BEDTools suite (Quinlan and Hall 2010) and homemade scripts. All scripts and formats used are available at: https://github.com/U1245/ExtremelyRareCNVContributingToADRisk (see also supplementary information).

Identification of Mendelian pathogenic CNVs

Based on current literature, we considered partial deletions of *PSEN1* involving exclusively exon 9 (NM_000021.4) or both exons 9 and 10 (Karlstrom et al. 2008; Le Guennec et al. 2017), complete duplications of *APP* (Rovelet-Lecrux et al. 2006), *MAPT* (Le Guennec et al. 2017), complete duplications or triplications of the *SNCA* gene (Miller et al. 2004; Singleton et al. 2003) as well as any coding deletion of the *GRN* gene (Rovelet-Lecrux et al. 2008) as pathogenic CNVs (causing either AD or a differential diagnosis). Individuals carrying such a pathogenic CNV were then excluded from case-control analyses.

medRxiv preprint doi: https://doi.org/10.1101/2024.10.28.24314051; this version posted November 1, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

For this analysis, we favored sensitivity over specificity and we kept and considered all calls from CANOES, contrary to the case-control analysis (see below).

Case-control analyses

CNV Filtering

This study focused on rare (frequency <1%) CNVs for two main reasons. First, rare CNVs, similar to rare SNVs and indels, have higher chances to be associated with a moderate to high AD risk (Holstege et al. 2022). Second, CNV-calling from exome sequencing data based on read-depth comparison has been essentially validated for rare CNVs (Quenez et al. 2021). Indeed, CANOES (and other read-depth comparison tools adapted to sequencing data obtained following capture) is based on the comparison of read depth on a specific target to a matrix computed from samples from a same batch. Thus, a very common variant (several dozens of percent in frequency) would likely be missed because of natural variability in each batch. Even if CANOES can detect some common variants, among those with a rather low frequency (e.g. 1-10%), its performances have not been assessed precisely on such variants.

To focus on rare CNVs, we filtered out all CNVs showing at least a 70% overlap with a common CNV (frequency >1% in the above-mentioned public databases). Similarly, we filtered out CNVs with a ≥50% overlap with segmental duplication regions, as these regions are considered as highly variable across individuals and hardly callable. Because of the complexity of X-linked models in a case-control analysis, we focused here on autosomes.

Finally, we removed CNVs overlapping only one target to decrease the risk of false positive CNV calls. Indeed, in our validation study (Quenez et al. 2021), the rate of false positive calls among CNVs overlapping one target reached 30% against 9.8% in CNVs called by ≥ 2 targets.

Transcripts filtering

To avoid the case-control analysis of rare CNVs affecting transcripts frequently deleted or duplicated, we built two sets of transcript, based on Refseq protein-coding transcripts (assessed, 26/10/2022): (i) a set of transcripts with a cumulative frequency of deletions (partial or complete) of less than 1% in public databases (set A, corresponding to non-frequently deleted transcripts) and (ii) a set of transcripts with a cumulative frequency of complete duplications of less than 1% in public databases (set B, corresponding to non-frequently duplicated transcripts), based on the DGV gold Standard section (MacDonald et al. 2014) and the non-neuro non-Finnish European section of the gnomAD database v2.1 (Chen et al. 2022; Karczewski et al. 2019).

medRxiv preprint doi: https://doi.org/10.1101/2024.10.28.24314051; this version posted November 1, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Building a copy number matrix

For the case-control analysis, we built a copy number matrix per autosomal protein-coding transcript, in order to harmonize heterogeneous data at the transcript level and to determine the missingness and hence which transcripts were eligible for case-control analysis. For each sample and each transcript remaining after filtration, five possible states were determined: (i) missing information, i.e. the corresponding capture kit does not target the transcript or the region was not covered enough (≤10 reads) for 90% of the samples and no CNV encompassing this transcript has been detected from targets on neighboring genes or (ii) no CNV overlaps the transcript [copy number = 2] or (iii) a complete or (iv) a partial duplication overlaps this transcript or (v) a complete or a partial deletion overlaps this transcript. In our five-state categorization, we worked under the assumption that complete and partial deletions likely result in haploinsufficiency. To explore the dosage effect, states (ii), (iii) and (v) were merged into a transcript copy number information (see supplementary methods for more details).

Statistical analyses

Given the previous knowledge obtained in gene-based rare variant analyses showing a larger effect among EOAD cases and given the expected extreme rarity of CNVs limiting power, we decided to set the EOAD cases versus controls as the primary analysis, to perform a gene dosage analysis, and we did so at the transcript level. For each transcript and each individual, the dosage information refers to the number of copies. Dosage equals 2 for individuals without any deletion nor full duplication affecting the transcript. Each duplication affecting the entire transcript increases the dosage by 1, whereas any deletion affecting partially or fully the transcript decreases the dosage by 1. To perform an EOAD cases vs controls transcript-based association study on protein-coding genes using a dosage strategy as the primary endpoint, we worked on the fusion of set A and set B and used rare partial and complete deletions and complete duplications. Dosage analyses were restricted to transcripts affected by a CNV in at least four carriers (including at least one deletion and one complete duplication) in our dataset and to the subset of individuals with available information relative to the transcript, i.e. individuals without missing information in the copy number matrix. Because of the rarity of CNVs, association was tested via firth logistic regression with status (EOAD vs controls) as dependent variable and dosage information (transcript copy number by individual) as independent variable. If the number of carriers of the different transcripts of a gene were similar, they were tested together. Otherwise, the transcripts were tested separately. We set a threshold at FDR=10% in this dosage EOAD analysis to consider signals as suggestive.

medRxiv preprint doi: https://doi.org/10.1101/2024.10.28.24314051; this version posted November 1, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Complementary analyses consisted in comparing dosages between all AD and controls at the *loci* prioritized in the EOAD versus control analysis and performing the following: (i) transcript-based analyses in a deletions-only (complete and partial deletions, on set A) and complete duplications-only (on set B) and (ii) a gene-based analysis of loss-of-function variants (deletion-CNVs plus truncating SNVs/indels, see supplementary methods) in a list of AD-associated genes and in genes prioritized by the dosage analysis. The list of AD-associated genes consisted of genes significantly associated with AD in a GWAS study (Bellenguez et al. 2022) and in the latest large gene-based case-control study on rare variants (Holstege et al. 2022), to which we added the recently identified *TYROBP* gene (Stefansson et al. 2024). For analyses mentioned in (i) above, regression was performed for each transcript from set A or B overlapping at least four deletions or four complete duplications in our dataset. In a sensitivity analysis, we adjusted for APOE4 or population stratification as captured by 10 PCs. See supplementary methods for more details.

Finally, we performed a cumulate CNV analysis of (i) all protein-coding genes, (ii) a list of AD associated genes in a GWAS (Bellenguez et al. 2022) and (iii) a list of genes having function in Aβ network (Campion et al. 2016). For that purpose, we compared EOAD versus controls in a Firth's logistic regression model. To overcome the variability in sequencing techniques, we excluded all transcripts with a significant differential missingness between disease status ($p<1E-5$ in a χ^2 test) from this analysis. (Supplemental methods)

CNV confirmation

The main analysis provided a list of prioritized transcripts/genes. Each transcript with an FDR below 10% was carefully checked in the UCSC genome browser showing all filtered and unfiltered CNVs separately at each transcript of interest, in cases and controls, along with the study and the capture kit information (Figure S3 to S7). This allowed us to detect signals driven by genes overlapping with repeats or duplicated gene in the genome despite CNVs not overlapping the 50% threshold set for repeats in the filtration steps, or genes for which the signal dosage analysis was driven by deletions whereas the transcript was not in set A, thus not relevant.

From each locus prioritized in the main analysis, we performed targeted validation of CNVs when DNA was available to us, using either ddPCR as previously described (Cassinari et al. 2019) or Quantitative Multiplex PCR of Short Fluorescent fragments (QMPSF) (Charbonnier et al. 2000).

Replication

medRxiv preprint doi: https://doi.org/10.1101/2024.10.28.24314051; this version posted November 1, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Loci of interest from the dosage analysis were assessed in a replication analysis based on genotyping data from the European Alzheimer & Dementia Biobank (EADB) and the UK biobank (UKBB).

The EADB study has previously been described in details (Bellenguez et al. 2022). From the EADB genotyping array data, we included 15,952 cases (2669 EOAD, 13,283 LOAD) and 20,335 controls not overlapping with ADES participants and after exclusion of relatives with a kinship >0.1 (supplementary information). CNV calling, QC and filtration procedures are described in supplementary information.

The UKBB has previously been described (Halldorsson et al. 2022). We selected Caucasian unrelated participants with available AD status (supplementary information). From the UKBB, we included 2050 cases (111 EOAD, 1939 LOAD) and 253,644 controls. CNV calling, QC and filtration procedures are described in supplementary information.

CNVs from EADB and UKBB were annotated for frequency similarly to our primary analysis. CNV data from EADB and UKBB were analyzed jointly in Firth's logistic regression models. Replication finally included the analysis of *ADI1* duplications and *FADS6* deletions as well as a dosage analysis for the 22q11 locus (*SCARF2-MED15-KLHL22*). Then, results from discovery and replication analyses were meta-analyzed with a fixed effect and the inverse variance weighted method.

medRxiv preprint doi: https://doi.org/10.1101/2024.10.28.24314051; this version posted November 1, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

RESULTS

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

Dataset, CNV calling and QC

We initially included 22,319 exomes from a previously described dataset (Holstege et al. 2022) (4,150 EOAD, 8,519 LOAD and 9,650 controls) (Figure 1). A description is available in supplementary information)

Following QC, we excluded 148 samples with excessive calls (≥ 50), 55 carriers of large likely somatic CNVs (26 controls, 28 LOAD cases, 1 EOAD case, see supplementary information, Table S2, Table S3), and 6 samples with biased amplification due to a custom kit. There was no significant difference when comparing large likely somatic CNVs between all cases and controls and likely somatic CNVs were linked to age at last visit (used as a proxy for age at blood sampling) (logistic regression; Age: p-value=6.78×10^{-6}; AD status: p-value=0.12). This suggests that such events are likely linked to clonal hematopoiesis and should thus be excluded from our germline CNV analysis study.

Among the 22,110 remaining individuals, we detected 261,190 CNVs (260,684 after exclusion of Mendelian pathogenic CNVs, see below). After excluding common CNVs (frequency>1%) and those overlapping with segmental duplication regions (see Supplementary information), a total of 61,053 rare CNVs remained, including 24,262 duplications and 36,791 deletions (Table 1, Figure 1).

As a supplement (extended dataset), we provide a catalog of genes with partial/complete deletions and duplications and frequencies among EOAD cases, LOAD cases, and controls.

Detection of Mendelian pathogenic CNVs

We observed 10 carriers of a complete duplication of *APP*, all with EOAD (mean age at onset (AAO) = 49.6 years, SD = 5.4, range: 43 to 58, three of them were previously reported (Lanoiselée et al. 2017; Rovelet-Lecrux et al. 2015). One patient carried a partial in-frame deletion in *PSEN1* (AAO=56 years old) (already reported (Le Guennec et al. 2017)). Six patients carried a complete duplication of *MAPT* (four of have been described previously (Le Guennec et al. 2017; Wallon et al. 2021), thus leading to a diagnosis a primary tauopathy (mean AAO = 58.2 years, SD = 14.9, range: 45 to 87). Of note, no carrier of a deletion of the *GRN* or any other AD differential diagnosis gene was detected, with the exception of the above-mentioned *MAPT* duplication carriers. All pathogenic CNV carriers are reported in Table 2.

medRxiv preprint doi: https://doi.org/10.1101/2024.10.28.24314051; this version posted November 1, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Case-control analyses

CNV analysis on the discovery exome sequencing dataset

After QC and exclusion of Mendelian CNV carriers, we kept 22,093 samples for case-control analyses, including 12,534 cases (4077 EOAD, 8457 LOAD) and 9559 controls and a total of 45,964 rare CNVs overlapping ≥2 targets of exome capture, outside of segmental duplication regions and affecting at least 1 coding transcript as described in RefSeq (23,134 deletions and 22,830 duplications).

To perform analyses based on transcript information, we focused on two sets of transcripts. Of 61,238 Refseq autosomal transcripts of 18,426 genes, set A contains 58,602 non-frequently deleted transcripts (related to 17,302 genes), of which 3299 (1040 genes) are affected by at least four deletions. Set B contains 60,220 non-frequently duplicated transcripts (17,921 genes), of which 3392 (1252 genes) are affected by at least four complete duplications. The fusion of sets A and B contains 60,633 transcripts (18,141 genes), including 4989 transcripts (1716 genes) affected by at least four CNVs (including at least one deletion and one complete duplication).

CNV-dosage EOAD-control analysis

Dosage analysis prioritized 56 transcripts from 21 genes with FDR ≤10% (Table 3, displaying one transcript per gene when several transcripts exhibit the same p-values). Three genes were then excluded because they either overlapped with repeats or duplicated regions in the genome, or the signal was driven by deletions while the deleted transcript was not in set A (corresponding to non-frequently deleted transcripts). Of the 44 remaining transcripts from 18 genes, five independent *loci* were identified, as 13 genes mapped to the 22q11.21 region (FDR from 0.0271 to 0.0386), and 2 other contiguous genes mapped to chromosome 19q13.32 (FDR from 0.0271 to 0.0412). All loci identified in the EOAD-control analysis were examined in an all AD versus control analysis, as depicted in Table S4.

The 22q11.21 region is known as a locus with rare recurrent deletions and duplications driven by non-allelic homologous recombination (NAHR) through low copy repeats (LCR) labelled LCR 22A, 22B, 22C, 22D and 22E (Figure 2). All carriers of a CNV encompassing this region are reported in table S5. The region where all transcripts with FDR≤10% clustered is located between LCR 22B and 22D. This region does not encompass the critical region for DiGeorge (velo-cardio-facial) 22q11.21 microdeletion syndrome (22A-22B including *TBX1*). Similar LCR 22B-22D deletions are often labelled *central deletions* and considered as a risk factor for developmental disorders, with incomplete penetrance and variable expression (Rump et al. 2014; Woodward et al. 2019). We observed deletions in this region in 4 EOAD cases, and this region was also prioritized in the deletion-only analysis (Table S6 displaying the top transcripts in the deletion-only analysis with FDR ≤20%). Of note, while 3/4 EOAD cases exhibited a

medRxiv preprint doi: https://doi.org/10.1101/2024.10.28.24314051; this version posted November 1, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

central deletion only, the fourth EOAD case actually carried a larger LCR 22A-22D deletion, typical of DiGeorge syndrome. This case has no specific history of learning disabilities or neurodevelopmental disorders, but he was born with congenital heart disease, without any other known features of this syndrome (Figure 2C). He presented first signs of cognitive impairment before the age of 45 and the diagnosis of AD was confirmed with CSF biomarkers (Aβ42, Tau, and 181-P-Tau levels all in abnormal ranges). DNA analysis using dPCR in this patient (EFA-429-001) confirmed the presence of the deletion. DNA from both unaffected parents was available and the deletion was absent, which, after parenthood confirmation using informative polymorphisms, allowed us to identify that the deletion occurred *de novo* in the proband (Figure 2). DNA of one of the other three EOAD cases (ROU-0165-001), carrying a central deletion, was also available and we confirmed the presence of the deletion by ddPCR (Table S8).

Interestingly, a smaller deletion was found in an LOAD case (LCR 22C-22D) and, strikingly, 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 (depending on the CNV coordinates) and in 4 to 5 LOAD cases (Figure 2), although the duplication signal was not significant enough to reach the 20% FDR threshold in the duplication-only analysis (Table S7). Overall, this suggests a dosage effect with deletions possibly increasing AD risk and duplications possibly decreasing the risk or delaying ages at onset.

Overall, the region with highest level of association based on deletions in EOAD cases was the LCR 22B-22D region, as was the case for the dosage analysis, including the mirror duplications. If considering all rare CNVs from this locus, a smaller minimal region could be defined between LCR 22C and 22D but this would be based on a single LOAD case and a single control. Thus, we conservatively consider that the region of interest, to be refined, remains the LCR 22B-22D region.

