TU Delft

Delft University of Technology

Document Version
Final published version

Licence
CC BY

Citation (APA)
Zarea, A., Cassinari, K., Lecoquierre, F., Quenez, O., Charbonnier, C., Schramm, C., Lacour, M., Holstege, H., Hulsman, M., & More Authors (2026). Assessing the de novo paradigm in sporadic early-onset Alzheimer disease trios. *Molecular Psychiatry*, *31*(10), 5879-5888. https://doi.org/10.1038/s41380-026-03665-6

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
In case the licence states “Dutch Copyright Act (Article 25fa)”, this publication was made available Green Open Access via the TU Delft Institutional Repository pursuant to Dutch Copyright Act (Article 25fa, the Taverne amendment). This provision does not affect copyright ownership.
Unless copyright is transferred by contract or statute, it remains with the copyright holder.

Sharing and reuse
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.

ARTICLE OPEN

Check for updates

Assessing the de novo paradigm in sporadic early-onset Alzheimer disease trios

Aline Zarea[1,43], Kevin Cassinari[2,43], François Lecoquierre[2,43], Olivier Quenez[2,43], Camille Charbonnier[3,43], Catherine Schramm[4], Morgane Lacour[1], Stéphane Rousseau[2], Anne-Claire Richard[2], Anne Rovelet-Lecrux[4], Magalie Lecourtois[4], Robert Olaso[5], Anne Boland[5], Jean-François Deleuze[5], Christian Gilissen[6,7], Joris A. Veltman[8], Lisenka ELM Vissers[6,7], Céline Bellenguez[9], Orio Dols-Icardo[10,11], John Hardy[12], Henne Holstege[13,14,15,16], Marc Hulsman[13,14,15,16], Jean-Charles Lambert[9], Simon Mead[17], Alfredo Ramirez[18,19,20,21], Rebecca Sims[22], John van Swieten[23], Michael Wagner[18,24], Julie Williams[22,25], Stéphanie Bombois[26], Claire Boutoleau-Bretonniere[27], Ludivine Charmard-Witkowski[28], Vincent de la Sayette[29], Vincent Deramecourt[30], Frédérique Etcharry-Bouyx[31], Audrey Gabelle[32], Claude Gueriot[33], Gwenaël Le Guyader[34], Isabelle Le Ber[35], Thibaud Lebouvier[30], Olivier Martinaud[36], Agnès Michon[37], Chloé Quelin[38], Marie Sarazin[39,40], Mathieu Sévin[41], Christel Thauvin-Robinet[42], David Wallon[1] and Gaël Nicolas[2]✉

© The Author(s) 2026

The genetic architecture of sporadic Early-Onset Alzheimer Disease (sEOAD, onset ≤65 years) remains largely unknown. To assess the de novo mutation (DNM) hypothesis, we performed a nationwide recruitment of 37 novel sEOAD patients–unaffected parents trios. After assessing known monogenic genes, we performed trio-based exome sequencing and jointly analyzed novel trios with 12 previously reported ones. Of these, we selected 16 trios for genome sequencing. We identified three patients with a pathogenic DNM in *APP* or *PSEN1*. Then, from the 46 remaining trios, we identified 38 non-synonymous coding DNM and 4 de novo copy number variants (CNVs) in exome data. Four DNM (2 novel, in *SPHK2* and *DDR1*) and bi-allelic inherited variants in two genes affected Alzheimer disease-related genes. No significant burden of rare coding variants in exome/genome data from 5643 EOAD cases and 16097 controls was identified using nested windows centered on each DNM position, at the transcript level. From genome data, one non-coding DNM was predicted to affect splicing in an AD-associated gene, *PINX1*. Overall, 48% probands carried ≥1 inherited risk factor with odds ratio (OR) > 1.5 and GWAS-defined Genetic Risk Scores (GRS) distribution was more consistent with random distribution than enrichment in higher scores in probands. We confirm that DNMs in known monogenic genes explain sEOAD in a minority of cases, while candidate DNMs in other genes might account for a small proportion of additional cases. The majority of sEOAD patients may have a complex etiology including multiple inherited variants, however, GRS might not explain most of its genetic component.

Molecular Psychiatry; https://doi.org/10.1038/s41380-026-03665-6

INTRODUCTION

De novo mutations (DNMs) are a major cause of genetic disorders. Beyond the well-established example of developmental disorders [1, 2], DNMs can also account for a proportion of sporadic cases of classically dominantly inherited conditions, i.e., cases with a negative family history. In adult-onset neurodegenerative disorders, DNMs have been implicated in early-onset, sporadic cases of amyotrophic lateral sclerosis, fronto-temporal dementia, or Alzheimer disease (AD), for example [3].

Trio-based exome or genome sequencing, consisting in sequencing a proband and both unaffected parents, is a powerful method to identify DNMs in known Mendelian genes as well as novel candidate genes. However, its application to adult-onset neurodegenerative disorders is challenged by three main factors: (1) difficulty obtaining DNA from elderly unaffected parents, (2) the need of a large number of trios to establish significant recurrence, and (3) disease etiology heterogeneity, as only some cases are monogenic and a majority involve complex etiology combining genetic and non-genetic factors. DNMs could theoretically contribute to the disease etiology at any level, from monogenic causes with full penetrance, to diverse risk factor effects, thus explaining part of the disease etiology or an earlier onset in a predisposed individual.

About 12% of early-onset AD (EOAD, onset ≤65 years) cases are caused by autosomal dominant pathogenic variants in *APP*, *PSEN1*, *PSEN2* or by *APP* duplications, with average ages at onset (AAO) of 44.4, 50.9, 53.9, and 51.1, respectively [4, 5]. In prior work, we showed that 12.3% of sporadic EOAD patients with onset before 51 years carried such variants [4, 6]. When parental DNA was available, we confirmed their de novo origin. Building on this, we hypothesized that DNMs in other genes may also contribute to EOAD. We previously applied trio-based exome sequencing to 12 sporadic EOAD trios without pathogenic variants in monogenic AD genes and found two coding DNMs with functional effects in

A full list of author affiliations appears at the end of the paper.

Received: 2 October 2025 Revised: 24 April 2026 Accepted: 20 May 2026
Published online: 29 May 2026

line with known AD pathophysiological processes, in *VPS35* and *MARK4* [7]. Importantly, these effects depended on variant position, as the *VPS35* p.L625P variant showed distinct consequences from the Parkinson linked D620N variant, and the *MARK4* delins altered a short linker domain. However, these cases remain unique, so far.

Most AD cases are not monogenic but show a multifactorial origin, with a substantial genetic component [8], including non-monogenic EOAD [9]. Rare coding variants in *SORL1*, *TREM2*, *ABCA7*, *ABCA1* or *ATP8B4*, along with the common *APOE* ε4 allele, are considered the main moderate-to-strong AD risk factors [5]. They may be inherited from either symptomatic or asymptomatic parents, as they are not fully penetrant. Their presence is thus not unexpected in sporadic EOAD, though inheritance is rarely assessed. In addition, about 80 common variants and a handful of rare recurrent variants confer modest AD risk [10]. Their combined effect is commonly assessed by genetic scores, although the contribution of these scores to AD susceptibility is likely lower than the above-mentioned moderate-to-strong risk factors. Importantly, from a mechanistic point of view, autosomal dominant AD genes, moderate-to-strong risk factors, and a number of common risk variants clearly place the aggregation of the Amyloid β (Aβ) peptide as a critical factor in the pathophysiology.

To investigate the role of DNMs in sporadic EOAD, we conducted a nationwide recruitment of sporadic EOAD trios in France. We applied a sequential genetic screen starting with known monogenic genes analysis, followed by trio-based exome sequencing with a joint analysis with previously reported trios. We further selected a subset of them for trio-based genome sequencing (GS). After interpretation of DNMs, we performed burden testing of rare coding variants centered on the positions of the non-synonymous DNMs, in a large case-control dataset. Finally, we assessed the role of inherited rare risk variants and genetic risk scores.