Another locus was identified with two contiguous genes involved, namely *ERCC2* (p-value=1.09×10^{-4}, FDR=0.0271) and *KLC3* (p-value=4.28×10^{-4}, FDR=0.0412). Despite showing FDR ≤10% in the dosage analysis, the signal was clearly driven by an enrichment of complete duplications in cases with an EOAD>LOAD>controls pattern. Most of the duplications at this locus shared similar coordinates, also encompassing the *PPP1R13L* gene but transcripts of this gene were not considered in the analysis because they were not predicted to be entirely duplicated (Figure S3). This locus is located on chromosome 19q13.32, ~440-kb away from the *APOE* gene. We observed that the duplications were significantly more frequently carried by *APOE*4+ individuals, independently of the disease status (Table S9), suggesting that rare duplications may have occurred on an *APOE4*-linked haplotype (Fisher's exact test p-value=4.26×10^{-13}). Consistent with this hypothesis, the association signal did not remain significant after adjusting for *APOE*-ε4 dosage (OR=1.58 [0.85; 2.99], p=0.148) while such adjustment

medRxiv preprint doi: https://doi.org/10.1101/2024.10.28.24314051; this version posted November 1, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

did not change the results on the other transcripts (Table S10). Of note, DNA from 12 carriers was available and duplications were confirmed in all of them (Table S8).

In addition to these two multi-gene loci, three single-gene loci showed an FDR≤10%, namely *MBL2* (FDR=0.0271), *FADS6* (FDR=0.0271) and *ADI1* (FDR=0.0916). For *MBL2* and *FADS6*, the dosage signal was mainly driven by deletions. *MBL2* deletions appeared to be enriched in controls whereas *FADS6* deletions appeared to be enriched in EOAD cases (Supplemental information, Figures S4, S6). However, enrichment observed for *MBL2* did not remain after adjusting for population structure (Supplemental results, Figure S14 Table S11). For *ADI1*, complete duplications appeared to be enriched in cases (Supplemental information, Figure S5). DNA from 13 carriers (1 *MBL2* carrier, 3 *ADI1* and 9 *FADS6*) was available and CNVs were confirmed in 12 of them (one *FADS6* false positive) (Table S8). After removing the latter one, EOAD versus controls dosage analysis remained similar (OR=0.20 [0.07; 0.49], p-value=4×10^{-4}).

Joint CNV and loss-of-function indels and SNVs analyses

Then, we hypothesized that the signal driven by deletions could be reinforced by LOF SNVs and indels or that such small variants could help refine genes of interest at the 22q11.21 locus. We thus performed a joint analysis of CNVs (deletions only, complete and partial) with LOF SNVs and indels among (i) the genes belonging to the prioritized *loci* in the dosage analysis and (ii) a list of AD-associated genes (Table S12 and S13).

Among the candidate *loci* identified in the dosage analysis, the dosage signal seemed to be driven by duplications regarding *ADI1* and the *ERCC2-KLC3* genes. For all three genes, there was no LOF association overall in the joint analysis, which is consistent with the *ERCC2-KLC3* duplications found as enriched in APOE4+ individuals, thus not directly linked to these genes. No LOF SNV/indel was identified in the *FADS6* gene.

In the 22q11.21 central deletion locus with 13 genes involved, the joint LOF analysis trended to narrow the locus of interest to *SCARF2*, *MED15* and *KLHL22*. Although the added LOF SNV/indels did not significantly reinforce these genes, with a single splice site short delins in *SCARF2* in an EOAD case (NM_153334.4:c.1424+1_1424+2delinsTC) and 2 *MED15* LOF SNV/indel variants in 2 LOAD cases (one nonsense, one frameshift 1-bp deletion), and no control carrying LOF SNV/indels (as for deletions), but all the other genes appeared as less significant now with a few LOF variant carriers in controls and not significantly more in cases.

We then focused on known AD-associated genes. Among them, we identified a significant enrichment of LOF variants of *ABCA1* (uncorrected p=0.0002) and confirmed the association of *ABCA7* LOF variants (p=0.0006), now adding deletion-CNVs to the known association of *ABCA7* LOF variants with AD

medRxiv preprint doi: https://doi.org/10.1101/2024.10.28.24314051; this version posted November 1, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

(Steinberg et al. 2015), while the *CTSB* gene, mapping to a GWAS locus, also appeared as a candidate (p=0.0089). In *ABCA1*, we observed 3 carriers of 3 distinct deletions in *ABCA1,* impacting respectively exons 16 to 18, 32 to 34 and 47 to 50 (Figure S7, all confirmed by ddPCR). All were EOAD patients (AAO: 49, 51 and 55). The joint OR of *ABCA1* LOF variants was 5.77 [2.25; 17.06], suggesting that LOF of *ABCA1* is a moderate EOAD risk factor, while burden tests gathering LOF SNV/indels with missense variants showed more modest odds ratios (Holstege et al. 2022). Overall, LOF variants in *ABCA1* remained extremely rare with only 13 EOAD cases (0.32%), 12 LOAD (0.14%) and 5 controls (0.05%) carrying such a variant. Deletions represented 10% (3/30) of the *ABCA1*-LOF alleles. Deletions in *ABCA7* were observed in 4 EOAD cases (0.09%), 3 LOAD (0.04%) and 3 controls (0.03%). They represented 8.6% (10/115) of the LOF alleles of this gene, whereas a complete duplication of *ABCA7* was observed in one control (0.01%) and 2 LOAD cases (0.02%) (Figure S8). Deletions did not affect the order of magnitude of known AD or EOAD association of *ABCA7* with LOF variants in terms of OR, given the higher frequency of *ABCA7* LOF variants overall. DNA was available for all 3 *ABCA1* and 2 *ABCA7* deletion carriers and confirmed the existence of each deletion. In addition, the recently identified *TYROBP* gene showed 4 deletions in 4 LOAD cases and none in controls, although no EOAD was a carrier and no LOF SNV/indel was identified (Figure S9).

Interestingly, LOF variants of *MAF*, *PLEKHA1* and *EPDR1* genes (incl. one deletion for each gene) were only detected in controls, leading to the hypothesis of a protective effect of LOF of these genes, which cannot be confirmed here as such events are extremely rare (Table S13, Figure S10). CNVs covering *CTSB* and *APH1B* genes suggested a dosage effect with deletions of these genes observed in EOAD patients only whereas duplications were observed in controls only; although these genes did not appear in the top 10% FDR in the main dosage exome-wide analysis. Extending the analysis of these genes to LOAD, we observed 4 LOAD patients carrying a deletion of *CTSB* as well as 1 LOAD patient (AAO=90) carrying a duplication (Figure S11). Of note, there were no carriers of CNVs encompassing the *APH1B* gene among LOAD, and three controls carried a LOF SNV/indel.

Cumulate CNV analysis of all transcripts

Considering CNVs overall in all transcripts from sets A and B, EOAD patients were more likely to carry at least one deleted or one duplicated gene than controls (burden of total and partial deletions: OR=1.15 [1.06; 1.23], p-value=3×10^{-4}; burden of complete duplications: OR=1.10 [1.02; 1.19], p-value=0.0125). Restricting the analysis to the GWAS list of genes (Bellenguez et al. 2022), difference between EOAD and controls was increased for deletions (OR=2.67 [1.51; 4.74], p-value=8×10^{-4}) whereas the difference trended to be reversed for duplications (OR=0.58 [0.30; 1.04], p-value=6.9×10^{-3}) suggesting that a deletion in the GWAS list of genes might be a risk factor for EOAD, that complete

medRxiv preprint doi: https://doi.org/10.1101/2024.10.28.24314051; this version posted November 1, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

duplications of some of these genes could be protective (Table S14), and that duplications in other genes remain to be identified. Finally, analysis of the Aβ network list of genes (Campion et al. 2016) suggested that having a deletion encompassing genes from this list might increase the risk for EOAD (OR=1.44 [1.09; 1.90], p=0.0114), and more particularly genes involved in the processing and trafficking of APP and genes involved in calcium homeostasis (OR=1.64 [1.04; 2.57], p=0.0336 and OR=16.42 [1.59; 2208.35], p=0.0161, respectively).

Replication

Joint analysis of EADB and UKBB datasets included DNA chip data from 2,780 EOAD patients, 15,222 LOAD patients and 273,979 controls. Replication of three signals was sought: dosage at the chr22q11 locus (*SCARF2*, *KLHL22* and *MED15* genes), enrichment of duplications in *ADI1* in cases, and enrichment of deletions of *FADS6* in cases. Overall, replication results were consistent with the discovery analysis, but the rarity of CNVs and the relatively small sample size of EOAD cases in the replication analysis prevented us to replicate our findings in an EOAD versus controls analysis (Supplementary table S15).

In the all AD versus controls analysis, dosage at the 22q11.21 showed that the risk of AD decreases when copy number increases (p-values ranging from 3.44×10^{-3} and 2.15×10^{-2} depending on the genes in the replication analysis, meta-analysis p-values ranging from 8.11×10^{-4} to 2.67×10^{-3} Figure 1, Supplementary table S15). Of note, 9/273,979 controls carried a deletion encompassing *SCARF2*, *KLHL22* and *MED15* genes, all coming from the UKBB cohort.

FADS6 deletions, which were enriched in cases in the discovery analysis, were extremely rare in the replication analysis. However, the meta-analysis still showed an OR of 6.92 [2.72; 17.61] ($p=4.99\times10^{-5}$) among EOAD patients and 4.86 [2.15; 10.98] ($p= 1.47\times10^{-4}$) among all AD patients. Importantly, the signal in the discovery analysis was driven by a rather small recurrent deletion (<10kb), which is too small to be accurately detected in DNA chip data.

Finally, although the odds ratio of *ADI1* duplications remained greater than 1, as in the discovery, the difference was not significant in the replication analysis (EOAD vs Controls meta-analysis: OR=6.33 [1.53; 26.27], $p=1.10\times10^{-2}$; all AD vs controls meta-analysis OR= 2.23 [0.88; 5.64], $p=8.98\times10^{-2}$).

medRxiv preprint doi: https://doi.org/10.1101/2024.10.28.24314051; this version posted November 1, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

DISCUSSION

Rare CNVs have been recognized as a disease mechanism for Mendelian disorder for decades, in the context of multiple diverse human disorders (e.g. (Hehir-Kwa et al. 2015; Lupski et al. 1991; Shen et al. 2010)) and this includes neurodegenerative disorders. Duplications or triplications of the *SNCA* gene are responsible for autosomal dominant Parkinson disease or Dementia with Lewy Bodies (Singleton et al. 2003), and *APP* duplications or triplications, EOAD with or without cerebral amyloid angiopathy (Grangeon et al. 2021; Rovelet-Lecrux et al. 2006), for example. However, in most Mendelian neurodegenerative disorders, CNVs represent an extremely rare mechanism. For example, only one *APP* triplication has been reported so far (Grangeon et al. 2021) while *APP* duplications represent 9% of all solved autosomal dominant EOAD families in the national French series (update data from (Lanoiselée et al. 2017)) and *MAPT* duplications represent a very rare cause of primary tauopathy that can mimic AD (Wallon et al. 2021), as found in a few patients here.

In line with Mendelian dementia genes, we show that rare CNVs can also contribute to AD risk in non-monogenic forms. They also represent a rare mechanism, as in the know AD risk genes *ABCA1* and *ABCA7*, where deletions represented a small proportion of loss-of-function (LOF) alleles. In *ABCA7*, the burden of LOF SNV/indels was already known to be associated with AD. Joint analysis of CNVs and LOF SNV/indels did not modify the odds ratios much, remaining in the order of magnitude of 2 to 3 (Holstege et al. 2022; Le Guennec et al. 2016), because (i) LOF alleles are not so rare (0.37% in controls; 0.86% in EOAD cases) and (ii) CNVs represented only 8.6% of all LOF alleles in the joint analysis. In *ABCA1*, where LOF variants are much rarer, we identified three distinct deletions, adding up to 27 LOF SNVs/indels. Deletions thus represent 10% of all LOF alleles. This analysis on *ABCA1* now allows us to consider LOF of *ABCA1* as a stronger risk factor for EOAD, compared to the known average odds ratios obtained by gathering LOF variants with missense, predicted damaging variants (OR=2.2 [1.6; 2.9] for LOF SNV/indels and missense variants with REVEL score >0.75 in ref. (Holstege et al. 2022) and that also include missense variants with a demonstrated effect on AD risk, Tangier disease and HDL cholesterol deficiency (Frikke-Schmidt et al. 2008; Nordestgaard et al. 2015), and OR=5.77 [2.25; 17.06] here in the joint LOF SNV/indels/CNV analysis. These results suggest that (i) some CNVs represent an extremely rare mechanism increasing the risk of AD but also that (ii) at the individual level, given the effect on AD risk of such LOF alleles, CNVs should not be ignored. Although risk variants are not used for genetic counseling, they may be used for the future of AD prevention in the context of precision medicine, along with other factors (Nicolas et al. 2024). In addition, we assessed the presence of LOF alleles in the recently AD-associated gene *TYROBP* (Stefansson et al. 2024). Four deletions were identified in LOAD cases, none in EOAD cases and none in controls. Three of the deletions likely shared similar breakpoints as a recurrent deletion described in Finland and also enriched in AD cases

medRxiv preprint doi: https://doi.org/10.1101/2024.10.28.24314051; this version posted November 1, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

(Martiskainen et al. 2024; Stefansson et al. 2024). Interestingly, we also detected a suggestive association of *CSTB* deletions with AD risk. *CTSB* is a known GWAS locus and this gene encodes the Cathepsin B protein, which is known as an antiamyloidogenic enzyme through proteolytic cleavage of Aβ peptides(Mueller-Steiner et al. 2006).

Beyond known AD risk genes, we identified several novel candidate loci here. Of note, in the protein-coding genome-wide analysis, one of the suggestive associations pointed to the *APOE* locus, which was not expected from a rare CNV analysis. However, this unexpected result, suggesting that rare duplications occurred on an *APOE4*-associated haplotype, further strengthen the validity of our global analysis. It is also a rare example of a rare recurrent CNV in linkage disequilibrium with a common GWAS locus, highlighting that this type of event can actually happen and should be considered in genome-wide analyses.

In the dosage analysis at the protein-coding genome level, the 22q11.21 locus showed a dosage effect with an enrichment of deletions in EOAD cases and duplications in controls, while LOAD cases remained in between (Figure 2B). Although this region is not known to be associated with AD risk, a family-based analysis showed a suggestive signal (Lee et al. 2008). Our results suggest that dosage of some of the encompassed genes may influence AD risk or age at onset. Two main hypotheses can be made to explain such an association. First, the association could be nonspecific to AD mechanisms but rather a modifier of ages of onset in individuals prone to develop AD. Indeed, carriers of the so-called central 22q11.21 deletion have a higher risk of neurodevelopmental disorders (Rump et al. 2014; Woodward et al. 2019), as do the carriers of the full 22q11.21 deletion with higher penetrance (DiGeorge syndrome) and lower average cognitive performances (Huguet et al. 2018). Thus, one could hypothesize that carriers may have a lower cognitive reserve, even if not reaching the threshold for a formal diagnosis of a neurodevelopmental disorder, but still sufficient to decrease the age at onset of AD in the carriers (Nelson et al. 2021; Soldan et al. 2017; Stern 2012). However, this hypothesis would require that duplication carriers have better average cognitive performances than non-carriers, which is not known up to now, B-D duplications being currently considered as variants of uncertain significance. Second, dosage of genes in this interval could be directly influencing the pathophysiology of AD. There is no gene obviously linked to AD mechanisms here, but we can still hypothesize that genes like *MED15* or *SCARF2* may be related somehow to AD pathophysiology. *SCARF2* encodes a scavenger receptor (class F scavenger receptor family) expressed in the brain and in macrophages, although a scavenger receptor role for SCARF2 is yet to be demonstrated, contrary to other members of this family including SCARF1. Interestingly, another member of this family, MEGF10, has been shown to be a receptor for Aβ in the brain (Singh et al. 2010; Wilkinson and El Khoury 2012). MED15 is also expressed in the brain and belongs to a family of proteins related to other genes' expression, including

medRxiv preprint doi: https://doi.org/10.1101/2024.10.28.24314051; this version posted November 1, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

inflammation and TGFβ and SMAD2/3 signal transduction, both involved in AD pathophysiology (von Bernhardi et al. 2015).

Importantly, our results on the 22q11.2 locus may be of interest for patients carrying larger 22q11.2 deletions (DiGeorge syndrome, the critical region of which is A-B), among those with large deletions encompassing the *SCARF2-KLHL22-MED15* region (A-C deletions or larger). Patients with DiGeorge syndrome present a variety of symptoms with incomplete penetrance, including neurodevelopmental disorders ranging from mild intellectual disability, psychosis, and autism spectrum disorder to normal intelligence, associated or not with congenital heart defects, palatal malformations, velopharyngeal incompetence, phosphocalcic metabolism disorders, immune deficiency and some morphological features. In addition, increased risk for early-onset Parkinson disease has been reported in DiGeorge syndrome patients (Butcher et al. 2013; Mok et al. 2016). However, not much is known about the aging of patients with DiGeorge syndrome, as most reports in adults involve individuals in their mid-30s on average and probability of survival to age 45 is 95% and 72% respectively for patients with or without major congenital heart disease (Boot et al. 2023). AD-related changes have already been reported in the neuropathological examination of at least one DiGeorge syndrome patient at the age of 56 years (Butcher et al. 2013), but neuropathological examinations of adults remain scarce. Overall, our results suggest that, as care of DiGeorge syndrome is improving, enabling patients to age, a cognitive follow-up should be proposed, especially to those with a large deletion (with distal breakpoint being C at least).