METHODS

Statistics were performed using R version 3.6.3 (2020-02-29) [11]

Patients

Following our 2015 study on EOAD trios [7], we considered for inclusion as a potential trio, every patient with early-onset AD (EOAD, onset before 65 years), living unaffected parents, and no family history of EOAD, recruited nationally across France.

Prior to any genetic testing request, patients are first diagnosed in local memory clinics, in one of more than 30 hospitals, based on clinical examination, brain imaging and cerebrospinal fluid (CSF) AD biomarkers (Aβ42 and Aβ42/40 when locally available, Tau, Phosphorylated Tau at position 181 (P-Tau)) and follow the IWG-2 AD research criteria [12]. CSF AD biomarkers are considered as AD-supportive in cases of A) decreased Aβ42 levels or [Aβ42 /Aβ40] ratio, and B) increased Tau or P-Tau level or decreased [P-Tau/ Aβ42] ratio (0.122 cutoff [13]), local laboratory normative values being used for each biomarker per guidelines. A pedigree is drawn based on direct patient and family informant interview by the referring physician. Then, clinical information is sent to the national reference center for young Alzheimer patients (CNRMAJ) of Rouen together with the patient's written consent for genetic analysis, and a blood sample (stored in EDTA), as part of a request for genetic analysis in a diagnostic and research context. Every request is reviewed by an expert neurologist from the CNRMAJ.

In case of identification of a potential EOAD trio following this first screen based on the first pedigree, a contact with parents is proposed by the referring physician or directly by one neurologist of the CNRMAJ of Rouen and, if they agree, a medical interview is performed, possibly including mini mental state examination (MMSE) assessment or equivalent phone-based interview, when possible, and further family history information is sought. Based on this interview and on available medical information on the proband, the following inclusion criteria are finally considered for an inclusion as a trio: (i) proband with probable EOAD and positive CSF biomarkers (see above), (ii) living, unaffected parents, i.e. no personal history of cognitive decline or neurodegenerative disease by interview, (iii) parents agree to perform genetic analyses and (iv) absence of family history of EOAD in first and second-degree relatives. A blood sample of the parents is taken after providing informed written consent in a medical setting including a query on research, and samples are stored for further genetic analyses.

Genetic analysis strategy

Probands are screened by targeted screening (Sanger sequencing of *APP*, *PSEN1*, and copy number evaluation of *APP* by Quantitative Multiplex PCR of Short Fluorescent fragments, QMPSF) or by first-tier exome sequencing. If targeted screening is negative, second-tier exome sequencing is performed. From first-tier or second-tier exome sequencing data, rare variants from a list of genes associated with monogenic causes of dementia [14] are interpreted.

In cases of identification of a likely pathogenic or pathogenic variant, parental inheritance is assessed by a targeted technique. In negative patients, parental exomes are obtained.

Parenthood is confirmed through 4 informative microsatellites PCR and/or from trio-based exome/GS data.

The analysis of sequencing data was approved by the CERDE ethics committee from the Rouen University Hospital (notification 2019-055).

Exome sequencing

Exomes from new trios (recruitment: 2015-2022) was obtained using Agilent Sureselect human all exons kit V5 ($n = 9$), V5 + UTR ($n = 15$) or V6 + UTR ($n = 11$) (Table S1) capture then Illumina sequencing (Hiseq4000 or Novaseq6000, $n = 4$, >100x average depth), in the CNRGH center (*Centre National de Recherche en Génomique Humaine*, Evry, France).

Bioinformatics analyses were performed in the CNRMAJ and Inserm U1245, Rouen. Fastq files were aligned to GRCh37 using BWA 0.7.17, duplicates flagged using GATK 4.2.1.0, SNVs and indels called by DeepVariant v1.5 and merged using GLnexus v1.2.7. CNV screening was performed using the CANOES tool [15] based on our previously published workflow [16].

Previously reported negative trios (i.e., no likely pathogenic variant in monogenic gene, $n = 12$ [7]) were reprocessed using the same pipeline.

Genome sequencing

Sixteen trios were randomly selected for genome sequencing (GS), among those who also consented to the RBM-0259 study, approved by the CPP Ile de France II ethics committee and with enough remaining DNA available at the time of the sequencing (year 2017). Fifty-base pair paired-end GS was performed on a BGIseq500 sequencer at the Beijing Genetics Institute, Shenzhen, China, following DNA nanoball generation. Raw data extraction, quality filtration (reads with more than 10% low quality bases, read with >1% non-sequenced) and mapping were performed in the Department of Human Genetics, Radboud UMC, Nijmegen (The Netherlands) using BWA and GATK. SNVs and indels were called using GATK HaplotypeCaller. Because of the short reads (50 bp) used here, we were not able to accurately detect structural variations from GS data.

Identification of de novo variants

De novo SNVs and indels were extracted from genome and exome data using similar methods. We applied sequential steps of filtration to the two multi vcf files using Bcftools through a custom script (https://github.com/francois-lecoquierre/de_novo_tools/blob/main/DeNovoMiner.py). Filters applied on both variant sets included (i) depth (DP) > 10 in all three individuals of the trio, (ii) variant allele fraction (VAF) > 0.2 in the proband, (iii) allelic ratio in child divided by allelic ratio of each parent greater than 4, and (iv) exclusion of multiallelic sites where sum of allelic depths (AD) AD1 + AD2 < 0.7 x DP. The resulting candidate de novo variants were then systematically reviewed on IGV using a semi-automated method (https://github.com/francois-lecoquierre/genomics_shortcuts/blob/main/classify_vcf_from_igv_v1.0.py). This script allows for the manual review of large numbers of variants by adding a minimal interface to IGV, which presents the variants of a VCF file to the user, gathers their interpretation and exports new analysed VCF files. Sanger sequencing was performed for all candidate DNMs in exons and splice regions (−20, + 20).

CNVs were considered as candidate de novo if they (i) encompassed ≥two capture targets, (ii) were not polymorphic in the database of genomic variants, (iii) did not overlap with >50% with segmental duplications and (iv) were not detected in parents (> 70% overlap with

a parental CNV). A targeted PCR-based (either QMPSF or digital droplet PCR) assay was performed to confirm CNVs and de novo status.

Variant annotation

We used a combination of homemade scripts and annotation software for the annotation of SNV and indels coding and non-coding variants. Coding variants were annotated using the SNPEff/SNPsift tool suite and the CRAVAT [17] software to predict the impact of variants from several bioinformatics prediction tools, including VEST4 [18], VEST4-Indel and REVEL [19]. Variant frequencies from gnomAD were retrieved using bcftools annotate. For non-coding variants, we used the VEP software [20] and its plugins to focus on (i) 5'UTR annotations using the UTRannotator tool [21], (ii) splicing predictions using SpliceAI [22] and (iii) the distance of any DNM to a list of AD genes followed by visualization in the UCSC genome browser. We then filtered and extracted the results using in-house scripts.

Gene-based and nested-window rare variant burden testing

We tested whether genes with DNMs were enriched in rare variants in AD patients using the mega dataset of ADES-ADSP, made of exomes and genomes of cases and controls from Europe (Alzheimer Disease European Sequencing, ADES) and the United States of America (Alzheimer Disease Sequencing Project, ADSP), among patients with a European ancestry, similarly to our trios [23], following an extensive quality control, gathering step 1 (discovery) with step 2 (replication), after exclusion of individuals with only BAM extracts available. We assessed a total of 15,808 cases (5643 EOAD) and 16,097 non-demented controls.

We performed a series of nested burden tests centered around every exonic DNM position (except *MKI67* and *COL19A1* that map to STR-rich regions with low exome quality) using a Firth logistic regression to account for low cumulative dosages. For such case-control analyses in the ADES-ADSP dataset, we selected variants mapping to a 50-bp, 100-bp, or 500-bp window around the position where a DNM had been identified, excluding intronic sequences, up to the whole transcript. Different to what has been previously done in most gene-based analyses, we worked up to the transcript level and also considered short in-frame deletions and insertions (indels) together with missense variants. As the consequences of such variants are difficult to predict, they are indeed generally excluded from exome-wide case-control analyses, including in our previous work [23]. To sort missense variants, we used the VEST4 tool that allows a transcript-specific annotation as well as scores for in-frame indels. All in all, 6387 nested unilateral burden tests were computed to test whether gene regions surrounding DNM positions were enriched in either rare loss-of-function variants, rare missense or in frame indel variants of high VEST4 score or the combination of both categories among all AD or EOAD patients. Strict Bonferroni correction for this number of tests would require to declare statistical significance below 7.8×10^{-6}. As many tests are actually nested within each other and/or strongly correlated within each gene of interest, we also considered a minimal but imperative correction for 35 genes times two separate categories of variants, setting this lenient statistical significance threshold at 0.0007.

CNVs were assessed using logistic regression, in exomes from ADES-ADSP based on partial or full deletions and full duplications, as in the extended dataset in ref. [24], after excluding trio probands.

Analysis of inherited variants

To further expand our study based on EOAD trios, we assessed the role of inherited variants in a multigenic hypothesis context.

First, exome data were screened for bi-allelic variants (homozygous or compound heterozygous) affecting coding regions or canonical splice sites (annotated as HIGH, MODERATE or LOW by SNPEff). Mendelian errors, false calls and multiple nucleotide variants (MNVs or delins) were manually curated.

Second, rare definite risk variants in *SORL1*, *ABCA7*, *TREM2*, *ABCA1* and *ATP8B4* were extracted following our previously reported classification algorithm that clinically classifies variants based on gene and variant-level evidence and sorts them into three categories of effect on AD risk, i.e. modest ($1.5 < \text{Odds Ratio (OR)} \leq 2$), moderate ($2 < \text{OR} < 5$) or strong ($\text{OR} \geq 5$) [5], and evaluated them for parental inheritance.

Finally, Genetic Risk Scores (GRS) excluding *APOE* were computed based on GS data, in every member of all 16 trios. For that purpose, we computed GRS as described in ref. [10], where GRS are divided into deciles according to thresholds defined from a population-based cohort. The GRS is based on 83 variants associated with AD and obtained from the following formula: $\text{GRS}=\frac{83}{\sum \log(OR)}\sum(\log(OR)\times dosage)$. A Mann-Whitney paired test was used to compare GRS of each proband with the mean of GRS from their parents. From parents, a theoretical distribution of possible GRS was generated for each proband and plotted. For each proband, we simulated a theoretical distribution of possible GRS, based on 500 random transmissions of the 83 SNPs. Each distribution was plotted with the corresponding mean ± 2 standard deviation (SD) interval.

RESULTS

Between March 2015 and September 2022, we preselected 102 probands with (i) sporadic EOAD, i.e. a diagnosis of probable AD, an AAO before 65 years, and no family history of EOAD and (ii) living, unaffected parents. After interviews and validation of inclusion criteria, we succeeded in collecting blood samples of 37 trios (Fig. 1). AAO ranged from 36–60 years (average 49.9, SD 4.9), 29 probands were female and 8 male (Table 1, Table S1). Non-included patients had similar AAO (50.8 years, NS, t-test).

Identification of *APP* and *PSEN1* de novo pathogenic single nucleotide variants

Monogenic AD genes were screened either by targeted analysis of *APP* and *PSEN1* followed by second-tier exome sequencing (i.e. targeted screening was performed first, and, if negative, exome was sequenced), or by first-tier exome sequencing (exome being the first-line genetic analysis in these cases). Three probands carried a pathogenic variant, one in *APP* and two in *PSEN1*. Sanger sequencing in the parents revealed that all three variants occurred de novo (Table S1).

One case with logopenic variant of primary progressive aphasia and apraxia from the age of 46 years and positive CSF AD biomarkers (EXT-1868) carried a likely pathogenic (ACMG-AMP Class 4) *PSEN1* chr14(GRCh37):73664820C>G, NM_000021.4:c.851C>G, p.(Pro284Arg) DNM, affecting a conserved residue, where other pathogenic missense changes have been reported. This variant is predicted to be damaging by multiple tools (REVEL score: 0.92). Her unaffected parents were aged 78 and 77 when blood samples were taken.

The other two variants were reported in previous articles from our group, without further details [4, 5]. One pathogenic (class 5) *APP* DNM, chr21(GRCh37):27264105T>C, NM_000484.4:c.2140A>G p.(Thr714Ala) was identified by first-tier exome sequencing in a patient with progressive memory impairment since the age of 41 years (MMSE at inclusion: 8/30) (EFA-0084). CSF biomarkers revealed a typical AD profile, brain MRI was considered as normal, and 18-FDG PET imaging showed biparietal hypometabolism. Her unaffected parents were aged 74 and 72 at the time of blood sample. This *APP* variant, called Iranian *APP*, is located in a mutational hotspot in the γ-secretase cleavage site. It has already been reported in three families from diverse ancestries with autosomal-dominant EOAD [25–27], with AAOs ranging from 44 to 52.

The *PSEN1* chr14(GRCh37):73653568A>G, NM_000021.4:c.488A>G, p.(His163Arg) variant is a recurrent pathogenic (class 5) variant reported in multiple families worldwide. It was identified by Sanger sequencing in patient with a frontal presentation and myoclonic seizures [4, 6] (EXT-1242). The patient developed behavioral changes and difficulties with planning at the age of 49. Brain MRI was normal but 18-FDG-PETscan showed a predominant frontal and parieto-temporal hypometabolism. CSF biomarkers revealed a typical AD profile. Her unaffected parents were aged 77 and 73 at the time of blood sample.

De novo variants in 46 EOAD trio exomes

None of the 34 remaining probands exhibited a pathogenic variant in a monogenic dementia gene. We thus sequenced the

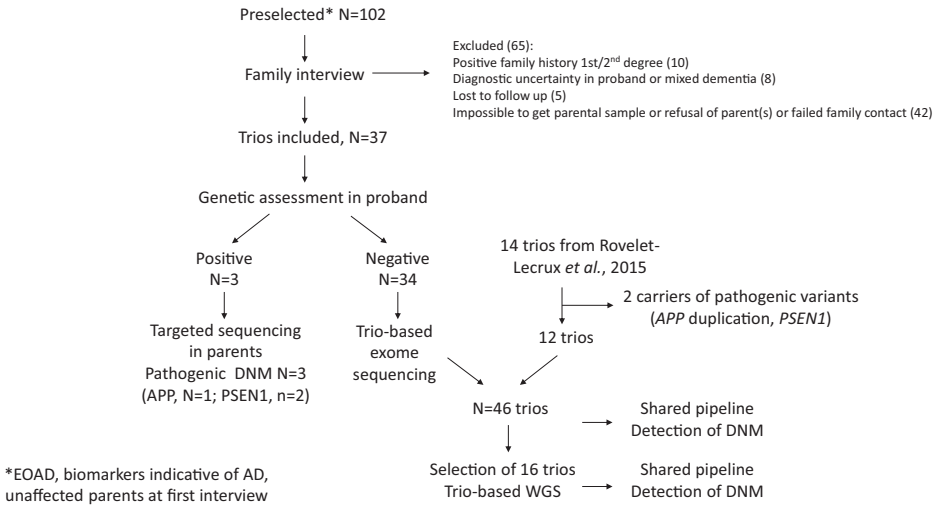


Fig. 1 Study flowchart. Since our first EOAD trio paper in 2015, patients with EOAD and apparently living parents based on the pedigree information provided by the referring physician upon first medical request of a genetic test, we systematically propose to contact the family for a second family interview. A total of 102 patients were preselected, 37 of them were finally included, as parents were actually unaffected by dementia, had a negative family history, accepted a blood sample and signed the informed written consent, and the blood drawing was eventually performed. Three patients presented a pathogenic *APP* or *PSEN1* variant, which was confirmed to be a DNM, and the other 34 patients underwent trio-based exome sequencing. Exome analyses were performed jointly with previously reported trios. Of these 46 trios, we selected 16 for trio-based genome sequencing.