This finding of 22q11.2 association with AD is somehow reminiscent of the Down syndrome patients with three copies of the *APP* gene on chromosome 21 and risk of EOAD, similar to patients with microduplications at the *APP* locus not encompassing the critical Down syndrome region (Grangeon et al. 2021; Rovelet-Lecrux et al. 2006). However, the impact of *APP* duplications may be much stronger, with almost full penetrance by the age of 65 for *APP* duplication carriers and Down syndrome patients (Fortea et al. 2020; Grangeon et al. 2023; Iulita et al. 2022). This can probably be explained by the obvious and direct link between the overexpression of the *APP* gene encoding the precursor of the Aβ peptide and AD pathophysiology, while it remains to be determined what explains AD association/protection at the 22q11.2 locus.

Notably, some controls carrying at least central 22q11.2 deletions (B-C or B-D) were identified in the UK biobank. Among those having available information for parents, 28.57% (2/7) of them 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 are recurrently found as *de novo* events, hence with no meaning of family history. Beyond the fact that

medRxiv preprint doi: https://doi.org/10.1101/2024.10.28.24314051; this version posted November 1, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

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.

In our analysis, *FADS6* deletions were associated with AD risk. Although the signal was driven by a recurrent deletion that was too small to be replicated in the DNA chip dataset, the results remained significant in the meta-analysis thanks to larger deletions identified in the replication dataset. *FADS6* encodes the fatty acid desaturase 6, which is expressed in the brain. Its pathophysiological roles remain unclear, but it has been shown that FADS6 expression is necessary for neuronal survival and neuroprotection against Tau phosphorylation after BACE1 silencing in cortical primary cultures (Villamil-Ortiz and Cardona-Gómez 2018).

Regarding ADI1 duplications, which were enriched in cases, the rarity and the relatively low number of EOAD cases in the replication dataset precluded the replication, although frequencies of duplications in cases and controls remained consistent with the discovery.

These results suggests that these loci should be replicated in even larger EOAD datasets with available sequencing data.

In the analysis at the protein-coding genome level, we identified an enrichment of deletions and of duplications in cases. After focusing on the GWAS list of genes, the signal based on duplications trended to be reversed. This suggests that additional genes, not identified here, may play a role in AD determinism, and this highlights a limitation of analyzing CNVs as a whole, at the protein-coding genome or at gene-list levels, as increasing expression or decreasing expression of genes may show opposite effects on AD pathophysiology. For example, *APP*, *ABCA7* and *ABCA1* are all AD GWAS hits, but gathering their effect into a same analysis (if *APP* duplications would not be excluded prior to case-control studies), would lead to non-interpretable results, as duplications of *APP* are expected to be enriched in cases but, conversely, deletions of *ABCA1* or *ABCA7* are expected to be enriched in cases. Similarly, we investigated genes related to the Aβ network. Although deletions suggestively contribute to AD, signal is probably confused by opposite effects of a decreased expression of these genes on AD. Overall, our analysis suggest that additional genes remain to be identified despite power limitations when working with rare CNVs.

In this study, we managed to detect and analyze jointly CNVs from a heterogeneous exome dataset of 22,319 individuals. We propose a quality control and analysis strategy based first on an extensive individual and sample QC from previous work (Holstege et al. 2022) and, second, on CNV-specific features. Our QC allowed us to obtain a dataset harmonized at the transcript level that allowed (i) the identification of high-quality CNV calls in known AD genes (Mendelian and risk factor genes) and (ii) identification of candidate loci, despite the extreme rarity of the individual CNVs under study. Despite

medRxiv preprint doi: https://doi.org/10.1101/2024.10.28.24314051; this version posted November 1, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

this design was set to identify effects at the transcript/gene level with a diversity of CNVs putatively affecting a same gene, the most significant finding was a multiple-gene recurrent CNV mediated by NAHR. This is probably related to a power issue in finding a burden of ultra-rare CNVs in a given gene, but this study design may be successful in finding such signals in future larger studies.

Here, we used a large existing dataset of exome sequencing data. Whole genome sequencing (WGS) with short reads may also be used to replicate such findings, although at a high computational cost. Such an analysis has been recently run on WGS of 6,646 AD cases (average age at onset: 74.6 years) and 6,938 controls from the ADSP consortium (Lee et al. 2021; Wang et al. 2023). A moderate but significant burden of (coding and non-coding) CNVs was associated with AD status overall and appeared higher for rarest events (singletons). In addition, an aggregated analysis of structural variants affecting a list of known AD genes also unveiled an association with AD status. Interestingly, these structural variants included one coding partial deletion of *SORL1* and 5 coding partial deletions of *ABCA7* (all in cases), among a few other genes of interest. None of the CNVs was significantly associated with AD due to power limitations, and possibly because this cohort is not enriched in EOAD cases. Interestingly, non-coding structural variants, which are not detectable by exome sequencing, were also analyzed, prioritizing for example an intronic deletion in the *ADD3* gene (11 cases and no control) and another intronic deletion in the *ITPR2* gene (33 cases, 7 controls), suggesting that increasing sample sizes with WGS data should unveil novel associations. Novel long read sequencing technologies may also offer new opportunities. In addition to CNVs, such novel technologies will enable the analysis of mobile element insertions, repeat variations, and balanced structural variants with an unprecedented accuracy (Logsdon, Vollger, and Eichler 2020; Miller et al. 2021). However, sequencing costs remain high, limiting the expectations that we can have from such a promising technology given the rarity of the events with a putatively strong effect.

In conclusion, we conducted an exome-wide CNV screen from sequencing data and identified ultra-rare CNVs that contribute to AD risk. Beyond autosomal dominant known CNVs, we highlight rare deletions in *ABCA1* and *ABCA7* as participating to AD risk in carriers, we identified loss-of-function of *CTSB* as a candidate mechanism increasing AD risk and, more importantly, a small region at the 22q11.21 locus, of which deletions increase EOAD risk and duplications decrease the risk of AD. This is a rare example of a dosage effect with opposite associated phenotypic features, which may also be important for the follow up of some patients with DiGeorge syndrome. While additional CNVs may contribute to AD risk, as suggested by the gene-list and protein-coding genome-wide analyses, such CNVs remain extremely rare events. We used a transcript point of view to reduce the recurrence issue at the CNV level and to reduce the heterogeneity among datasets, but the rarity of such events remains an issue even in datasets as large as the one used here. Finally, we are providing a catalog of genes

medRxiv preprint doi: https://doi.org/10.1101/2024.10.28.24314051; this version posted November 1, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

with rare CNVs in EOAD, LOAD and controls, which can be reused for meta-analyses and combination with other datasets.

ACKNOWLEDGEMENTS

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

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.

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.

This research has been conducted using data from UK Biobank, a major biomedical database.

medRxiv preprint doi: https://doi.org/10.1101/2024.10.28.24314051; this version posted November 1, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

REFERENCES

Backenroth, Daniel, Jason Homsy, Laura R. Murillo, Joe Glessner, Edwin Lin, Martina Brueckner, Richard Lifton, Elizabeth Goldmuntz, Wendy K. Chung, and Yufeng Shen. 2014. 'CANOES: Detecting Rare Copy Number Variants from Whole Exome Sequencing Data'. *Nucleic Acids Research* 42(12):e97. doi: 10.1093/nar/gku345.

Bellenguez, Céline, Fahri Küçükali, Iris E. Jansen, Luca Kleineidam, Sonia Moreno-Grau, Najaf Amin, Adam C. Naj, Rafael Campos-Martin, Benjamin Grenier-Boley, Victor Andrade, Peter A. Holmans, Anne Boland, Vincent Damotte, Sven J. van der Lee, Marcos R. Costa, Teemu Kuulasmaa, Qiong Yang, Itziar de Rojas, Joshua C. Bis, Amber Yaqub, Ivana Prokic, Julien Chapuis, Shahzad Ahmad, Vilmantas Giedraitis, Dag Aarsland, Pablo Garcia-Gonzalez, Carla Abdelnour, Emilio Alarcón-Martín, Daniel Alcolea, Montserrat Alegret, Ignacio Alvarez, Victoria Álvarez, Nicola J. Armstrong, Anthoula Tsolaki, Carmen Antúnez, Ildebrando Appollonio, Marina Arcaro, Silvana Archetti, Alfonso Arias Pastor, Beatrice Arosio, Lavinia Athanasiu, Henri Bailly, Nerisa Banaj, Miquel Baquero, Sandra Barral, Alexa Beiser, Ana Belén Pastor, Jennifer E. Below, Penelope Benchek, Luisa Benussi, Claudine Berr, Céline Besse, Valentina Bessi, Giuliano Binetti, Alessandra Bizarro, Rafael Blesa, Mercè Boada, Eric Boerwinkle, Barbara Borroni, Silvia Boschi, Paola Bossù, Geir Bråthen, Jan Bressler, Catherine Bresner, Henry Brodaty, Keeley J. Brookes, Luis Ignacio Brusco, Dolores Buiza-Rueda, Katharina Bûrger, Vanessa Burholt, William S. Bush, Miguel Calero, Laura B. Cantwell, Geneviève Chene, Jaeyoon Chung, Michael L. Cuccaro, Ángel Carracedo, Roberta Cecchetti, Laura Cervera-Carles, Camille Charbonnier, Hung-Hsin Chen, Caterina Chillotti, Simona Ciccone, Jurgen A. H. R. Claassen, Christopher Clark, Elisa Conti, Anaïs Corma-Gómez, Emanuele Costantini, Carlo Custodero, Delphine Daian, Maria Carolina Dalmasso, Antonio Daniele, Efthimios Dardiotis, Jean-François Dartigues, Peter Paul de Deyn, Katia de Paiva Lopes, Lot D. de Witte, Stéphanie Debette, Jürgen Deckert, Teodoro Del Ser, Nicola Denning, Anita DeStefano, Martin Dichgans, Janine Diehl-Schmid, Mónica Diez-Fairen, Paolo Dionigi Rossi, Srdjan Djurovic, Emmanuelle Duron, Emrah Düzel, Carole Dufouil, Gudny Eiriksdottir, Sebastiaan Engelborghs, Valentina Escott-Price, Ana Espinosa, Michael Ewers, Kelley M. Faber, Tagliavini Fabrizio, Sune Fallgaard Nielsen, David W. Fardo, Lucia Farotti, Chiara Fenoglio, Marta Fernández-Fuertes, Raffaele Ferrari, Catarina B. Ferreira, Evelyn Ferri, Bertrand Fin, Peter Fischer, Tormod Fladby, Klaus Fließbach, Bernard Fongang, Myriam Fornage, Juan Fortea, Tatiana M. Foroud, Silvia Fostinelli, Nick C. Fox, Emlio Franco-Macías, María J. Bullido, Ana Frank-García, Lutz Froelich, Brian Fulton-Howard, Daniela Galimberti, Jose Maria García-Alberca, Pablo García-González, Sebastian Garcia-Madrona, Guillermo Garcia-Ribas, Roberta Ghidoni, Ina Giegling, Giaccone Giorgio, Alison M. Goate, Oliver Goldhardt, Duber Gomez-Fonseca, Antonio González-Pérez, Caroline Graff, Giulia Grande, Emma Green, Timo Grimmer, Edna Grünblatt, Michelle Grunin, Vilmundur Gudnason, Tamar Guetta-Baranes, Annakaisa Haapasalo, Georgios Hadjigeorgiou, Jonathan L. Haines, Kara L. Hamilton-Nelson, Harald Hampel, Olivier Hanon, John Hardy, Annette M. Hartmann, Lucrezia Hausner, Janet Harwood, Stefanie Heilmann-Heimbach, Seppo Helisalmi, Michael T. Heneka, Isabel Hernández, Martin J. Herrmann, Per Hoffmann, Clive Holmes, Henne Holstege, Raquel Huerto Vilas, Marc Hulsman, Jack Humphrey, Geert Jan Biessels, Xueqiu Jian, Charlotte Johansson, Gyungah R. Jun, Yuriko Kastumata, John Kauwe, Patrick G. Kehoe, Lena Kilander, Anne Kinhult Ståhlbom, Miia Kivipelto, Anne Koivisto, Johannes Kornhuber, Mary H. Kosmidis, Walter A. Kukull, Pavel P. Kuksa, Brian W. Kunkle, Amanda B. Kuzma, Carmen Lage, Erika J. Laukka, Lenore Launer, Alessandra Lauria, Chien-Yueh Lee, Jenni Lehtisalo, Ondrej Lerch, Alberto Lleó, William Longstreth, Oscar Lopez, Adolfo Lopez de Munain, Seth Love, Malin Löwemark, Lauren Luckcuck, Kathryn L. Lunetta, Yiyi Ma, Juan Macías, Catherine A.

medRxiv preprint doi: https://doi.org/10.1101/2024.10.28.24314051; this version posted November 1, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

MacLeod, Wolfgang Maier, Francesca Mangialasche, Marco Spallazzi, Marta Marquié, Rachel Marshall, Eden R. Martin, Angel Martín Montes, Carmen Martínez Rodríguez, Carlo Masullo, Richard Mayeux, Simon Mead, Patrizia Mecocci, Miguel Medina, Alun Meggy, Shima Mehrabian, Silvia Mendoza, Manuel Menéndez-González, Pablo Mir, Susanne Moebus, Merel Mol, Laura Molina-Porcel, Laura Montrreal, Laura Morelli, Fermin Moreno, Kevin Morgan, Thomas Mosley, Markus M. Nöthen, Carolina Muchnik, Shubhabrata Mukherjee, Benedetta Nacmias, Tiia Ngandu, Gael Nicolas, Børge G. Nordestgaard, Robert Olaso, Adelina Orellana, Michela Orsini, Gemma Ortega, Alessandro Padovani, Caffarra Paolo, Goran Papenberg, Lucilla Parnetti, Florence Pasquier, Pau Pastor, Gina Peloso, Alba Pérez-Cordón, Jordi Pérez-Tur, Pierre Pericard, Oliver Peters, Yolande A. L. Pijnenburg, Juan A. Pineda, Gerard Piñol-Ripoll, Claudia Pisanu, Thomas Polak, Julius Popp, Danielle Posthuma, Josef Priller, Raquel Puerta, Olivier Quenez, Inés Quintela, Jesper Qvist Thomassen, Alberto Rábano, Innocenzo Rainero, Farid Rajabli, Inez Ramakers, Luis M. Real, Marcel J. T. Reinders, Christiane Reitz, Dolly Reyes-Dumeyer, Perry Ridge, Steffi Riedel-Heller, Peter Riederer, Natalia Roberto, Eloy Rodriguez-Rodriguez, Arvid Rongve, Irene Rosas Allende, Maitée Rosende-Roca, Jose Luis Royo, Elisa Rubino, Dan Rujescu, María Eugenia Sáez, Paraskevi Sakka, Ingvild Saltvedt, Ángela Sanabria, María Bernal Sánchez-Arjona, Florentino Sanchez-Garcia, Pascual Sánchez Juan, Raquel Sánchez-Valle, Sigrid B. Sando, Chloé Sarnowski, Claudia L. Satizabal, Michela Scamosci, Nikolaos Scarmeas, Elio Scarpini, Philip Scheltens, Norbert Scherbaum, Martin Scherer, Matthias Schmid, Anja Schneider, Jonathan M. Schott, Geir Selbæk, Davide Seripa, Manuel Serrano, Jin Sha, Alexey A. Shadrin, Olivia Skrobot, Susan Slifer, Gijsje J. L. Snijders, Hilkka Soininen, Vincenzo Solfrizzi, Alina Solomon, Yeunjoo Song, Sandro Sorbi, Oscar Sotolongo-Grau, Gianfranco Spalletta, Annika Spottke, Alessio Squassina, Eystein Stordal, Juan Pablo Tartan, Lluís Tárraga, Niccolo Tesí, Anbupalam Thalamuthu, Tegos Thomas, Giuseppe Tosto, Latchezar Traykov, Lucio Tremolizzo, Anne Tybjærg-Hansen, Andre Uitterlinden, Abbe Ullgren, Ingun Ulstein, Sergi Valero, Otto Valladares, Christine Van Broeckhoven, Jeffery Vance, Badri N. Vardarajan, Aad van der Lugt, Jasper Van Dongen, Jeroen van Rooij, John van Swieten, Rik Vandenberghe, Frans Verhey, Jean-Sébastien Vidal, Jonathan Vogelgsang, Martin Vyhnalek, Michael Wagner, David Wallon, Li-San Wang, Ruiqi Wang, Leonie Weinhold, Jens Wiltfang, Gill Windle, Bob Woods, Mary Yannakoulia, Habil Zare, Yi Zhao, Xiaoling Zhang, Congcong Zhu, Miren Zulaica, EADB, GR@ACE, DEGESCO, EADI, GERAD, Demgene, FinnGen, ADGC, CHARGE, Lindsay A. Farrer, Bruce M. Psaty, Mohsen Ghanbari, Towfique Raj, Perminder Sachdev, Karen Mather, Frank Jessen, M. Arfan Ikram, Alexandre de Mendonça, Jakub Hort, Magda Tsolaki, Margaret A. Pericak-Vance, Philippe Amouyel, Julie Williams, Ruth Frikke-Schmidt, Jordi Clarimon, Jean-François Deleuze, Giacomina Rossi, Sudha Seshadri, Ole A. Andreassen, Martin Ingelsson, Mikko Hiltunen, Kristel Sleegers, Gerard D. Schellenberg, Cornelia M. van Duijn, Rebecca Sims, Wiesje M. van der Flier, Agustín Ruiz, Alfredo Ramirez, and Jean-Charles Lambert. 2022. 'New Insights into the Genetic Etiology of Alzheimer's Disease and Related Dementias'. *Nature Genetics* 54(4):412–36. doi: 10.1038/s41588-022-01024-z.

Benjamini, Yuval, and Terence P. Speed. 2012. 'Summarizing and Correcting the GC Content Bias in High-Throughput Sequencing'. *Nucleic Acids Research* 40(10):e72. doi: 10.1093/nar/gks001.

von Bernhardi, Rommy, Francisca Cornejo, Guillermo E. Parada, and Jaime Eugenín. 2015. 'Role of TGFβ Signaling in the Pathogenesis of Alzheimer's Disease'. *Frontiers in Cellular Neuroscience* 9:426. doi: 10.3389/fncel.2015.00426.

Boot, Erik, Sólveig Óskarsdóttir, Joanne C. Y. Loo, Terrence Blaine Crowley, Ani Orchanian-Cheff, Danielle M. Andrade, Jill M. Arganbright, René M. Castelein, Christine Cserti-Gazdewich, Steven de Reuver, Ania M. Fiksinski, Gunilla Klingberg, Anthony E. Lang, Maria R. Mascarenhas, Edward M. Moss, Beata Anna Nowakowska, Erwin Oechslin, Lisa Palmer,

medRxiv preprint doi: https://doi.org/10.1101/2024.10.28.24314051; this version posted November 1, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Gabriela M. Repetto, Nikolai Gil D. Reyes, Maude Schneider, Candice Silversides, Kathleen E. Sullivan, Ann Swillen, Therese A. M. J. van Amelsvoort, Jason P. Van Batavia, Claudia Vingerhoets, Donna M. McDonald-McGinn, and Anne S. Bassett. 2023. 'Updated Clinical Practice Recommendations for Managing Adults with 22q11.2 Deletion Syndrome'. *Genetics in Medicine: Official Journal of the American College of Medical Genetics* 25(3):100344. doi: 10.1016/j.gim.2022.11.012.

Bouzid, Hind, Julia A. Belk, Max Jan, Yanyan Qi, Chloé Sarnowski, Sara Wirth, Lisa Ma, Matthew R. Chrostek, Herra Ahmad, Daniel Nachun, Winnie Yao, NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium, Alexa Beiser, Alexander G. Bick, Joshua C. Bis, Myriam Fornage, William T. Longstreth, Oscar L. Lopez, Pradeep Natarajan, Bruce M. Psaty, Claudia L. Satizabal, Joshua Weinstock, Eric B. Larson, Paul K. Crane, C. Dirk Keene, Sudha Seshadri, Ansuman T. Satpathy, Thomas J. Montine, and Siddhartha Jaiswal. 2023. 'Clonal Hematopoiesis Is Associated with Protection from Alzheimer's Disease'. *Nature Medicine*. doi: 10.1038/s41591-023-02397-2.

Brouwers, N., C. Van Cauwenberghe, S. Engelborghs, J. C. Lambert, K. Bettens, N. Le Bastard, F. Pasquier, A. Gil Montoya, K. Peeters, M. Mattheijssens, R. Vandenberghe, P. P. De Deyn, M. Cruts, P. Amouyel, K. Sleegers, and C. Van Broeckhoven. 2012. 'Alzheimer Risk Associated with a Copy Number Variation in the Complement Receptor 1 Increasing C3b/C4b Binding Sites'. *Molecular Psychiatry* 17(2):223–33. doi: 10.1038/mp.2011.24.

Butcher, Nancy J., Tim-Rasmus Kiehl, Lili-Naz Hazrati, Eva W. C. Chow, Ekaterina Rogaeva, Anthony E. Lang, and Anne S. Bassett. 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 Neurology* 70(11):1359. doi: 10.1001/jamaneurol.2013.3646.

Campion, D., C. Pottier, G. Nicolas, K. Le Guennec, and A. Rovelet-Lecrux. 2016. 'Alzheimer Disease: Modeling an Aβ-Centered Biological Network'. *Molecular Psychiatry* 21(7):861–71. doi: 10.1038/mp.2016.38.

Cassinari, Kévin, Olivier Quenez, Géraldine Joly-Hélas, Ludivine Beaussire, Nathalie Le Meur, Mathieu Castelain, Alice Goldenberg, Anne-Marie Guerrot, Anne-Claire Brehin, Jean-François Deleuze, Anne Boland, Anne Rovelet-Lecrux, Dominique Campion, Pascale Saugier-Veber, Nicolas Gruchy, Thierry Frebourg, Gaël Nicolas, Nasrin Sarafan-Vasseur, and Pascal Chambon. 2019. 'A Simple, Universal, and Cost-Efficient Digital PCR Method for the Targeted Analysis of Copy Number Variations'. *Clinical Chemistry*. doi: 10.1373/clinchem.2019.304246.

Charbonnier, F., G. Raux, Q. Wang, N. Drouot, F. Cordier, J. M. Limacher, J. C. Saurin, A. Puisieux, S. Olschwang, and T. Frebourg. 2000. 'Detection of Exon Deletions and Duplications of the Mismatch Repair Genes in Hereditary Nonpolyposis Colorectal Cancer Families Using Multiplex Polymerase Chain Reaction of Short Fluorescent Fragments'. *Cancer Research* 60(11):2760–63.

Chen, Siwei, Laurent C. Francioli, Julia K. Goodrich, Ryan L. Collins, Masahiro Kanai, Qingbo Wang, Jessica Alföldi, Nicholas A. Watts, Christopher Vittal, Laura D. Gauthier, Timothy Poterba, Michael W. Wilson, Yekaterina Tarasova, William Phu, Mary T. Yohannes, Zan Koenig, Yossi Farjoun, Eric Banks, Stacey Donnelly, Stacey Gabriel, Namrata Gupta, Steven Ferriera, Charlotte Tolonen, Sam Novod, Louis Bergelson, David Roazen, Valentin Ruano-Rubio, Miguel Covarrubias, Christopher Llanwarne, Nikelle Petrillo, Gordon Wade, Thibault Jeandet, Ruchi Munshi, Kathleen Tibbetts, gnomAD Project Consortium, Anne O'Donnell-Luria, Matthew Solomonson, Cotton Seed, Alicia R. Martin, Michael E. Talkowski, Heidi L. Rehm, Mark J. Daly,

medRxiv preprint doi: https://doi.org/10.1101/2024.10.28.24314051; this version posted November 1, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Grace Tiao, Benjamin M. Neale, Daniel G. MacArthur, and Konrad J. Karczewski. 2022. *A Genome-Wide Mutational Constraint Map Quantified from Variation in 76,156 Human Genomes*. *preprint*. Genetics. doi: 10.1101/2022.03.20.485034.

Crook, R., A. Verkkoniemi, J. Perez-Tur, N. Mehta, M. Baker, H. Houlden, M. Farrer, M. Hutton, S. Lincoln, J. Hardy, K. Gwinn, M. Somer, A. Paetau, H. Kalimo, R. Ylikoski, M. Pöyhönen, S. Kucera, and M. Haltia. 1998. 'A Variant of Alzheimer's Disease with Spastic Paraparesis and Unusual Plaques Due to Deletion of Exon 9 of Presenilin 1'. *Nature Medicine* 4(4):452–55. doi: 10.1038/nm0498-452.

Fortea, Juan, Eduard Vilaplana, Maria Carmona-Iragui, Bessy Benejam, Laura Videla, Isabel Barroeta, Susana Fernández, Miren Altuna, Jordi Pegueroles, Víctor Montal, Silvia Valldeneu, Sandra Giménez, Sofía González-Ortiz, Laia Muñoz, Teresa Estellés, Ignacio Illán-Gala, Olivia Belbin, Valle Camacho, Liam Reese Wilson, Tiina Annus, Ricardo S. Osorio, Sebastián Videla, Sylvain Lehmann, Anthony J. Holland, Daniel Alcolea, Jordi Clarimón, Shahid H. Zaman, Rafael Blesa, and Alberto Lleó. 2020. 'Clinical and Biomarker Changes of Alzheimer's Disease in Adults with Down Syndrome: A Cross-Sectional Study'. *Lancet (London, England)* 395(10242):1988–97. doi: 10.1016/S0140-6736(20)30689-9.

Frikke-Schmidt, Ruth, Børge G. Nordestgaard, Maria C. A. Stene, Amar A. Sethi, Alan T. Remaley, Peter Schnohr, Peer Grande, and Anne Tybjaerg-Hansen. 2008. 'Association of Loss-of-Function Mutations in the ABCA1 Gene with High-Density Lipoprotein Cholesterol Levels and Risk of Ischemic Heart Disease'. *JAMA* 299(21):2524–32. doi: 10.1001/jama.299.21.2524.

Gabrielaite, Migle, Mathias Husted Torp, Malthe Sebro Rasmussen, Sergio Andreu-Sánchez, Filipe Garrett Vieira, Christina Bligaard Pedersen, Savvas Kinalis, Majbritt Busk Madsen, Miyako Kodama, Gül Sude Demircan, Arman Simonyan, Christina Westmose Yde, Lars Rønn Olsen, Rasmus L. Marvig, Olga Østrup, Maria Rossing, Finn Cilius Nielsen, Ole Winther, and Frederik Otzen Bagger. 2021. 'A Comparison of Tools for Copy-Number Variation Detection in Germline Whole Exome and Whole Genome Sequencing Data'. *Cancers* 13(24):6283. doi: 10.3390/cancers13246283.

Gatz, Margaret, Chandra A. Reynolds, Laura Fratiglioni, Boo Johansson, James A. Mortimer, Stig Berg, Amy Fiske, and Nancy L. Pedersen. 2006. 'Role of Genes and Environments for Explaining Alzheimer Disease'. *Archives of General Psychiatry* 63(2):168–74. doi: 10.1001/archpsyc.63.2.168.

Genin, E., D. Hannequin, D. Wallon, K. Sleegers, M. Hiltunen, O. Combarros, M. J. Bullido, S. Engelborghs, P. De Deyn, C. Berr, F. Pasquier, B. Dubois, G. Tognoni, N. Fiévet, N. Brouwers, K. Bettens, B. Arosio, E. Coto, M. Del Zompo, I. Mateo, J. Epelbaum, A. Frank-Garcia, S. Helisalmi, E. Porcellini, A. Pilotto, P. Forti, R. Ferri, E. Scarpini, G. Siciliano, V. Solfrizzi, S. Sorbi, G. Spalletta, F. Valdivieso, S. Vepsäläinen, V. Alvarez, P. Bosco, M. Mancuso, F. Panza, B. Nacmias, P. Bossù, O. Hanon, P. Piccardi, G. Annoni, D. Seripa, D. Galimberti, F. Licastro, H. Soininen, J. F. Dartigues, M. I. Kamboh, C. Van Broeckhoven, J. C. Lambert, P. Amouyel, and D. Campion. 2011. 'APOE and Alzheimer Disease: A Major Gene with Semi-Dominant Inheritance'. *Molecular Psychiatry* 16(9):903–7. doi: 10.1038/mp.2011.52.

Grangeon, Lou, Kévin Cassinari, Stéphane Rousseau, Bernard Croisile, Maïté Formaglio, Olivier Moreaud, Jean Boutonnat, Nathalie Le Meur, Manuele Miné, Thibault Coste, Eva Pipiras, Elisabeth Tournier-Lasserve, Anne Rovelet-Lecrux, Dominique Campion, David Wallon, and Gael Nicolas. 2021. 'Early-Onset Cerebral Amyloid Angiopathy and Alzheimer Disease Related

medRxiv preprint doi: https://doi.org/10.1101/2024.10.28.24314051; this version posted November 1, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

to an APP Locus Triplication'. *Neurology. Genetics* 7(5):e609. doi: 10.1212/NXG.0000000000000609.

Grangeon, Lou, Camille Charbonnier, Aline Zarea, Stephane Rousseau, Anne Rovelet-Lecrux, David Bendetowicz, Marion Lemaitre, Cécile Malrain, Muriel Quillard-Muraine, Kevin Cassinari, David Maltete, Jeremie Pariente, Olivier Moreaud, Eloi Magnin, Benjamin Cretin, Marie-Anne Mackowiak, Adeline Rollin Sillaire, Martine Vercelletto, Elsa Dionet, Olivier Felician, Pauline Rod-Olivieri, Catherine Thomas-Antérion, Gaelle Godeneche, Mathilde Sauvée, Leslie Cartz-Piver, Isabelle Le Ber, Valérie Chauvire, Therèse Jonveaux, Anna-Chloé Balageas, Annie Laquerriere, Charles Duyckaerts, Anne Vital, Andre Maues de Paula, David Meyronet, Lucie Guyant-Marechal, Didier Hannequin, Elisabeth Tournier-Lasserve, Dominique Campion, CNR-MAJ collaborators, Gaël Nicolas, and David Wallon. 2023. 'Phenotype and Imaging Features Associated with APP Duplications'. *Alzheimer's Research & Therapy* 15(1):93. doi: 10.1186/s13195-023-01172-2.

Halldorsson, Bjarni V., Hannes P. Eggertsson, Kristjan H. S. Moore, Hannes Hauswedell, Ogmundur Eiriksson, Magnus O. Ulfarsson, Gunnar Palsson, Marteinn T. Hardarson, Asmundur Oddsson, Brynjar O. Jensson, Snaedis Kristmundsdottir, Brynja D. Sigurpalsdottir, Olafur A. Stefansson, Doruk Beyter, Guillaume Holley, Vinicius Tragante, Arnaldur Gylfason, Pall I. Olason, Florian Zink, Margret Asgeirsdottir, Sverrir T. Sverrisson, Brynjar Sigurdsson, Sigurjon A. Gudjonsson, Gunnar T. Sigurdsson, Gisli H. Halldorsson, Gardar Sveinbjornsson, Kristjan Norland, Unnur Styrkarsdottir, Droplaug N. Magnusdottir, Steinunn Snorradottir, Kari Kristinsson, Emilia Sobech, Helgi Jonsson, Arni J. Geirsson, Isleifur Olafsson, Palmi Jonsson, Ole Birger Pedersen, Christian Erikstrup, Søren Brunak, Sisse Rye Ostrowski, DBDS Genetic Consortium, Steffen Andersen, Karina Banasik, Kristoffer Burgdorf, Maria Didriksen, Khoa Manh Dinh, Christian Erikstrup, Daniel Gudbjartsson, Thomas Folkmann Hansen, Henrik Hjalgrim, Gregor Jemec, Poul Jennum, Pär Ingemar Johansson, Margit Anita Hørup Larsen, Susan Mikkelsen, Kasper Rene Nielsen, Mette Nyegaard, Sisse Rye Ostrowski, Susanne Sækmose, Erik Sørensen, Unnur Thorsteinsdottir, Mie Topholm Brun, Henrik Ullum, Thomas Werge, Gudmar Thorleifsson, Frosti Jonsson, Pall Melsted, Ingileif Jonsdottir, Thorunn Rafnar, Hilma Holm, Hreinn Stefansson, Jona Saemundsdottir, Daniel F. Gudbjartsson, Olafur T. Magnusson, Gisli Masson, Unnur Thorsteinsdottir, Agnar Helgason, Hakon Jonsson, Patrick Sulem, and Kari Stefansson. 2022. 'The Sequences of 150,119 Genomes in the UK Biobank'. *Nature* 607(7920):732–40. doi: 10.1038/s41586-022-04965-x.