Table 1. Summary information on all novel 37 trios.

Average age at onset [range]	*49.89* [36–60]
Males/Females	*8/29*
N pathogenic variants	*3 (APP, n = 1; PSEN1:n = 2)*
APOE genotypes	*2-3 : N = 2*
	3-3 : N = 24
	3-4 : N = 9
	4-4 : N = 2

exomes of the parents and jointly analyzed those with 12 trios from ref. [7] that did not contain pathogenic variants carriers in monogenic dementia genes either (negative trios) (Fig. 1). All 46 trios were processed using the same bioinformatics pipeline, revealing 53 DNM (1.15 per trio, on average): 38 coding non-synonymous variants, 9 synonymous variants, and 6 splice region variants (all independently confirmed by Sanger sequencing, Table 2). Of note, from previously published exomes, 3 additional DNM were identified by the novel deep-variant-based pipeline (in *LPIN3*, *ANO7* and *COL19A1*).

At the gene level, there was only one recurrence, in the large gene *SPEG* (two DNMs), including one synonymous variant with no predicted consequences on splicing. One patient carried 5 DNMs including 4 with allelic ratios ranging from 24 to 31%, one being a *TET2* truncating variant, suggesting a hematopoietic clone. His blood cell count was normal, suggesting clonal hematopoiesis of indeterminate potential (CHIP). The corresponding DNMs were further excluded from subsequent analyses.

Among the non-synonymous DNMs, four affected Aβ or AD-related genes: the *VPS35* and *MARK4* DNMs in both previously reported patients, a novel *DDR1* missense variant, and an in frame indel in the *SPHK2* gene [7, 28, 29].

Among candidate de novo CNVs, three were confirmed by independent techniques and were actually de novo (Table S2): a 45-kb deletion at 7p22.1 involving the *RSPH10B2* gene and part of the *CCZB1* gene; a deletion at 18q21.33 involving the *SERPINB4* gene; and a unique duplication at 1p34.3 involving the *SNIP1* gene.

De novo variants in 16 EOAD trio genomes

GS (quality metrics: Table S3) retrieved all exome-identified DNMs and detected an additional DNM in *CEP112*, a postzygotic missense variant with a 21% allelic ratio in GS data and 12% in exome alignments.

Overall, we identified 1232 DNMs (1153 SNVs, 79 de novo indels; average 77 DNMs per trio, range:[58-164]). To assess the role of non-coding DNM, (i) we performed a recurrence analysis at the gene level, (ii) we assessed consequences on splicing and (iii) other consequences on UTRs, and (iv) we assessed whether DNMs mapping in the genomic regions of a known AD locus affected an expression regulatory element. Only one candidate variant emerged (see supplementary information and Table S4 for detailed results), a deep intronic *PINX1* variant predicted to create a novel 5′ splicing site. *PINX1* has been linked to AD in a multi-ethnic study through a gene-based case-control analysis on uncommon (1–5% frequency) and rare coding variants, mainly through missense variants, with only three truncating variants in the last exon [30]. Unfortunately, no RNA was available for our patient.

Rare variants nested burden testing centered on DNM positions

We assessed whether DNMs could highlight regions with rare variant local enrichment in a case-control dataset, using nested-window burden testing across 35 genes, centered on positions of all coding non-synonymous DNMs (after exclusion of the 4 CHIP-related variants). Gathering missense variants with in-frame indels prioritized based on the VEST4 score, with or without truncating variants, no region showed significant enrichment after correction (Table S5 and Table S6).

CNV burden testing in exomes from the same dataset was performed as described in ref. [24] *CCZ1B*, *RSPH10B2* and *SERPINB4* genes overlap regions of segmental duplications and could not be analyzed. The de novo *SNIP1* duplication detected here was partial and only one other partial duplication was detected, in a control, while there was no enrichment in complete duplications or deletions (Table S7).

Inherited variants in 46 EOAD trio exomes

Autosomal recessive analysis. Rare coding bi-allelic variants were extracted from exome data of all 46 negative trios. On average,

Table 2. List of all confirmed exonic and splice region de novo variants.