Hehir-Kwa, Jayne Y., Rolph Pfundt, and Joris A. Veltman. 2015. 'Exome Sequencing and Whole Genome Sequencing for the Detection of Copy Number Variation'. *Expert Review of Molecular Diagnostics* 15(8):1023–32. doi: 10.1586/14737159.2015.1053467.

Holstege, Henne, Marc Hulsman, Camille Charbonnier, Benjamin Grenier-Boley, Olivier Quenez, Detelina Grozeva, Jeroen G. J. van Rooij, Rebecca Sims, Shahzad Ahmad, Najaf Amin, Penny J. Norsworthy, Oriol Dols-Icardo, Holger Hummerich, Amit Kawalia, Philippe Amouyel, Gary W. Beecham, Claudine Berr, Joshua C. Bis, Anne Boland, Paola Bossù, Femke Bouwman, Jose Bras, Dominique Campion, J. Nicholas Cochran, Antonio Daniele, Jean-François Dartigues, Stéphanie Debette, Jean-François Deleuze, Nicola Denning, Anita L. DeStefano, Lindsay A. Farrer, Maria Victoria Fernández, Nick C. Fox, Daniela Galimberti, Emmanuelle Genin, Johan J. P. Gille, Yann Le Guen, Rita Guerreiro, Jonathan L. Haines, Clive Holmes, M. Arfan Ikram, M. Kamran Ikram, Iris E. Jansen, Robert Kraaij, Marc Lathrop, Afina W. Lemstra, Alberto Lleó, Lauren Luckcuck, Marcel M. A. M. Mannens, Rachel Marshall, Eden R. Martin, Carlo Masullo, Richard Mayeux, Patrizia Mecocci, Alun Meggy, Merel O. Mol, Kevin Morgan, Richard M. Myers, Benedetta Nacmias, Adam C. Naj, Valerio Napolioni, Florence Pasquier, Pau Pastor, Margaret A. Pericak-Vance, Rachel Raybould, Richard Redon, Marcel J. T. Reinders, Anne-

medRxiv preprint doi: https://doi.org/10.1101/2024.10.28.24314051; this version posted November 1, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Claire Richard, Steffi G. Riedel-Heller, Fernando Rivadeneira, Stéphane Rousseau, Natalie S. Ryan, Salha Saad, Pascual Sanchez-Juan, Gerard D. Schellenberg, Philip Scheltens, Jonathan M. Schott, Davide Seripa, Sudha Seshadri, Daoud Sie, Erik A. Sistermans, Sandro Sorbi, Resie van Spaendonk, Gianfranco Spalletta, Niccolo' Tesi, Betty Tijms, André G. Uitterlinden, Sven J. van der Lee, Pieter Jelle Visser, Michael Wagner, David Wallon, Li-San Wang, Aline Zarea, Jordi Clarimon, John C. van Swieten, Michael D. Greicius, Jennifer S. Yokoyama, Carlos Cruchaga, John Hardy, Alfredo Ramirez, Simon Mead, Wiesje M. van der Flier, Cornelia M. van Duijn, Julie Williams, Gaël Nicolas, Céline Bellenguez, and Jean-Charles Lambert. 2022. 'Exome Sequencing Identifies Rare Damaging Variants in ATP8B4 and ABCA1 as Risk Factors for Alzheimer's Disease'. *Nature Genetics* 54(12):1786–94. doi: 10.1038/s41588-022-01208-7.

Hooli, B. V., Z. M. Kovacs-Vajna, K. Mullin, M. A. Blumenthal, M. Mattheisen, C. Zhang, C. Lange, G. Mohapatra, L. Bertram, and R. E. Tanzi. 2014. 'Rare Autosomal Copy Number Variations in Early-Onset Familial Alzheimer's Disease'. *Molecular Psychiatry* 19(6):676–81. doi: 10.1038/mp.2013.77.

Huguet, Guillaume, Catherine Schramm, Elise Douard, Lai Jiang, Aurélie Labbe, Frédérique Tihy, Géraldine Mathonnet, Sonia Nizard, Emmanuelle Lemyre, Alexandre Mathieu, Jean-Baptiste Poline, Eva Loth, Roberto Toro, Gunter Schumann, Patricia Conrod, Zdenka Pausova, Celia Greenwood, Tomas Paus, Thomas Bourgeron, Sébastien Jacquemont, and IMAGEN Consortium. 2018. 'Measuring and Estimating the Effect Sizes of Copy Number Variants on General Intelligence in Community-Based Samples'. *JAMA Psychiatry* 75(5):447–57. doi: 10.1001/jamapsychiatry.2018.0039.

Iulita, Maria Florencia, Diana Garzón Chavez, Maria Klitgaard Christensen, Natalia Valle Tamayo, Oleguer Plana-Ripoll, Sonja A. Rasmussen, Marta Roqué Figuls, Daniel Alcolea, Laura Videla, Isabel Barroeta, Bessy Benejam, Miren Altuna, Concepción Padilla, Jordi Pegueroles, Susana Fernandez, Olivia Belbin, María Carmona-Iragui, Rafael Blesa, Alberto Lleó, Alexandre Bejanin, and Juan Fortea. 2022. 'Association of Alzheimer Disease With Life Expectancy in People With Down Syndrome'. *JAMA Network Open* 5(5):e2212910. doi: 10.1001/jamanetworkopen.2022.12910.

Jonsson, Thorlakur, Hreinn Stefansson, Stacy Steinberg, Ingileif Jonsdottir, Palmi V. Jonsson, Jon Snaedal, Sigurbjorn Bjornsson, Johanna Huttenlocher, Allan I. Levey, James J. Lah, Dan Rujescu, Harald Hampel, Ina Giegling, Ole A. Andreassen, Knut Engedal, Ingun Ulstein, Srdjan Djurovic, Carla Ibrahim-Verbaas, Albert Hofman, M. Arfan Ikram, Cornelia M. van Duijn, Unnur Thorsteinsdottir, Augustine Kong, and Kari Stefansson. 2013. 'Variant of TREM2 Associated with the Risk of Alzheimer's Disease'. *The New England Journal of Medicine* 368(2):107–16. doi: 10.1056/NEJMoa1211103.

Karczewski, Konrad J., Laurent C. Francioli, Grace Tiao, Beryl B. Cummings, Jessica Alföldi, Qingbo Wang, Ryan L. Collins, Kristen M. Laricchia, Andrea Ganna, Daniel P. Birnbaum, Laura D. Gauthier, Harrison Brand, Matthew Solomonson, Nicholas A. Watts, Daniel Rhodes, Moriel Singer-Berk, Eleanor G. Seaby, Jack A. Kosmicki, Raymond K. Walters, Katherine Tashman, Yossi Farjoun, Eric Banks, Timothy Poterba, Arcturus Wang, Cotton Seed, Nicola Whiffin, Jessica X. Chong, Kaitlin E. Samocha, Emma Pierce-Hoffman, Zachary Zappala, Anne H. O'Donnell-Luria, Eric Vallabh Minikel, Ben Weisburd, Monkol Lek, James S. Ware, Christopher Vittal, Irina M. Armean, Louis Bergelson, Kristian Cibulskis, Kristen M. Connolly, Miguel Covarrubias, Stacey Donnelly, Steven Ferriera, Stacey Gabriel, Jeff Gentry, Namrata Gupta, Thibault Jeandet, Diane Kaplan, Christopher Llanwarne, Ruchi Munshi, Sam Novod, Nikelle Petrillo, David Roazen, Valentin Ruano-Rubio, Andrea Saltzman, Molly Schleicher, Jose Soto,

medRxiv preprint doi: https://doi.org/10.1101/2024.10.28.24314051; this version posted November 1, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Kathleen Tibbetts, Charlotte Tolonen, Gordon Wade, Michael E. Talkowski, The Genome Aggregation Database Consortium, Benjamin M. Neale, Mark J. Daly, and Daniel G. MacArthur. 2019. *Variation across 141,456 Human Exomes and Genomes Reveals the Spectrum of Loss-of-Function Intolerance across Human Protein-Coding Genes*. *preprint*. Genomics. doi: 10.1101/531210.

Karlstrom, Helena, William S. Brooks, John B. J. Kwok, G. Anthony Broe, Jillian J. Kril, Heather McCann, Glenda M. Halliday, and Peter R. Schofield. 2008. 'Variable Phenotype of Alzheimer's Disease with Spastic Paraparesis'. *Journal of Neurochemistry* 104(3):573–83. doi: 10.1111/j.1471-4159.2007.05038.x.

Lanoiselée, Hélène-Marie, Gaël Nicolas, David Wallon, Anne Rovelet-Lecrux, Morgane Lacour, Stéphane Rousseau, Anne-Claire Richard, Florence Pasquier, Adeline Rollin-Sillaire, Olivier Martinaud, Muriel Quillard-Muraine, Vincent de la Sayette, Claire Boutoleau-Bretonniere, Frédérique Etcharry-Bouyx, Valérie Chauviré, Marie Sarazin, Isabelle le Ber, Stéphane Epelbaum, Thérèse Jonveaux, Olivier Rouaud, Mathieu Ceccaldi, Olivier Félician, Olivier Godefroy, Maite Formaglio, Bernard Croisile, Sophie Auriacombe, Ludivine Chamard, Jean-Louis Vincent, Mathilde Sauvée, Cecilia Marelli-Tosi, Audrey Gabelle, Canan Ozsancak, Jérémie Pariente, Claire Paquet, Didier Hannequin, Dominique Campion, and collaborators of the CNR-MAJ project. 2017. 'APP, PSEN1, and PSEN2 Mutations in Early-Onset Alzheimer Disease: A Genetic Screening Study of Familial and Sporadic Cases'. *PLoS Medicine* 14(3):e1002270. doi: 10.1371/journal.pmed.1002270.

Le Guennec, K, O. Quenez, G. Nicolas, D. Wallon, S. Rousseau, A. C. Richard, J. Alexander, P. Paschou, C. Charbonnier, C. Bellenguez, B. Grenier-Boley, D. Lechner, M. T. Bihoreau, R. Olaso, A. Boland, V. Meyer, J. F. Deleuze, P. Amouyel, H. M. Munter, G. Bourque, M. Lathrop, T. Frebourg, R. Redon, L. Letenneur, J. F. Dartigues, O. Martinaud, O. Kalev, S. Mehrabian, L. Traykov, T. Ströbel, I. Le Ber, P. Caroppo, S. Epelbaum, T. Jonveaux, F. Pasquier, A. Rollin-Sillaire, E. Génin, L. Guyant-Maréchal, G. G. Kovacs, J. C. Lambert, D. Hannequin, D. Campion, and A. Rovelet-Lecrux. 2017. '17q21.31 Duplication Causes Prominent Tau-Related Dementia with Increased MAPT Expression'. *Molecular Psychiatry* 22(8):1119–25. doi: 10.1038/mp.2016.226.

Le Guennec, Kilan, Gaël Nicolas, Olivier Quenez, Camille Charbonnier, David Wallon, Céline Bellenguez, Benjamin Grenier-Boley, Stéphane Rousseau, Anne-Claire Richard, Anne Rovelet-Lecrux, Delphine Bacq, Jean-Guillaume Garnier, Robert Olaso, Anne Boland, Vincent Meyer, Jean-François Deleuze, Philippe Amouyel, Hans Markus Munter, Guillaume Bourque, Mark Lathrop, Thierry Frebourg, Richard Redon, Luc Letenneur, Jean-François Dartigues, Florence Pasquier, Adeline Rollin-Sillaire, Emmanuelle Génin, Jean-Charles Lambert, Didier Hannequin, Dominique Campion, and CNR-MAJ collaborators. 2016. 'ABCA7 Rare Variants and Alzheimer Disease Risk'. *Neurology* 86(23):2134–37. doi: 10.1212/WNL.0000000000002627.

Le Guennec, Kilan, Sarah Veugelen, Olivier Quenez, Maria Szaruga, Stéphane Rousseau, Gaël Nicolas, David Wallon, Frédérique Fluchere, Thierry Frébourg, Bart De Strooper, Dominique Campion, Lucía Chávez-Gutiérrez, and Anne Rovelet-Lecrux. 2017. '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* 104:97–103. doi: 10.1016/j.nbd.2017.04.020.

Lee, Joseph H., Rong Cheng, Neill Graff-Radford, Tatiana Foroud, Richard Mayeux, 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'. *Archives of Neurology* 65(11):1518–26. doi: 10.1001/archneur.65.11.1518.

medRxiv preprint doi: https://doi.org/10.1101/2024.10.28.24314051; this version posted November 1, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Lee, Wan-Ping, Albert A. Tucci, Mitchell Conery, Yuk Yee Leung, Amanda B. Kuzma, Otto Valladares, Yi-Fan Chou, Wenbin Lu, Li-San Wang, Gerard D. Schellenberg, and Jung-Ying Tzeng. 2021. 'Copy Number Variation Identification on 3,800 Alzheimer's Disease Whole Genome Sequencing Data from the Alzheimer's Disease Sequencing Project'. *Frontiers in Genetics* 12:752390. doi: 10.3389/fgene.2021.752390.

Logsdon, Glennis A., Mitchell R. Vollger, and Evan E. Eichler. 2020. 'Long-Read Human Genome Sequencing and Its Applications'. *Nature Reviews. Genetics* 21(10):597–614. doi: 10.1038/s41576-020-0236-x.

Lupski, J. R., R. M. de Oca-Luna, S. Slaugenhaupt, L. Pentao, V. Guzzetta, B. J. Trask, O. Saucedo-Cardenas, D. F. Barker, J. M. Killian, C. A. Garcia, A. Chakravarti, and P. I. Patel. 1991. 'DNA Duplication Associated with Charcot-Marie-Tooth Disease Type 1A'. *Cell* 66(2):219–32. doi: 10.1016/0092-8674(91)90613-4.

MacDonald, Jeffrey R., Robert Ziman, Ryan K. C. Yuen, Lars Feuk, and Stephen W. Scherer. 2014. 'The Database of Genomic Variants: A Curated Collection of Structural Variation in the Human Genome'. *Nucleic Acids Research* 42(Database issue):D986-992. doi: 10.1093/nar/gkt958.

Machiela, Mitchell J., Weiyin Zhou, Neil Caporaso, Michael Dean, Susan M. Gapstur, Lynn Goldin, Nathaniel Rothman, Victoria L. Stevens, Meredith Yeager, and Stephen J. Chanock. 2017. 'Mosaic Chromosome 20q Deletions Are More Frequent in the Aging Population'. *Blood Advances* 1(6):380–85. doi: 10.1182/bloodadvances.2016003129.

Martiskainen, Henna, Roosa-Maria Willman, Sami Heikkinen, Stephan A. Müller, Rosa Sinisalo, Mari Takalo, Petra Mäkinen, Teemu Kuulasmaa, Viivi Pekkala, Ana Galván Del Rey, Päivi Harju, Sini-Pauliina Juopperi, Heli Jeskanen, Inka Kervinen, Kirsi Saastamoinen, FinnGen, Marja Niiranen, Sami V. Heikkinen, Mitja I. Kurki, Jarkko Marttila, Petri I. Mäkinen, Tiia Ngandu, Jenni Lehtisalo, Céline Bellenguez, Jean-Charles Lambert, Christian Haass, Juha Rinne, Juhana Hakumäki, Tuomas Rauramaa, Johanna Krüger, Hilkka Soininen, Annakaisa Haapasalo, Stefan F. Lichtenthaler, Ville Leinonen, Eino Solje, and Mikko Hiltunen. 2024. 'Monoallelic *TYROBP* Deletion Is a Novel Risk Factor for Alzheimer's Disease'.

McKhann, G., D. Drachman, M. Folstein, R. Katzman, D. Price, and E. M. Stadlan. 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(7):939–44. doi: 10.1212/wnl.34.7.939.

McKhann, Guy M., David S. Knopman, Howard Chertkow, Bradley T. Hyman, Clifford R. Jack, Claudia H. Kawas, William E. Klunk, Walter J. Koroshetz, Jennifer J. Manly, Richard Mayeux, Richard C. Mohs, John C. Morris, Martin N. Rossor, Philip Scheltens, Maria C. Carrillo, Bill Thies, Sandra Weintraub, and Creighton H. Phelps. 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 & Dementia: The Journal of the Alzheimer's Association* 7(3):263–69. doi: 10.1016/j.jalz.2011.03.005.

Miller, D. W., S. M. Hague, J. Clarimon, M. Baptista, K. Gwinn-Hardy, M. R. Cookson, and A. B. Singleton. 2004. 'Alpha-Synuclein in Blood and Brain from Familial Parkinson Disease with SNCA Locus Triplication'. *Neurology* 62(10):1835–38. doi: 10.1212/01.wnl.0000127517.33208.f4.

Miller, Danny E., Arvis Sulovari, Tianyun Wang, Hailey Loucks, Kendra Hoekzema, Katherine M. Munson, Alexandra P. Lewis, Edith P. Almanza Fuerte, Catherine R. Paschal, Tom Walsh,

medRxiv preprint doi: https://doi.org/10.1101/2024.10.28.24314051; this version posted November 1, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Jenny Thies, James T. Bennett, Ian Glass, Katrina M. Dipple, Karynne Patterson, Emily S. Bonkowski, Zoe Nelson, Audrey Squire, Megan Sikes, Erika Beckman, Robin L. Bennett, Dawn Earl, Winston Lee, Rando Allikmets, Seth J. Perlman, Penny Chow, Anne V. Hing, Tara L. Wenger, Margaret P. Adam, Angela Sun, Christina Lam, Irene Chang, Xue Zou, Stephanie L. Austin, Erin Huggins, Alexias Safi, Apoorva K. Iyengar, Timothy E. Reddy, William H. Majoros, Andrew S. Allen, Gregory E. Crawford, Priya S. Kishnani, University of Washington Center for Mendelian Genomics, Mary-Claire King, Tim Cherry, Jessica X. Chong, Michael J. Bamshad, Deborah A. Nickerson, Heather C. Mefford, Dan Doherty, and Evan E. Eichler. 2021. 'Targeted Long-Read Sequencing Identifies Missing Disease-Causing Variation'. *American Journal of Human Genetics* 108(8):1436–49. doi: 10.1016/j.ajhg.2021.06.006.

Mok, Kin Y., Una Sheerin, Javier Simón-Sánchez, Afnan Salaka, Lucy Chester, Valentina Escott-Price, Kiran Mantripragada, Karen M. Doherty, Alastair J. Noyce, Niccolo E. Mencacci, Steven J. Lubbe, Caroline H. Williams-Gray, Roger A. Barker, Karin D. Van Dijk, Henk W. Berendse, Peter Heutink, Jean-Christophe Corvol, Florence Cormier, Suzanne Lesage, Alexis Brice, Kathrin Brockmann, Claudia Schulte, Thomas Gasser, Thomas Foltynie, Patricia Limousin, Karen E. Morrison, Carl E. Clarke, Stephen Sawcer, Tom T. Warner, Andrew J. Lees, Huw R. Morris, Mike A. Nalls, Andrew B. Singleton, John Hardy, Andrey Y. Abramov, Vincent Plagnol, Nigel M. Williams, and Nicholas W. Wood. 2016. 'Deletions at 22q11.2 in Idiopathic Parkinson's Disease: A Combined Analysis of Genome-Wide Association Data'. *The Lancet Neurology* 15(6):585–96. doi: 10.1016/S1474-4422(16)00071-5.

Mueller-Steiner, Sarah, Yungui Zhou, Hideaki Arai, Erik D. Roberson, Binggui Sun, Jennifer Chen, Xin Wang, Guiqiu Yu, Luke Esposito, Lennart Mucke, and Li Gan. 2006. 'Antiamyloidogenic and Neuroprotective Functions of Cathepsin B: Implications for Alzheimer's Disease'. *Neuron* 51(6):703–14. doi: 10.1016/j.neuron.2006.07.027.

Nelson, Monica E., Dylan J. Jester, Andrew J. Petkus, and Ross Andel. 2021. 'Cognitive Reserve, Alzheimer's Neuropathology, and Risk of Dementia: A Systematic Review and Meta-Analysis'. *Neuropsychology Review* 31(2):233–50. doi: 10.1007/s11065-021-09478-4.

Nicolas, G., C. Charbonnier, D. Wallon, O. Quenez, C. Bellenguez, B. Grenier-Boley, S. Rousseau, A. C. Richard, A. Rovelet-Lecrux, K. Le Guennec, D. Bacq, J. G. Garnier, R. Olaso, A. Boland, V. Meyer, J. F. Deleuze, P. Amouyel, H. M. Munter, G. Bourque, M. Lathrop, T. Frebourg, R. Redon, L. Letenneur, J. F. Dartigues, E. Génin, J. C. Lambert, D. Hannequin, D. Campion, and CNR-MAJ collaborators. 2016. 'SORL1 Rare Variants: A Major Risk Factor for Familial Early-Onset Alzheimer's Disease'. *Molecular Psychiatry* 21(6):831–36. doi: 10.1038/mp.2015.121.

Nicolas, Gaël. 2024. 'Recent Advances in Alzheimer Disease Genetics'. *Current Opinion in Neurology* 37(2):154–65. doi: 10.1097/WCO.0000000000001242.

Nicolas, Gaël, Aline Zaréa, Morgane Lacour, Olivier Quenez, Stéphane Rousseau, Anne-Claire Richard, Antoine Bonnevalle, Catherine Schramm, Robert Olaso, Florian Sandron, Anne Boland, Jean-François Deleuze, Daniela Andriuta, Pierre Anthony, Sophie Auriacombe, Anna-Chloé Balageas, Guillaume Ballan, Mélanie Barbay, Yannick Béjot, Serge Belliard, Marie Benaiteau, Karim Bennys, Stéphanie Bombois, Claire Boutoleau-Bretonnière, Pierre Branger, Jasmine Carlier, Leslie Cartz-Piver, Pascaline Cassagnaud, Mathieu-Pierre Ceccaldi, Valérie Chauviré, Yaohua Chen, Julien Cogez, Emmanuel Cognat, Fabienne Contegal-Callier, Léa Corneille, Philippe Couratier, Benjamin Cretin, Charlotte Crinquette, Benjamin Dauriat, Sophie Dautricourt, Vincent de la Sayette, Astrid de Liège, Didier Deffond, Florence Demurger, Vincent Deramecourt, Céline Derollez, Elsa Dionet, Martine Doco Fenzy, Julien Dumurgier, Anaïs Dutray, Frédérique Etcharry-Bouyx, Maïté Formaglio, Audrey Gabelle, Anne Gainche-

medRxiv preprint doi: https://doi.org/10.1101/2024.10.28.24314051; this version posted November 1, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Salmon, Olivier Godefroy, Mathilde Graber, Chloé Gregoire, Stephan Grimaldi, Julien Gueniat, Claude Gueriot, Virginie Guillet-Pichon, Sophie Haffen, Cezara-Roxana Hanta, Clémence Hardy, Geoffroy Hautecloque, Camille Heitz, Claire Hourregue, Thérèse Jonveaux, Snejana Jurici, Lejla Koric, Pierre Krolak-Salmon, Julien Lagarde, Hélène-Marie Lanoiselée, Brice Laurens, Isabelle Le Ber, Gwenaël Le Guyader, Amélie Leblanc, Thibaud Lebouvier, Richard Levy, Anaïs Lippi, Marie-Anne Mackowiak, Eloi Magnin, Cecilia Marelli, Olivier Martinaud, Aurelien Maureille, Raffaella Migliaccio, Emilie Milongo-Rigal, Sophie Mohr, Hélène Mollion, Alexandre Morin, Julia Nivelle, Camille Noiray, Pauline Olivieri, Claire Paquet, Jérémie Pariente, Florence Pasquier, Alexandre Perron, Nathalie Philippi, Vincent Planche, Hélène Pouclet-Courtemanche, Marie Rafiq, Adeline Rollin-Sillaire, Carole Roué-Jagot, Dario Saracino, Marie Sarazin, Mathilde Sauvée, François Sellal, Marc Teichmann, Christel Thauvin, Quentin Thomas, Camille Tisserand, Cédric Turpinat, Laurène Van Damme, Olivier Vercruysse, Nicolas Villain, Nathalie Wagemann, Camille Charbonnier, and David Wallon. 2024. '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* 26(5):101082. doi: 10.1016/j.gim.2024.101082.

Nordestgaard, Liv Tybjærg, Anne Tybjærg-Hansen, Børge G. Nordestgaard, and Ruth Frikke-Schmidt. 2015. 'Loss-of-Function Mutation in ABCA1 and Risk of Alzheimer's Disease and Cerebrovascular Disease'. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association* 11(12):1430–38. doi: 10.1016/j.jalz.2015.04.006.

Quenez, Olivier, Kevin Cassinari, Sophie Coutant, François Lecoquierre, Kilan Le Guennec, Stéphane Rousseau, Anne-Claire Richard, Stéphanie Vasseur, Emilie Bouvignies, Jacqueline Bou, Gwendoline Lienard, Sandrine Manase, Steeve Fourneaux, Nathalie Drouot, Virginie Nguyen-Viet, Myriam Vezain, Pascal Chambon, Géraldine Joly-Helas, Nathalie Le Meur, Mathieu Castelain, Anne Boland, Jean-François Deleuze, FREX Consortium, Isabelle Tournier, Françoise Charbonnier, Edwige Kasper, Gaëlle Bougeard, Thierry Frebourg, Pascale Saugier-Veber, Stéphanie Baert-Desurmont, Dominique Campion, Anne Rovelet-Lecrux, and Gaël Nicolas. 2021. 'Detection of Copy-Number Variations from NGS Data Using Read Depth Information: A Diagnostic Performance Evaluation'. *European Journal of Human Genetics: EJHG* 29(1):99–109. doi: 10.1038/s41431-020-0672-2.

Quinlan, Aaron R., and Ira M. Hall. 2010. 'BEDTools: A Flexible Suite of Utilities for Comparing Genomic Features'. *Bioinformatics (Oxford, England)* 26(6):841–42. doi: 10.1093/bioinformatics/btq033.

Rovelet-Lecrux, A., C. Charbonnier, D. Wallon, G. Nicolas, M. N. J. Seaman, C. Pottier, S. Y. Breusegem, P. P. Mathur, P. Jenardhanan, K. Le Guennec, A. S. Mukadam, O. Quenez, S. Coutant, S. Rousseau, A. C. Richard, A. Boland, J. F. Deleuze, T. Frebourg, D. Hannequin, D. Campion, and CNR-MAJ collaborators. 2015. 'De Novo Deleterious Genetic Variations Target a Biological Network Centered on Aβ Peptide in Early-Onset Alzheimer Disease'. *Molecular Psychiatry* 20(9):1046–56. doi: 10.1038/mp.2015.100.

Rovelet-Lecrux, Anne, Vincent Deramecourt, Solenn Legallic, Claude-Alain Maurage, Isabelle Le Ber, Alexis Brice, Jean-Charles Lambert, Thierry Frébourg, Didier Hannequin, Florence Pasquier, and Dominique Campion. 2008. 'Deletion of the Progranulin Gene in Patients with Frontotemporal Lobar Degeneration or Parkinson Disease'. *Neurobiology of Disease* 31(1):41–45. doi: 10.1016/j.nbd.2008.03.004.

medRxiv preprint doi: https://doi.org/10.1101/2024.10.28.24314051; this version posted November 1, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Rovelet-Lecrux, Anne, Didier Hannequin, Gregory Raux, Nathalie Le Meur, Annie Laquerrière, Anne Vital, Cécile Dumanchin, Sébastien Feuillette, Alexis Brice, Martine Vercelletto, Frédéric Dubas, Thierry Frebourg, and Dominique Campion. 2006. 'APP Locus Duplication Causes Autosomal Dominant Early-Onset Alzheimer Disease with Cerebral Amyloid Angiopathy'. *Nature Genetics* 38(1):24–26. doi: 10.1038/ng1718.

Rovelet-Lecrux, Anne, Solenn Legallic, David Wallon, Jean-Michel Flaman, Olivier Martinaud, Stéphanie Bombois, Adeline Rollin-Sillaire, Agnès Michon, Isabelle Le Ber, Jérémie Pariente, Michèle Puel, Claire Paquet, Bernard Croisile, Catherine Thomas-Antérion, Martine Vercelletto, Richard Lévy, Thierry Frébourg, Didier Hannequin, Dominique Campion, and Investigators of the GMAJ project. 2012. 'A Genome-Wide Study Reveals Rare CNVs Exclusive to Extreme Phenotypes of Alzheimer Disease'. *European Journal of Human Genetics: EJHG* 20(6):613–17. doi: 10.1038/ejhg.2011.225.

Rump, Patrick, Nicole de Leeuw, Anthonie J. van Essen, Corien C. Verschuuren-Bemelmans, Hermine E. Veenstra-Knol, Mariëlle E. M. Swinkels, Wilma Oostdijk, Claudia Ruivenkamp, Willie Reardon, Sonja de Munnik, Mariken Ruiter, Ayala Frumkin, Dorit Lev, Christina Evers, Birgit Sikkema-Raddatz, Trijnie Dijkhuizen, and Conny M. van Ravenswaaij-Arts. 2014. 'Central 22q11.2 Deletions'. *American Journal of Medical Genetics. Part A* 164A(11):2707–23. doi: 10.1002/ajmg.a.36711.

Schramm, Catherine, Camille Charbonnier, Aline Zaréa, Morgane Lacour, David Wallon, CNRMAJ collaborators, Anne Boland, Jean-François Deleuze, Robert Olaso, ADES consortium, Flora Alarcon, Dominique Campion, Grégory Nuel, and Gaël Nicolas. 2022. 'Penetrance Estimation of Alzheimer Disease in SORL1 Loss-of-Function Variant Carriers Using a Family-Based Strategy and Stratification by APOE Genotypes'. *Genome Medicine* 14(1):69. doi: 10.1186/s13073-022-01070-6.

Shen, Yiping, Kira A. Dies, Ingrid A. Holm, Carolyn Bridgemohan, Magdi M. Sobeih, Elizabeth B. Caronna, Karen J. Miller, Jean A. Frazier, Iris Silverstein, Jonathan Picker, Laura Weissman, Peter Raffalli, Shafali Jeste, Laurie A. Demmer, Heather K. Peters, Stephanie J. Brewster, Sara J. Kowalczyk, Beth Rosen-Sheidley, Caroline McGowan, Andrew W. Duda, Sharyn A. Lincoln, Kathryn R. Lowe, Alison Schonwald, Michael Robbins, Fuki Hisama, Robert Wolff, Ronald Becker, Ramzi Nasir, David K. Urion, Jeff M. Milunsky, Leonard Rappaport, James F. Gusella, Christopher A. Walsh, Bai-Lin Wu, David T. Miller, and Autism Consortium Clinical Genetics/DNA Diagnostics Collaboration. 2010. 'Clinical Genetic Testing for Patients with Autism Spectrum Disorders'. *Pediatrics* 125(4):e727-735. doi: 10.1542/peds.2009-1684.