Trio ID	Trio in Rovelet-Lecrux et al., [7]	Consequence	Gene symbol	Nomenclature (cDNA level)	Nomenclature (protein level)	Identifier of existing variations	CADD	REVEL	Allele frequency in gnomAD v2.1
EFA-362	No	missense	*ZSWIM8*	NM_001242488.2:c.5563A>G	NP_001229417.1:p.(Ile1855Val)	-	-	-	-
EFA-787	No	splice region	*NUP54*	NM_017426.4:c.67+5G>C	-	-	-	-	-
EFA-787	No	splice region	*CNTFR*	NM_147164.3:c.86-10C>T	-	rs573942894	-	-	6.894e-05
EFA-435	No	synonymous	*SSH3*	NM_017857.4:c.1260A>G	NP_060327.3:p.(Ser420Ser)	-	-	-	-
EFA-435	No	missense	*RRP7A*	NM_015703.5:c.211A>C	NP_056518.2:p.(Thr71Pro)	-	20.4	0.098	-
EFA-848	No	missense	*LOC400499*	NM_001370704.1:c.3319G>A	NP_001357633.1:p.(Val1107Ile)	rs1482886742	0.006	-	7.49e-06
EFA-848	No	missense	*SYNJ2*	NM_003898.4:c.3200C>T	NP_003889.1:p.(Thr1067Met)	rs140015862	23.8	0.437	3.987e-06
EFA-667	No	missense	*RREB1*	NM_001003699.4:c.1984G>A	NP_001003699.1:p.(Val662Met)	rs777379568	20.3	0.145	2.45e-05
EFA-667	No	missense	*SLC7A13*	NM_138817.3:c.906C>A	NP_620172.2:p.(Asn302Lys)	-	8.079	0.395	-
EXT-0395	Yes	missense	*CPM*	NM_001874.5:c.352C>T	NP_001865.1:p.(Arg118Trp)	-	25.0	0.589	-
EXT-0395	Yes	missense	*RNF213*	NM_001256071.3:c.5114C>T	NP_001243000.2:p.(Thr1705Met)	rs147868237,COSV60393253	24.6	0.277	3.527e-05
EXT-0941	No	missense	*EME2*	NM_001257370.2:c.1063G>C	NP_001244299.1:p.(Gly355Arg)	-	24.5	-	-
EXT-0941	No	inframe insertion	*TBL1X*	NM_001139467.1:c.342_350dup	NP_001132939,1:p.(Ala115_Ala117dup)				
EXT-1340	No	missense	*IER5*	NM_016545.5:c.136G>A	NP_057629.2:p.(Val46Ile)	-	24.7	0.128	-
EXT-1340	No	missense	*MCTP1*	NM_024717.6:c.326C>T	NP_078993.4:p.(Pro109Leu)	-	22.4	0.211	-
EXT-0222	No	missense	*KCNJ3*	NM_002239.4:c.1214A>G	NP_002230.1:p.(Lys405Arg)	-	22.2	0.221	-
EXT-1289	No	missense	*LPAR6*,***	NM_005767.7:c.614G>A	NP_005758.2:p.(Ser205Asn)	rs780139124	23.0	0.375	3.998e-06
EXT-1289	No	synonymous	*GABRB3*,***	NM_000814.6:c.249C>T	NP_000805.1:p.(Thr83Thr)	rs1476625006	-	-	3.989e-06
EXT-1289	No	missense	*STARD9*	NM_020759.3:c.2909 C > G	NP_065810.2:p.(Ser970Cys)	-	18.61	0.053	-
EXT-1289	No	missense	*TOMM70*,***	NM_014820.5:c.1795 A > G	NP_055635.3:p.(Lys599Glu)	rs1270426552	22.6	0.130	-
EXT-1289	No	frameshift	*TET2*,***	NM_001127208.3:c.1328_1341del	NP_001120680.1:p.(Thr443ArgfsTer7)	-	-	-	-
EXT-1549	No	missense	*ATG2B*	NM_018036.7:c.5716_5717delinsAT	NP_060506.6:p.(Asp1906Ile)	-	33	0.732	-
EXT-1549	No	synonymous	*EHMT2*	NM_001363689.1:c.2232 G > A	NP_001350618.1:p.(Thr744Thr)	rs762523692	-	-	4.456e-05
EXT-0893	No	missense	*CEP112**	NM_001199165.4:c.2018 C > T	NP_001186094.1:p.(Thr673Met)				
EXT-0893	No	splice region	*TACC1*	NM_001352778.2:c.1887+8C > T	-	-	-	-	-
EXT-0848	No	missense	*FASN*	NM_004104.5:c.4874 C > T	NP_004095.4:p.(Ser1625Phe)	-	22.6	0.118	-
EXT-1244	No	missense	*TTN*	NM_001267550.2:c.88820 G > A	NP_001254479.2:p.(Arg29607Gln)	rs755212831	23.7	0.424	4.839e-05
EXT-1244	No	missense	*DDR1*	NM_013994.3:c.2155 C > T	NP_054700.2:p.(Leu719Phe)	-	28.2	0.874	-
EXT-1301	No	missense	*IL27RA*	NM_004843.4:c.1216 T > C	NP_004834.1:p.(Ser406Pro)	-	3.751	0.014	-
EXT-344	Yes	frameshift	*LPIN3*	NM_001301860.2:c.335dup	NP_001288789.1:p.(Leu113SerfsTer15)	rs756955559	-	-	0.0001159
EXT-0804	Yes	missense	*SPEG*	NM_005876.5:c.2929 G > A	NP_005867.3:p.(Val977Met)	rs770098499	25.6	0.592	8.299e-06
EXT-0854	No	missense	*MKI67*	NM_002417.5:c.5194 A > G	NP_002408.3:p.(Thr1732Ala)	rs768117451	6.276	0.036	2.388e-05
EXT-0854	No	synonymous	*TRABD2B*	NM_001194986.2:c.774 A > T	NP_001181915.1:p.(Gly258Gly)	-	-	-	-
EXT-0854	No	splice region	*RAP1GAP2*	NM_015085.5:c.45-20 C > T	-	rs1285394067	-	-	1.228e-05
EXT-186	Yes	splice region	*ANO7*	NM_001370694.2:c.1561+17 C > T	p.?				
EXT-1223	No	missense	*DPYD*	NM_000110.4:c.745 A > G	NP_000101.2:p.(Lys249Glu)	-	25.8	0.782	-
EXT-1223	No	missense	*TMEM131*	NM_015348.2:c.32 G > T	NP_056163.1:p.(Gly11Val)				
EXT-0488	Yes	missense	*SLC25A27*	NM_004277.5:c.811 A > G	NP_004268.3:p.(Lys271Glu)	-	26.8	0.728	-
EXT-0488	Yes	missense	*PHKA2*	NM_000292.3:c.869 G > A	NP_000283.1:p.(Arg290His)	rs186632999	26.2	0.829	1.091e-05
EXT-0687	Yes	missense	*RGS13*	NM_002927.5:c.227 G > A	NP_002918.1:p.(Arg76Gln)	rs1359297515	13.43	0.016	7.985e-06
EXT-0687	Yes	missense	*COL19A1*	NM_001858.6:c.3106 G > A	NP_001849.2:p.(Val1036Ile)	-	22.3	-	-
EXT-1403	No	synonymous	*SPEG*	NM_005876.5:c.9057 C > T	NP_005867.3:p.(His3019His)	rs200656875	-	-	0.0001569
EXT-1403	No	missense	*FAM13A*	NM_014883.4:c.481 C > A	NP_055698.2:p.(Pro161Thr)	-	25.6	0.675	-
EXT-1551	No	inframe deletion	*SPHK2*	NM_001204159.3:c.387_407del	NP_001191088.1:p.(Arg130_Arg136del)	-	-	-	-

Table 2. continued

Trio ID	Trio in Rovelet-Lecrux et al., [7]	Consequence	Gene symbol	Nomenclature (cDNA level)	Nomenclature (protein level)	Identifier of existing variations	CADD	REVEL	Allele frequency in gnomAD v2.1
EXT-1551	No	splice region	*COL4A3*	NM_000091.5:c.3211-15 T > A	-	-	-	-	-
ROU-682	Yes	missense	*VPS35*	NM_018206.6:c.1874T>C	NP_060676.2:p.(Leu625Pro)		29.4	0.746	-
ROU-682	Yes	synonymous	*OXTR*	NM_000916.4:c.714 G > T	NP_000907.2:p.(Ala238Ala)	-	-	-	-
ROU-1331	Yes	in frame delins	*MARK4*	NM_001199867.2:c.947_951delinsAT	NP_001186796.1:p.(Gly316_Glu317delinsAsp)	-	-	-	-
ROU-1331	Yes	missense	*ABCA12*	NM_173076.3:c.2067 C > A	NP_775099.2:p.(Asp689Glu)	-	16.65	0.204	-
ROU-1331	Yes	synonymous	*ACO2*	NM_001098.3:c.1443 C > T	NP_001089.1:p.(Asp481Asp)	rs576516374	-	-	2.785e-05
ROU-1331	Yes	missense	*MYCT1*	NM_025107.2:c.580 C > T	NP_079383.2:p.(Pro194Ser)	rs1034181147	1.859	0.036	3.979e-06
ROU-1322	Yes	synonymous	*ADH7*	NM_001166504.2:c.651 C > T	NP_001159976.1:p.(Val217Val)	rs765427932	-	-	7.979e-06
ROU-1597	No	synonymous	*MICALL1*	NM_033386.4:c.1473 C > T	NP_203744.1:p.(His491His)	rs374402603	-	-	0.0002476
ROU-1750	No	missense	*POP1*	NM_001145860.2:c.1252 A > G	NP_001139332.1:p.(Thr418Ala)	-	21.7	0.150	-

CADD combined annotation dependent depletion score, *REVEL* rare exome variant ensemble learner score.
*post-zygotic variant; **likely clone-specific variant, somatic (suspected CHIP).

each proband carried 0.6 [0–4] homozygous and 2.4 [0–6] compound heterozygous coding or canonical splice site rare variants (Table S8 and Table S9). Six genes showed ≥2 biallelic variants carriers (*MUC4*, $n = 5$ carriers, *SYNE1*, $n = 3$; *HLA-B*, $n = 3$; *TTN*, *FRAS1*, *ALPK3*, $n = 2$ each), most of them were large genes (all>4000 codons apart from *HLA-B* and *ALPK3*), some of them had a known high burden of benign polymorphic variants (e.g. *TTN*, *MUC4*) or repeated sequences in the human genome (e.g. *MUC4*, *HLA-B*) (see also supplementary information).