Sims, Rebecca, Sven J. van der Lee, Adam C. Naj, Céline Bellenguez, Nandini Badarinarayan, Johanna Jakobsdottir, Brian W. Kunkle, Anne Boland, Rachel Raybould, Joshua C. Bis, Eden R. Martin, Benjamin Grenier-Boley, Stefanie Heilmann-Heimbach, Vincent Chouraki, Amanda B. Kuzma, Kristel Sleegers, Maria Vronskaya, Agustin Ruiz, Robert R. Graham, Robert Olaso, Per Hoffmann, Megan L. Grove, Badri N. Vardarajan, Mikko Hiltunen, Markus M. Nöthen, Charles C. White, Kara L. Hamilton-Nelson, Jacques Epelbaum, Wolfgang Maier, Seung-Hoan Choi, Gary W. Beecham, Cécile Dulary, Stefan Herms, Albert V. Smith, Cory C. Funk, Céline Derbois, Andreas J. Forstner, Shahzad Ahmad, Hongdong Li, Delphine Bacq, Denise Harold, Claudia L. Satizabal, Otto Valladares, Alessio Squassina, Rhodri Thomas, Jennifer A. Brody, Liming Qu, Pascual Sánchez-Juan, Taniesha Morgan, Frank J. Wolters, Yi Zhao, Florentino Sanchez Garcia, Nicola Denning, Myriam Fornage, John Malamon, Maria Candida Deniz Naranjo, Elisa Majounie, Thomas H. Mosley, Beth Dombroski, David Wallon, Michelle K. Lupton, Josée Dupuis, Patrice Whitehead, Laura Fratiglioni, Christopher Medway, Xueqiu Jian, Shubhabrata Mukherjee, Lina Keller, Kristelle Brown, Honghuang Lin, Laura B. Cantwell, Francesco Panza,

medRxiv preprint doi: https://doi.org/10.1101/2024.10.28.24314051; this version posted November 1, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Bernadette McGuinness, Sonia Moreno-Grau, Jeremy D. Burgess, Vincenzo Solfrizzi, Petra Proitsi, Hieab H. Adams, Mariet Allen, Davide Seripa, Pau Pastor, L. Adrienne Cupples, Nathan D. Price, Didier Hannequin, Ana Frank-García, Daniel Levy, Paramita Chakrabarty, Paolo Caffarra, Ina Giegling, Alexa S. Beiser, Vilmantas Giedraitis, Harald Hampel, Melissa E. Garcia, Xue Wang, Lars Lannfelt, Patrizia Mecocci, Gudny Eiriksdottir, Paul K. Crane, Florence Pasquier, Virginia Boccardi, Isabel Henández, Robert C. Barber, Martin Scherer, Lluis Tarraga, Perrie M. Adams, Markus Leber, Yuning Chen, Marilyn S. Albert, Steffi Riedel-Heller, Valur Emilsson, Duane Beekly, Anne Braae, Reinhold Schmidt, Deborah Blacker, Carlo Masullo, Helena Schmidt, Rachelle S. Doody, Gianfranco Spalletta, W. T. Longstreth, Thomas J. Fairchild, Paola Bossù, Oscar L. Lopez, Matthew P. Frosch, Eleonora Sacchinelli, Bernardino Ghetti, Qiong Yang, Ryan M. Huebinger, Frank Jessen, Shuo Li, M. Ilyas Kamboh, John Morris, Oscar Sotolongo-Grau, Mindy J. Katz, Chris Corcoran, Melanie Dunstan, Amy Braddel, Charlene Thomas, Alun Meggy, Rachel Marshall, Amy Gerrish, Jade Chapman, Miquel Aguilar, Sarah Taylor, Matt Hill, Mònica Díez Fairén, Angela Hodges, Bruno Vellas, Hilkka Soininen, Iwona Kloszewska, Makrina Daniilidou, James Uphill, Yogen Patel, Joseph T. Hughes, Jenny Lord, James Turton, Annette M. Hartmann, Roberta Cecchetti, Chiara Fenoglio, Maria Serpente, Marina Arcaro, Carlo Caltagirone, Maria Donata Orfei, Antonio Ciaramella, Sabrina Pichler, Manuel Mayhaus, Wei Gu, Alberto Lleó, Juan Fortea, Rafael Blesa, Imelda S. Barber, Keeley Brookes, Chiara Cupidi, Raffaele Giovanni Maletta, David Carrell, Sandro Sorbi, Susanne Moebus, Maria Urbano, Alberto Pilotto, Johannes Kornhuber, Paolo Bosco, Stephen Todd, David Craig, Janet Johnston, Michael Gill, Brian Lawlor, Aoibhinn Lynch, Nick C. Fox, John Hardy, ARUK Consortium, Roger L. Albin, Liana G. Apostolova, Steven E. Arnold, Sanjay Asthana, Craig S. Atwood, Clinton T. Baldwin, Lisa L. Barnes, Sandra Barral, Thomas G. Beach, James T. Becker, Eileen H. Bigio, Thomas D. Bird, Bradley F. Boeve, James D. Bowen, Adam Boxer, James R. Burke, Jeffrey M. Burns, Joseph D. Buxbaum, Nigel J. Cairns, Chuanhai Cao, Chris S. Carlson, Cynthia M. Carlsson, Regina M. Carney, Minerva M. Carrasquillo, Steven L. Carroll, Carolina Ceballos Diaz, Helena C. Chui, David G. Clark, David H. Cribbs, Elizabeth A. Crocco, Charles DeCarli, Malcolm Dick, Ranjan Duara, Denis A. Evans, Kelley M. Faber, Kenneth B. Fallon, David W. Fardo, Martin R. Farlow, Steven Ferris, Tatiana M. Foroud, Douglas R. Galasko, Marla Gearing, Daniel H. Geschwind, John R. Gilbert, Neill R. Graff-Radford, Robert C. Green, John H. Growdon, Ronald L. Hamilton, Lindy E. Harrell, Lawrence S. Honig, Matthew J. Huentelman, Christine M. Hulette, Bradley T. Hyman, Gail P. Jarvik, Erin Abner, Lee-Way Jin, Gyungah Jun, Anna Karydas, Jeffrey A. Kaye, Ronald Kim, Neil W. Kowall, Joel H. Kramer, Frank M. LaFerla, James J. Lah, James B. Leverenz, Allan I. Levey, Ge Li, Andrew P. Lieberman, Kathryn L. Lunetta, Constantine G. Lyketsos, Daniel C. Marson, Frank Martiniuk, Deborah C. Mash, Eliezer Masliah, Wayne C. McCormick, Susan M. McCurry, Andrew N. McDavid, Ann C. McKee, Marsel Mesulam, Bruce L. Miller, Carol A. Miller, Joshua W. Miller, John C. Morris, Jill R. Murrell, Amanda J. Myers, Sid O'Bryant, John M. Olichney, Vernon S. Pankratz, Joseph E. Parisi, Henry L. Paulson, William Perry, Elaine Peskind, Aimee Pierce, Wayne W. Poon, Huntington Potter, Joseph F. Quinn, Ashok Raj, Murray Raskind, Barry Reisberg, Christiane Reitz, John M. Ringman, Erik D. Roberson, Ekaterina Rogaeva, Howard J. Rosen, Roger N. Rosenberg, Mark A. Sager, Andrew J. Saykin, Julie A. Schneider, Lon S. Schneider, William W. Seeley, Amanda G. Smith, Joshua A. Sonnen, Salvatore Spina, Robert A. Stern, Russell H. Swerdlow, Rudolph E. Tanzi, Tricia A. Thornton-Wells, John Q. Trojanowski, Juan C. Troncoso, Vivianna M. Van Deerlin, Linda J. Van Eldik, Harry V. Vinters, Jean Paul Vonsattel, Sandra Weintraub, Kathleen A. Welsh-Bohmer, Kirk C. Wilhelmsen, Jennifer Williamson, Thomas S. Wingo, Randall L. Woltjer, Clinton B. Wright, Chang-En Yu, Lei Yu, Fabienne Garzia, Feroze Golamaully, Gislain Septier, Sebastien Engelborghs, Rik Vandenberghe, Peter P. De Deyn, Carmen Muñoz Fernadez, Yoland Aladro Benito, Hakan Thonberg, Charlotte Forsell, Lena Lilius, Anne Kinhult-Stählbom, Lena Kilander, RoseMarie Brundin, Letizia Concari, Seppo Helisalmi, Anne Maria Koivisto, Annakaisa Haapasalo, Vincent Dermecourt, Nathalie Fievet, Olivier Hanon, Carole Dufouil, Alexis Brice, Karen Ritchie, Bruno

medRxiv preprint doi: https://doi.org/10.1101/2024.10.28.24314051; this version posted November 1, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Dubois, Jayanadra J. Himali, C. Dirk Keene, JoAnn Tschanz, Annette L. Fitzpatrick, Walter A. Kukull, Maria Norton, Thor Aspelund, Eric B. Larson, Ron Munger, Jerome I. Rotter, Richard B. Lipton, María J. Bullido, Albert Hofman, Thomas J. Montine, Eliecer Coto, Eric Boerwinkle, Ronald C. Petersen, Victoria Alvarez, Fernando Rivadeneira, Eric M. Reiman, Maura Gallo, Christopher J. O'Donnell, Joan S. Reisch, Amalia Cecilia Bruni, Donald R. Royall, Martin Dichgans, Mary Sano, Daniela Galimberti, Peter St George-Hyslop, Elio Scarpini, Debby W. Tsuang, Michelangelo Mancuso, Ubaldo Bonuccelli, Ashley R. Winslow, Antonio Daniele, Chuang-Kuo Wu, GERAD/PERADES, CHARGE, ADGC, EADI, Oliver Peters, Benedetta Nacmias, Matthias Riemenschneider, Reinhard Heun, Carol Brayne, David C. Rubinsztein, Jose Bras, Rita Guerreiro, Ammar Al-Chalabi, Christopher E. Shaw, John Collinge, David Mann, Magda Tsolaki, Jordi Clarimón, Rebecca Sussams, Simon Lovestone, Michael C. O'Donovan, Michael J. Owen, Timothy W. Behrens, Simon Mead, Alison M. Goate, Andre G. Uitterlinden, Clive Holmes, Carlos Cruchaga, Martin Ingelsson, David A. Bennett, John Powell, Todd E. Golde, Caroline Graff, Philip L. De Jager, Kevin Morgan, Nilufer Ertekin-Taner, Onofre Combarros, Bruce M. Psaty, Peter Passmore, Steven G. Younkin, Claudine Berr, Vilmundur Gudnason, Dan Rujescu, Dennis W. Dickson, Jean-François Dartigues, Anita L. DeStefano, Sara Ortega-Cubero, Hakon Hakonarson, Dominique Campion, Merce Boada, John Keoni Kauwe, Lindsay A. Farrer, Christine Van Broeckhoven, M. Arfan Ikram, Lesley Jones, Jonathan L. Haines, Christophe Tzourio, Lenore J. Launer, Valentina Escott-Price, Richard Mayeux, Jean-François Deleuze, Najaf Amin, Peter A. Holmans, Margaret A. Pericak-Vance, Philippe Amouyel, Cornelia M. van Duijn, Alfredo Ramirez, Li-San Wang, Jean-Charles Lambert, Sudha Seshadri, Julie Williams, and Gerard D. Schellenberg. 2017. 'Rare Coding Variants in PLCG2, ABI3, and TREM2 Implicate Microglial-Mediated Innate Immunity in Alzheimer's Disease'. *Nature Genetics* 49(9):1373–84. doi: 10.1038/ng.3916.

Singh, Thoudam Debraj, Seung-Yoon Park, Jae-sung Bae, Youngeun Yun, Yong-Chul Bae, Rang-Woon Park, and In-San Kim. 2010. 'MEGF10 Functions as a Receptor for the Uptake of Amyloid-β'. *FEBS Letters* 584(18):3936–42. doi: 10.1016/j.febslet.2010.08.050.

Singleton, A. B., M. Farrer, J. Johnson, A. Singleton, S. Hague, J. Kachergus, M. Hulihan, T. Peuralinna, A. Dutra, R. Nussbaum, S. Lincoln, A. Crawley, M. Hanson, D. Maraganore, C. Adler, M. R. Cookson, M. Muenter, M. Baptista, D. Miller, J. Blancato, J. Hardy, and K. Gwinn-Hardy. 2003. 'Alpha-Synuclein Locus Triplication Causes Parkinson's Disease'. *Science (New York, N.Y.)* 302(5646):841. doi: 10.1126/science.1090278.

Soldan, Anja, Corinne Pettigrew, Qing Cai, Jiangxia Wang, Mei-Cheng Wang, Abhay Moghekar, Michael I. Miller, Marilyn Albert, and BIOCARD Research Team. 2017. 'Cognitive Reserve and Long-Term Change in Cognition in Aging and Preclinical Alzheimer's Disease'. *Neurobiology of Aging* 60:164–72. doi: 10.1016/j.neurobiolaging.2017.09.002.

Stefansson, Hreinn, G. Bragi Walters, Gardar Sveinbjornsson, Vinicius Tragante, Gudmundur Einarsson, Hannes Helgason, Asgeir Sigurðsson, Doruk Beyter, Audunn S. Snaebjarnarson, Erna V. Ivarsdottir, Gudmar Thorleifsson, Bjarni V. Halldorsson, Gudmundur Norddahl, Unnur Styrkarsdottir, Arni Sturluson, Hilma Holm, Agnar Helgason, Kristjan Moore, Hannes P. Eggertsson, Asmundur H. Oddsson, Gudrun A. Jonsdottir, Arni F. Gunnarsson, Gyda Bjornsdottir, Rosa S. Gisladottir, Thorgeir E. Thorgeirsson, Astros Skuladottir, Daniel F. Gudbjartsson, Patrick Sulem, Palmi Jonsson, Steinunn Thordardottir, Jon Snaedal, Helga Eyjolfsdottir, Byron Creese, Clive Ballard, Anne Corbett, Miguel Vasconcelos Da Silva, Dag Aarsland, Ole A. Andreassen, DemGen Study Group, Geir Selbæk, Srdjan Djurovic, Eystein Stordal, Thormod Fladby, Jan Haavik, Jannicke Igland, Lasse M. Giil, Sture Eriksson, Göran Hallmans, Hugo Lövheim, Karin Lopatko Lindman, Miles Trupp, Lars Forsgren, Thomas Werge, Karina Banasik, Søren Brunak, Henrik Ullum, Ruth Frikke-Schmidt, Sisse R. Ostrowski, DBDS

medRxiv preprint doi: https://doi.org/10.1101/2024.10.28.24314051; this version posted November 1, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Genomic Consortium, Maria Didriksen, Erik Sørensen, Anja H. Simonsen, Jørgen E. Nielsen, Gunhild Waldemar, Ole B. Pedersen, Christian Erikstrup, Kirk U. Knowlton, Lincoln D. Nadauld, Kari Stefansson, and DemGen Study Group and DBDS Genomic Consortium. 2024. 'Homozygosity for R47H in TREM2 and the Risk of Alzheimer's Disease'. *The New England Journal of Medicine* 390(23):2217–19. doi: 10.1056/NEJMc2314334.

Steinberg, Stacy, Hreinn Stefansson, Thorlakur Jonsson, Hrefna Johannsdottir, Andres Ingason, Hannes Helgason, Patrick Sulem, Olafur Th Magnusson, Sigurjon A. Gudjonsson, Unnur Unnsteinsdottir, Augustine Kong, Seppo Helisalmi, Hilkka Soininen, James J. Lah, DemGene, Dag Aarsland, Tormod Fladby, Ingun D. Ulstein, Srdjan Djurovic, Sigrid B. Sando, Linda R. White, Gun-Peggy Knudsen, Lars T. Westlye, Geir Selbæk, Ina Giegling, Harald Hampel, Mikko Hiltunen, Allan I. Levey, Ole A. Andreassen, Dan Rujescu, Palmi V. Jonsson, Sigurbjorn Bjornsson, Jon Snaedal, and Kari Stefansson. 2015. 'Loss-of-Function Variants in ABCA7 Confer Risk of Alzheimer's Disease'. *Nature Genetics* 47(5):445–47. doi: 10.1038/ng.3246.

Stern, Yaakov. 2012. 'Cognitive Reserve in Ageing and Alzheimer's Disease'. *The Lancet. Neurology* 11(11):1006–12. doi: 10.1016/S1474-4422(12)70191-6.

Teo, Shu Mei, Yudi Pawitan, Chee Seng Ku, Kee Seng Chia, and Agus Salim. 2012. 'Statistical Challenges Associated with Detecting Copy Number Variations with Next-Generation Sequencing'. *Bioinformatics* 28(21):2711–18. doi: 10.1093/bioinformatics/bts535.

Villamil-Ortiz, Javier G., and Gloria Patricia Cardona-Gómez. 2018. 'cPLA2 and Desaturases Underlie the Tau Hyperphosphorylation Offset Induced by BACE Knock-down in Neuronal Primary Cultures'. *Biochimica Et Biophysica Acta. Molecular Basis of Disease* 1864(11):3696–3707. doi: 10.1016/j.bbadis.2018.08.028.

Wallon, David, Susana Boluda, Anne Rovelet-Lecrux, Manon Thierry, Julien Lagarde, Laetitia Miguel, Magalie Lecourtois, Antoine Bonnevalle, Marie Sarazin, Michel Bottlaender, Mathieu Mula, Serge Marty, Natsuko Nakamura, Catherine Schramm, François Sellal, Thérèse Jonveaux, Camille Heitz, Isabelle Le Ber, Stéphane Epelbaum, Eloi Magnin, Aline Zarea, Stéphane Rousseau, Olivier Quenez, Didier Hannequin, Florence Clavaguera, Dominique Campion, Charles Duyckaerts, and Gaël Nicolas. 2021. 'Clinical and Neuropathological Diversity of Tauopathy in MAPT Duplication Carriers'. *Acta Neuropathologica* 142(2):259–78. doi: 10.1007/s00401-021-02320-4.

Wang, Hui, Beth A. Dombroski, Po-Liang Cheng, Albert Tucci, Ya-Qin Si, John J. Farrell, Jung-Ying Tzeng, Yuk Yee Leung, John S. Malamon, Alzheimer's Disease Sequencing Project, Li-San Wang, Badri N. Vardarajan, Lindsay A. Farrer, Gerard D. Schellenberg, and Wan-Ping Lee. 2023. 'Structural Variation Detection and Association Analysis of Whole-Genome-Sequence Data from 16,905 Alzheimer's Diseases Sequencing Project Subjects'. *medRxiv: The Preprint Server for Health Sciences* 2023.09.13.23295505. doi: 10.1101/2023.09.13.23295505.

Wang, Hui, Li-San Wang, Gerard Schellenberg, and Wan-Ping Lee. 2022. 'The Role of Structural Variations in Alzheimer's Disease and Other Neurodegenerative Diseases'. *Frontiers in Aging Neuroscience* 14:1073905. doi: 10.3389/fnagi.2022.1073905.