Two probands carried compound heterozygous rare missense variants in either *ANK3* or *AGRN*, belonging to the AD-GWAS list of genes [10] or to the Aβ network [31], respectively, however without any recurrence among the other trios.

Overall, no compelling recessive candidate was found.

Inherited rare variants in risk factor genes. Among all 46 negative probands, 16 (35%) were *APOE*-ε4 heterozygotes and 2 (4%) homozygotes. Nine patients (20%) carried at least one rare definite risk variant using our clinical classification of AD risk variants [5] (Table S1), all inherited from asymptomatic parents. Two carried an *ABCA7* truncating variant (moderate effect on AD risk) and both probands also exhibited another risk variant: one also carried a *TREM2* R62H variant (modest effect), in an *APOE* 3-3 context, while the other patient was *APOE4* heterozygous. In addition, one proband carried an extremely rare truncating *ABCA1* variant (moderate effect), a CNV-deletion of 3 exons, in the context of an *APOE* 3-4 genotype. Two patients carried a *TREM2* R47H variant (moderate effect), in an *APOE* 3-4 and *APOE* 3-3 context, respectively. Finally, four carried a variant with a modest effect (*TREM2* R62H or *ATP8B4* G395S) (two being APOE 3-4 and two being 3-3), in addition to the above-mentioned *TREM2* R62H – *ABCA7* LOF double variant carrier. Overall, 22/46 (48%) probands carried at least one of these selected risk variants, all inherited from an asymptomatic parent, 6 of them (13%) carried two risk variants.

Genetic risk scores among 16 trio-based genomes. To assess the role of variants identified in GWAS, we computed a genetic risk score (GRS) based on ref. [10], which does not include most of the previously mentioned rare variants. Following Bellenguez et al. deciles, probands GRS were distributed all along GRS decile groups, suggesting that GRS do not explain most of the genetic component of AD etiology in these patients (Figure S1). In addition, GRS did not differ significantly between probands and parental mean (p-value = 0.63). Compared to the GRS of the corresponding parents at the trio level, 4 patients had a higher GRS than both parents, 4 had lower and 8 had a GRS between that of their father and mother. In addition, none of the probands' GRS deviated >2 SD from the theoretical simulated distribution of GRS based on parental SNPs (Figures S2-S17), suggesting that probands did not inherit more risk variants than random transmission, on average.

DISCUSSION

In this study of 49 sporadic EOAD trios, three carried a pathogenic de novo variant in *APP* or *PSEN1*. Among the 46 remaining trios, exome sequencing revealed 4 de novo CNVs and 37 non-synonymous DNMs, four of which were in genes linked to Aβ aggregation or toxicity. Further assessment of DNMs in 16 trios by GS prioritized only one candidate variant.

From the coding DNMs, *VPS35* and *MARK4* variants have previously been assessed in vitro, respectively showing a loss-of-function effect thus increasing Aβ secretion, and an increased phosphorylation of Tau by the Aβ-induced MARK4 kinase [7]. In addition, we identified DNMs in *DDR1* and *SPHK2* here.

DDR1 encodes the Discoidin Domain Receptor Tyrosine Kinase 1 protein, a collagen-activated receptor tyrosine kinase expressed in

the brain and involved in microglial cells and neuroinflammation [28]. Knockdown of *DDR1* by shRNA or the use of pharmacological inhibitors reduced the levels of secreted Aβ42 and P-Tau in vitro and in mice expressing *APP* or *MAPT* pathogenic variants [28, 32, 33]. In addition, in a small placebo-controlled clinical trial, patients with mild–moderate AD who were treated with the tyrosine kinase inhibitor nilotinib that preferentially targets discoidin domain receptors (DDRs) showed reduced CSF Aβ levels and CNS amyloid burden as measured with Positron Emission Tomography (PET) [34].

SPHK2 encodes the primary sphingosine kinase in the brain, responsible for generating sphingosine-1-phosphate (S1P), a key bioactive lipid involved in multiple signaling pathways [35]. Thanks to a nuclear localization sequence (NLS) and a nuclear export sequence (NES) [36], SPHK2 shuttles between the nucleus, where it regulates gene expression, DNA synthesis, and telomere maintenance and the cytoplasm, where it regulates apoptosis, mitochondrial function, and autophagy [37–40]. SPHK2 overexpression in neurons leads to histone acetylation changes and DNA double-strand breaks, contributing to neurotoxicity [41]. Additionally, astrocytes secrete S1P to modulate neuronal processes like migration and synapse formation [42]. SPHK2 activity has been linked to AD through studies on brain tissue, although with conflicting results [43, 44]. Another study reported an alteration of the SPHK2 protein subcellular localization in AD brains, with a shift of full length SPHK2 from cytosol to the nucleus and an accumulation of the cleaved SPHK2 in the nucleus [45]. In cellular and animal models, lowering neuronal SPHK2 expression decreases Aβ secretion, while its overexpression increases the levels of secreted Aβ [29, 43]. SPHK2 acts by regulating the β-cleavage of APP through direct interaction of S1P with BACE1 protein, a limiting enzyme for amyloidogenic processing of APP that results in the production of Aβ peptides. Additionally, nuclear SPHK2/S1P signaling regulates APOE production and Aβ uptake in astrocytes [46]. Interestingly, the p.(Arg130_Arg136del) DNM detected in a sporadic EOAD patient here is located next to the NLS. In a post-hoc analysis focusing on the 9 ± 7 codons that are predicted to act as an NLS based on Uniprot, there was a local enrichment – albeit non-significant after multiple testing correction – of ultra-rare (frequency $<10^{-4}$) non-synonymous, predicted damaging (VEST4 score >0.30) variants in this restricted region, with 5 variants among 10,436 cases and none among 10,623 controls ($p = 0.03$) (filters applied: use of genotypes, GQ for heterozygous variants >90, individuals with depth<10 set as missing). This suggests that this region might actually be functional regarding AD, but that the power of our study is still too low to detect such local enrichments of carriers in very circumscribed biologically-relevant regions.

Here, we used our nation-wide recruitment to include sporadic EOAD trios. Despite significant effort in systematically proposing genetic analyses to parents, we could only recruit 37 novel trios. We are not aware of any other study on novel EOAD trios and could thus not replicate our findings in other trios. Not surprisingly, no gene was recurrently hit by a non-synonymous DNM, consistent with findings in Amyotrophic Lateral Sclerosis (reviewed in ref. [3]) or autism spectrum disorders, which taught us that hundreds, if not thousands of trios are required to expect finding recurrences, in absence of a major recurrently hit gene [47]. This suggests that DNMs, beyond known autosomal dominant genes, likely play a minor role in sporadic EOAD. However, some DNMs might have a modest impact on disease etiology in the carriers and some others may lead to AD determinism with higher impact, but with an extreme heterogeneity at the population level. To circumvent this lack or recurrence of DNMs at the gene level, and because AD does not face the same transmission constraint as neurodevelopmental disorders, we assessed the burden of rare variants in genes hit by DNMs. However, our burden testing, even in such a large case-control dataset, did not reveal significant enrichment in DNM-targeted transcripts. Burden tests are more powerful (i) in case of homogeneous effects of selected variants (either all increasing risk or all decreasing risk) and (ii) in case of a loss-of-function effect driving the signal, while variance-based tests or its derivates (e.g., SKAT, SKAT-O) are very sensitive to subtle population stratification issues. Here, the DNMs of interest may have non-haploinsufficiency effects, first, and the putative effect may well be very specific to a small region. Given the rarity of such variants, even using the large ADES-ADSP dataset did not allow a firm conclusion on these genes.

From a clinical perspective, trio-based exome or genome sequencing appears unnecessary for most sporadic EOAD cases. Proband-only testing (panel, exome or genome sequencing focusing on dementia-related genes) appears more cost-effective. Indeed, the de novo status of the pathogenic variants here was not required to classify the variants as (likely) pathogenic. However, the de novo status is obviously of tremendous importance for genetic counseling, but the inheritance pattern can be assessed by a targeted analysis after the identification of the genetic cause in the proband. Importantly, the low frequency of pathogenic DNMs and the high rate (48%) of inherited risk variants, even in carefully selected sporadic cases, supports a complex genetic etiology for most cases, consistent with previous findings in sporadic EOAD or in EOAD overall [5, 6, 48], and to which some DNMs might erratically add to the genetic etiology. The identified inherited risk variants may only represent the tip of the iceberg, as it is likely that such young patients may exhibit other unknown genetic determinants, also knowing that no strong environmental factor has been identified in EOAD yet. Common variants, apart from the *APOE*-ε4 alleles, may not explain a significant part of EOAD risk, even when gathered in a GRS. Although GRS without *APOE* may explain a very small proportion of the variance of AD in population studies (less than 2.4% in ref. [49]), we may hypothesize that the extreme selection of our cases here (very young onset AD and negative family history, no pathogenic variant) could have led to select individuals with a very unlikely distribution of possible parental transmission of AD risk alleles among GWAS-defined common variants with low effect size. Here, our analyses suggest that variants included in the GRS except *APOE* do not contribute significantly to the disease etiology, as scores of the patients were not significantly higher than their parents and the transmission was not biased towards risk alleles.

Among the variant types not assessed here, short tandem repeats should certainly be assessed as candidates, as well as a more systematic study of structural variants. Unfortunately, the use of 50-bp reads GS did not allow any accurate analysis of structural variants in our study. The use of short-read GS, with longer read depths than here (100 or 150-bp reads) has recently been applied to a large dataset of AD and controls, showing significant enrichment in burdens of deletions or duplications, and rare structural variants affecting known AD genes [50]. In addition, using novel long-read sequencing techniques will be of great interest in the field of AD research, by enabling the detection of virtually any kind of structural variants and repeats [51]. Applying these novel technologies to EOAD trios represents a promising perspective for our study, along with large-scale case-control studies.

In conclusion, assessing the de novo paradigm confirmed that some *APP* or *PSEN1* pathogenic variants can explain a small proportion of sporadic EOAD, and unveiled *DDR1* and *SPHK2* as candidate genes that should be further assessed in future studies. Despite this, exome and short-read genome sequencing did not allow the identification of a high number of candidate variants regarding AD genetic determinism. Altogether, these results suggest that sporadic EOAD might be linked to the association of multiple, mostly inherited variants rather than monogenic

variants, in a large proportion of patients, although part of this genetic component still remains to be identified, as GWAS-related genetic risk did not explain it and moderate-to-strong risk variants were found in <50% of the probands. Novel long-read techniques should be applied to fully assess the de novo paradigm in sporadic EOAD, in parallel to the use of similar techniques in large case-control studies.

DATA AVAILABILITY

Summary statistics are available upon request. Access to deidentified raw data of the ADES study can be requested using the Alzheimer Genetics Hub. Authors do not own data from ADSP, which can be requested through their respective websites following specific procedures.