Wilkinson, Kim, and Joseph El Khoury. 2012. 'Microglial Scavenger Receptors and Their Roles in the Pathogenesis of Alzheimer's Disease'. *International Journal of Alzheimer's Disease* 2012:489456. doi: 10.1155/2012/489456.

Woodward, Karen J., Julie Stampalia, Hannah Vanyai, Hashika Rijhumal, Kim Potts, Fiona Taylor, Joanne Peverall, Tanya Grumball, Soruba Sivamoorthy, Hamid Alinejad-Rokny, John Wray,

medRxiv preprint doi: https://doi.org/10.1101/2024.10.28.24314051; this version posted November 1, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Andrew Whitehouse, Lakshmi Nagarajan, Jacqueline Scurlock, Sabine Afchani, Matthew Edwards, Ashleigh Murch, John Beilby, Gareth Baynam, Cathy Kiraly-Borri, Fiona McKenzie, and Julian I. T. Heng. 2019. 'Atypical Nested 22q11.2 Duplications between LCR22B and LCR22D Are Associated with Neurodevelopmental Phenotypes Including Autism Spectrum Disorder with Incomplete Penetrance'. *Molecular Genetics & Genomic Medicine* 7(2):e00507. doi: 10.1002/mgg3.507.

Zhao, Min, Qingguo Wang, Quan Wang, Peilin Jia, and Zhongming Zhao. 2013. 'Computational Tools for Copy Number Variation (CNV) Detection Using next-Generation Sequencing Data: Features and Perspectives'. *BMC Bioinformatics* 14 Suppl 11(Suppl 11):S1. doi: 10.1186/1471-2105-14-S11-S1.

medRxiv preprint doi: https://doi.org/10.1101/2024.10.28.24314051; this version posted November 1, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Tables

	EOAD	LOAD	All AD	Controls
N	4,077	8,457	12,534	9,559
Age y mean (SD)[1]	58.43 (5.09)	77.90 (7.48)	71.38 (11.42)	81.36 (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 analysis

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

medRxiv preprint doi: https://doi.org/10.1101/2024.10.28.24314051; this version posted November 1, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Sample	Cohort	AAO	Gender	CNV coordinate	HGVS nomenclature	Gene	Event
ALZ-0596-001[1]	ADES-FR	[45-50]	Male	chr17:43,679,169-44,145,099	NC_000017.10:g.43585921_44159772dup	MAPT	complete duplication
ALZ-0441-005[1]	ADES-FR	[45-50]	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**	[45-50]	**Male**	**chr21:22,370,838-28,338,827**	**NC_000021.8:g.20230358_30248689dup**	**APP**	**complete duplication**
TE17A00020a_DS1	**PERADES**	[45-50]	**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**	[45-50]	**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**	**[45-50]**	**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. Mendelien 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

medRxiv preprint doi: https://doi.org/10.1101/2024.10.28.24314051; this version posted November 1, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

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	6.62 [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

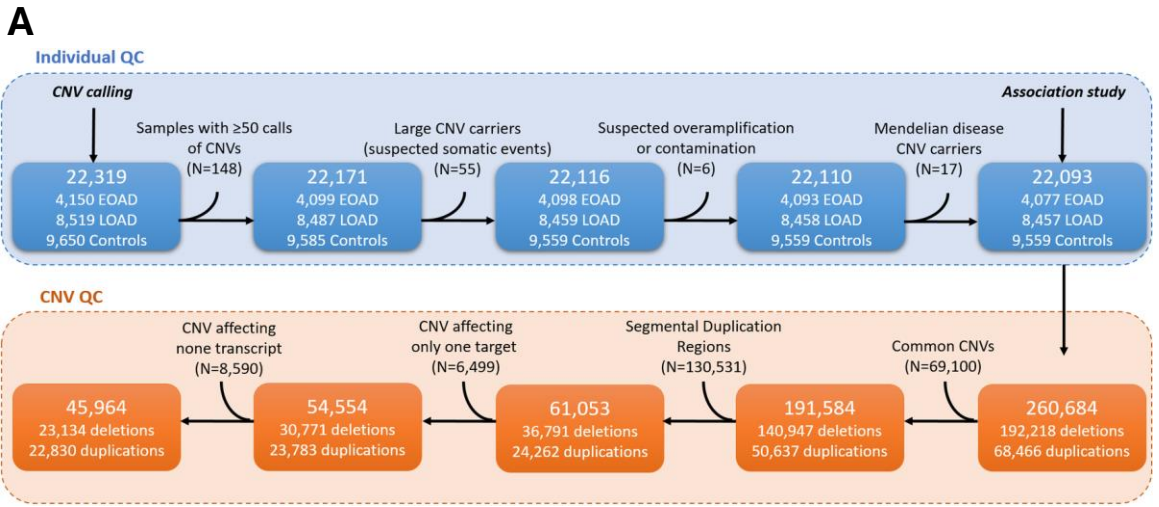
[1]OR should be interpreted for an increase of transcript copies by 1. i.e. a OR > 1 (resp. OR <1) means that duplications (resp. deletions) are associated with a higher disease risk

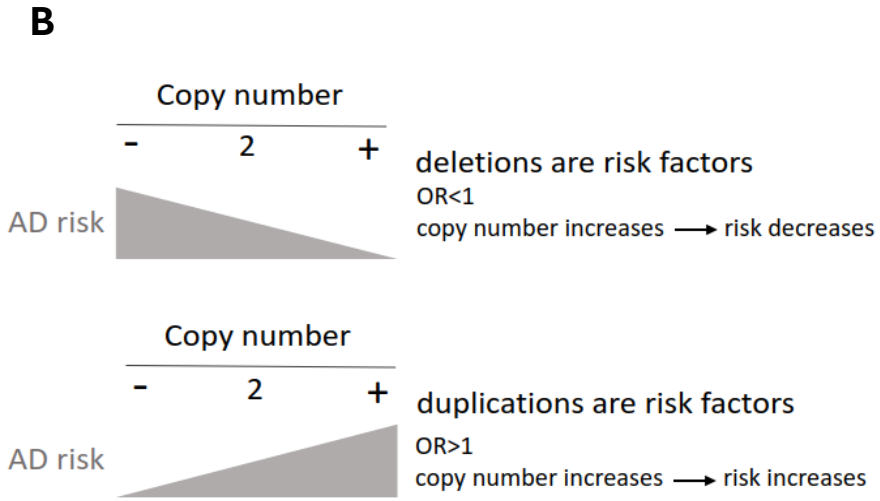
[2]non adjusted p-values. selected based on FRD < 10%

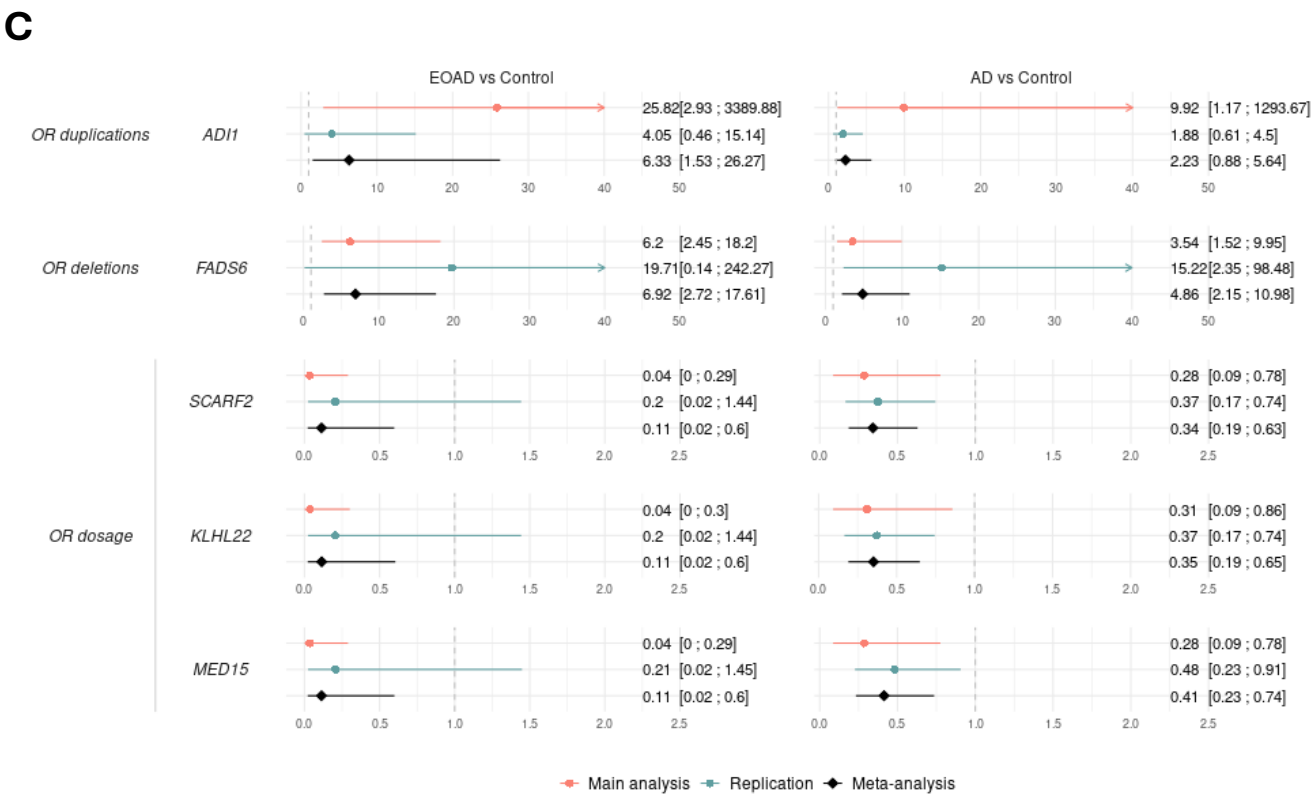
* gene overlapping with repeats or duplicated gene in the genome (despite CNVs not overlapping >50% with repeats)

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

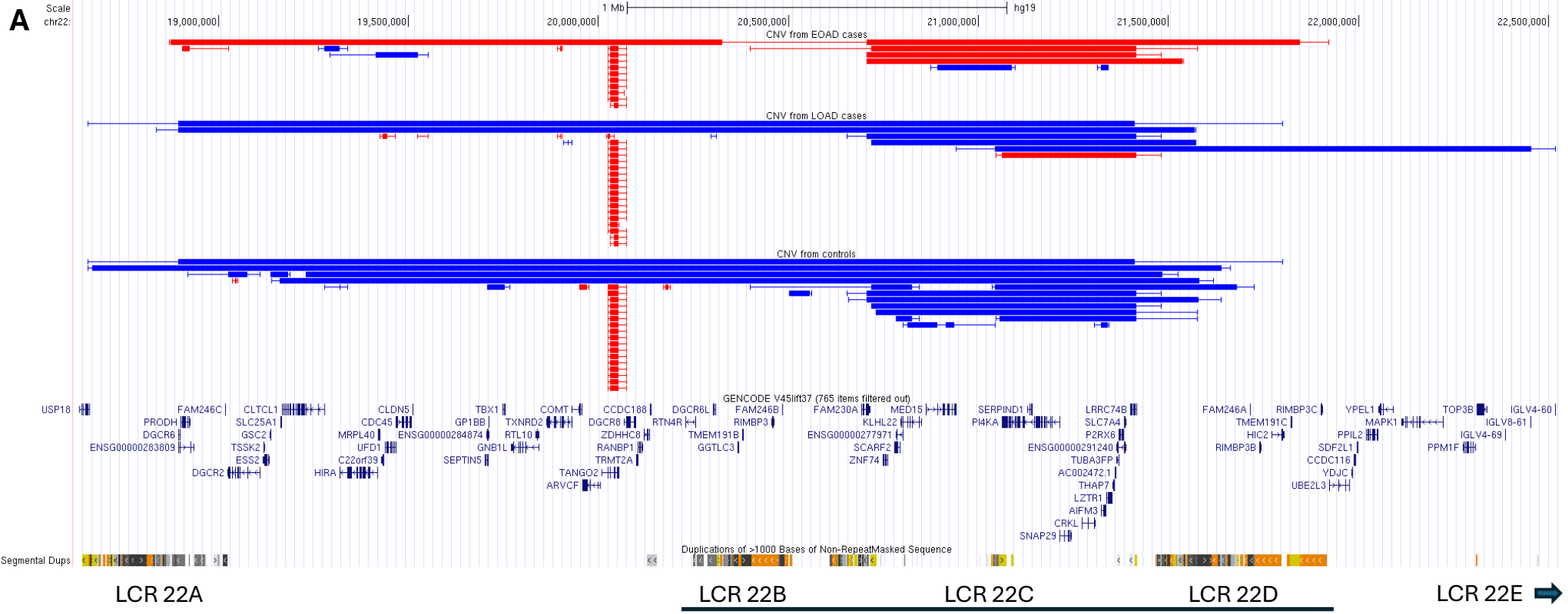
Genes in the same locus are marked by a colored block on the left


A
Individual QC
CNV calling
Association study
Samples with ≥50 calls of CNVs (N=148)
Large CNV carriers (suspected somatic events) (N=55)
Suspected overamplification or contamination (N=6)
Mendelian disease CNV carriers (N=17)
22,319 4,150 EOAD 8,519 LOAD 9,650 Controls
22,171 4,099 EOAD 8,487 LOAD 9,585 Controls
22,116 4,098 EOAD 8,459 LOAD 9,559 Controls
22,110 4,093 EOAD 8,458 LOAD 9,559 Controls
22,093 4,077 EOAD 8,457 LOAD 9,559 Controls
CNV QC
CNV affecting none transcript (N=8,590)
CNV affecting only one target (N=6,499)
Segmental Duplication Regions (N=130,531)
Common CNVs (N=69,100)
45,964 23,134 deletions 22,830 duplications
54,554 30,771 deletions 23,783 duplications
61,053 36,791 deletions 24,262 duplications
191,584 140,947 deletions 50,637 duplications
260,684 192,218 deletions 68,466 duplications

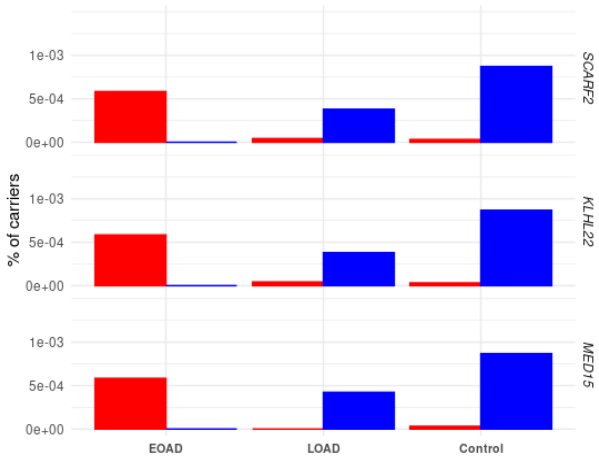

B
Copy number
-
2
+
AD risk
deletions are risk factors
OR<1
copy number increases → risk decreases
Copy number
-
2
+
AD risk
duplications are risk factors
OR>1
copy number increases → risk increases


C
EOAD vs Control
AD vs Control
OR duplications
ADI1
25.82[2.93 ; 3389.88]
4.05 [0.46 ; 15.14]
6.33 [1.53 ; 26.27]
9.92 [1.17 ; 1293.67]
1.88 [0.61 ; 4.5]
2.23 [0.88 ; 5.64]
OR deletions
FADS6
6.2 [2.45 ; 18.2]
19.71[0.14 ; 242.27]
6.92 [2.72 ; 17.61]
3.54 [1.52 ; 9.95]
15.22[2.35 ; 98.48]
4.86 [2.15 ; 10.98]
OR dosage
SCARF2
0.04 [0 ; 0.29]
0.2 [0.02 ; 1.44]
0.11 [0.02 ; 0.6]
0.28 [0.09 ; 0.78]
0.37 [0.17 ; 0.74]
0.34 [0.19 ; 0.63]
KLHL22
0.04 [0 ; 0.3]
0.2 [0.02 ; 1.44]
0.11 [0.02 ; 0.6]
0.31 [0.09 ; 0.86]
0.37 [0.17 ; 0.74]
0.35 [0.19 ; 0.65]
MED15
0.04 [0 ; 0.29]
0.21 [0.02 ; 1.45]
0.11 [0.02 ; 0.6]
0.28 [0.09 ; 0.78]
0.48 [0.23 ; 0.91]
0.41 [0.23 ; 0.74]
Main analysis
Replication
Meta-analysis

A



B



C

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