REFERENCES

1. Vissers LE, de Ligt J, Gilissen C, Janssen I, Steehouwer M, de Vries P, et al. A de novo paradigm for mental retardation. Nat Genet. 2010;42:1109–12.
2. Deciphering Developmental Disorders S. Prevalence and architecture of de novo mutations in developmental disorders. Nature. 2017;542:433–8.
3. Nicolas G, Veltman JA. The role of de novo mutations in adult-onset neurodegenerative disorders. Acta neuropathol. 2019;137:183–207.
4. Lanoiselee HM, Nicolas G, Wallon D, Rovelet-Lecrux A, Lacour M, Rousseau S, et al. APP, PSEN1, and PSEN2 mutations in early-onset alzheimer disease: a genetic screening study of familial and sporadic cases. PLoS Med. 2017;14:e1002270.
5. Nicolas G, Zarea A, Lacour M, Quenez O, Rousseau S, Richard AC, et al. Assessment of Mendelian and risk-factor genes in alzheimer disease: a prospective nationwide clinical utility study and recommendations for genetic screening. Genet Med. 2024;26:101082.
6. Lacour M, Quenez O, Rovelet-Lecrux A, Salomon B, Rousseau S, Richard AC, et al. Causative mutations and genetic risk factors in sporadic early onset alzheimer's disease before 51 years. J Alzheimers Dis. 2019;71:227–43.
7. Rovelet-Lecrux A, Charbonnier C, Wallon D, Nicolas G, Seaman MN, Pottier C, et al. De novo deleterious genetic variations target a biological network centered on abeta peptide in early-onset Alzheimer disease. Mol psychiatry. 2015;20:1046–56.
8. Gatz M, Reynolds CA, Fratiglioni L, Johansson B, Mortimer JA, Berg S, et al. Role of genes and environments for explaining alzheimer disease. Arch Gen psychiatry. 2006;63:168–74.
9. Nicolas G. Recent advances in alzheimer disease genetics. Curr Opin Neurol. 2024;37:154–65.
10. Bellenguez C, Kucukali F, Jansen IE, Kleineidam L, Moreno-Grau S, Amin N, et al. New insights into the genetic etiology of alzheimer's disease and related dementias. Nat Genet. 2022;54:412–36.
11. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, *Vienna, Austria* 2021. https://www.r-project.org/.
12. Dubois B, Feldman HH, Jacova C, Hampel H, Molinuevo JL, Blennow K, et al. Advancing research diagnostic criteria for alzheimer's disease: the IWG-2 criteria. Lancet Neurol. 2014;13:614–29.
13. Constantinides VC, Paraskevas GP, Boufidou F, Bourbouli M, Pyrgelis ES, Stefanis L, et al. CSF abeta42 and abeta42/Abeta40 ratio in alzheimer's disease and frontotemporal dementias. Diagnostics. 2023;13:783.
14. Nicolas G, Wallon D, Charbonnier C, Quenez O, Rousseau S, Richard AC, et al. Screening of dementia genes by whole-exome sequencing in early-onset alzheimer disease: input and lessons. Eur J Hum Genet. 2016;24:710–6.
15. Backenroth D, Homsy J, Murillo LR, Glessner J, Lin E, Brueckner M, et al. CANOES: detecting rare copy number variants from whole exome sequencing data. Nucleic acids Res. 2014;42:e97.
16. Quenez O, Cassinari K, Coutant S, Lecoquierre F, Le Guennec K, Rousseau S, et al. Detection of copy-number variations from NGS data using read depth information: a diagnostic performance evaluation. Eur J Hum Genet. 2021;29:99–109.
17. Pagel KA, Kim R, Moad K, Busby B, Zheng L, Tokheim C, et al. Integrated informatics analysis of cancer-related variants. JCO Clin Cancer Inform. 2020;4:310–7.
18. Carter H, Douville C, Stenson PD, Cooper DN, Karchin R. Identifying mendelian disease genes with the variant effect scoring tool. BMC Genomics. 2013;14:S3 https://doi.org/10.1186/1471-2164-14-S3-S3.
19. Ioannidis NM, Rothstein JH, Pejaver V, Middha S, McDonnell SK, Baheti S, et al. REVEL: an ensemble method for predicting the pathogenicity of rare missense variants. Am J Hum Genet. 2016;99:877–85.
20. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, et al. The ensembl variant effect predictor. Genome Biol. 2016;17:122.
21. Zhang X, Wakeling M, Ware J, Whiffin N. Annotating high-impact 5'untranslated region variants with the UTRannotator. Bioinformatics. 2021;37:1171–3.
22. Jaganathan K, Kyriazopoulou Panagiotopoulou S, McRae JF, Darbandi SF, Knowles D, Li YI, et al. Predicting splicing from primary sequence with deep learning. Cell. 2019;176:535–48 e24.
23. Holstege H, Hulsman M, Charbonnier C, Grenier-Boley B, Quenez O, Grozeva D, et al. Exome sequencing identifies rare damaging variants in ATP8B4 and ABCA1 as risk factors for alzheimer's disease. Nat Genet. 2022;54:1786–94.
24. Quenez O, Schramm C, Cassinari K, Nicolas A, Groeneveld J, Huguet G, et al. Extremely rare CNVs contributing to alzheimer disease risk: a case-control association analysis of exome sequencing data from 22,319 individuals. medRxiv 2024.10.28.24314051 [Preprint]. 2024. Available from: https://www.medrxiv.org/content/10.1101/2024.10.28.24314051v1.
25. Lindquist SG, Nielsen JE, Stokholm J, Schwartz M, Batbayli M, Ballegaard M, et al. Atypical early-onset alzheimer's disease caused by the Iranian APP mutation. J Neurological Sci. 2008;268:124–30.
26. Pasalar P, Najmabadi H, Noorian AR, Moghimi B, Jannati A, Soltanzadeh A, et al. An iranian family with alzheimer's disease caused by a novel APP mutation (Thr714Ala). Neurology. 2002;58:1574–5.
27. Zekanowski C, Styczynska M, Peplonska B, Gabryelewicz T, Religa D, Ilkowski J, et al. Mutations in presenilin 1, presenilin 2 and amyloid precursor protein genes in patients with early-onset alzheimer's disease in Poland. Exp Neurol. 2003;184:991–6.
28. Hebron M, Peyton M, Liu X, Gao X, Wang R, Lonskaya I, et al. Discoidin domain receptor inhibition reduces neuropathology and attenuates inflammation in neurodegeneration models. J Neuroimmunol. 2017;311:1–9.
29. Lei M, Teo JD, Song H, McEwen HP, Yup Lee J, Couttas TA, et al. sphingosine kinase 2 potentiates amyloid deposition but protects against hippocampal volume loss and demyelination in a mouse model of alzheimer's disease. J Neurosci. 2019;39:9645–59.
30. Tosto G, Vardarajan B, Sariya S, Brickman AM, Andrews H, Manly JJ, et al. Association of variants in PINX1 and TREM2 with late-onset alzheimer disease. JAMA Neurol. 2019;76:942–8.
31. Campion D, Pottier C, Nicolas G, Le Guennec K, Rovelet-Lecrux A. Alzheimer disease: modeling an abeta-centered biological network. Mol Psychiatry. 2016;21:861–71.
32. Fowler AJ, Hebron M, Balaraman K, Shi W, Missner AA, Greenzaid JD, et al. Discoidin domain receptor 1 is a therapeutic target for neurodegenerative diseases. Hum Mol Genet. 2020;29:2882–98.
33. Hebron ML, Javidnia M, Moussa CE. Tau clearance improves astrocytic function and brain glutamate-glutamine cycle. J Neurol Sci. 2018;391:90–9.
34. Turner RS, Hebron ML, Lawler A, Mundel EE, Yusuf N, Starr JN, et al. Nilotinib effects on safety, tolerability, and biomarkers in alzheimer's disease. Ann Neurol. 2020;88:183–94.
35. Spiegel S, Milstien S. Sphingosine-1-phosphate: an enigmatic signalling lipid. Nat Rev Mol Cell Biol. 2003;4:397–407.
36. Ding G, Sonoda H, Yu H, Kajimoto T, Goparaju SK, Jahangeer S, et al. Protein kinase D-mediated phosphorylation and nuclear export of sphingosine kinase 2. J Biol Chem. 2007;282:27493–502.
37. Liu H, Toman RE, Goparaju SK, Maceyka M, Nava VE, Sankala H, et al. Sphingosine kinase type 2 is a putative BH3-only protein that induces apoptosis. J Biol Chem. 2003;278:40330–6.
38. Strub GM, Paillard M, Liang J, Gomez L, Allegood JC, Hait NC, et al. Sphingosine-1-phosphate produced by sphingosine kinase 2 in mitochondria interacts with prohibitin 2 to regulate complex IV assembly and respiration. FASEB J. 2011;25:600–12.
39. Song DD, Zhang TT, Chen JL, Xia YF, Qin ZH, Waeber C, et al. Sphingosine kinase 2 activates autophagy and protects neurons against ischemic injury through interaction with Bcl-2 via its putative BH3 domain. Cell Death Dis. 2017;8:e2912.
40. Hait NC, Allegood J, Maceyka M, Strub GM, Harikumar KB, Singh SK, et al. Regulation of histone acetylation in the nucleus by sphingosine-1-phosphate. Science. 2009;325:1254–7.
41. Moruno-Manchon JF, Uzor NE, Blasco-Conesa MP, Mannuru S, Putluri N, Furr-Stimming EE, et al. Inhibiting sphingosine kinase 2 mitigates mutant huntingtin-induced neurodegeneration in neuron models of Huntington disease. Hum Mol Genet. 2017;26:1305–17.
42. Choi JW, Chun J. Lysophospholipids and their receptors in the central nervous system. Biochim Biophys Acta. 2013;1831:20–32.
43. Takasugi N, Sasaki T, Suzuki K, Osawa S, Isshiki H, Hori Y, et al. BACE1 activity is modulated by cell-associated sphingosine-1-phosphate. J Neurosci. 2011;31:6850–7.
44. Couttas TA, Kain N, Daniels B, Lim XY, Shepherd C, Kril J, et al. Loss of the neuroprotective factor sphingosine 1-phosphate early in alzheimer's disease pathogenesis. Acta Neuropathol Commun. 2014;2:9.
45. Dominguez G, Maddelein ML, Pucelle M, Nicaise Y, Maurage CA, Duyckaerts C, et al. Neuronal sphingosine kinase 2 subcellular localization is altered in alzheimer's disease brain. Acta Neuropathol Commun. 2018;6:25.

46. Komai M, Noda Y, Ikeda A, Kaneshiro N, Kamikubo Y, Sakurai T, et al. Nuclear SphK2/S1P signaling is a key regulator of ApoE production and abeta uptake in astrocytes. J Lipid Res. 2024;65:100510.
47. Iossifov I, O'Roak BJ, Sanders SJ, Ronemus M, Krumm N, Levy D, et al. The contribution of de novo coding mutations to autism spectrum disorder. Nature. 2014;515:216–21.
48. Schramm C, Charbonnier C, Zarea A, Lacour M, Wallon D, collaborators, et al. Penetrance estimation of Alzheimer disease in SORL1 loss-of-function variant carriers using a family-based strategy and stratification by APOE genotypes. Genome Med. 2022;14:69.
49. Karlsson IK, Escott-Price V, Gatz M, Hardy J, Pedersen NL, Shoai M, et al. Measuring heritable contributions to alzheimer's disease: polygenic risk score analysis with twins. Brain Commun. 2022;4:fcab308.
50. Wang H, Dombroski BA, Cheng PL, Tucci A, Si YQ, Farrell JJ, et al. Structural variation detection and association analysis of whole-genome-sequence data from 16,543 Alzheimer's disease sequencing project subjects. Alzheimers Dement. 2025;21:e70277.
51. Mokveld T, Dolzhenko E, Dashnow H, Nicholas TJ, Sasani T, van der Sanden B, et al. TRGT-denovo: accurate detection of de novo tandem repeat mutations. bioRxiv : the preprint server for biology [Preprint]. 2024. Available from: https://www.biorxiv.org/content/10.1101/2024.07.16.600745v1.

ACKNOWLEDGEMENTS

We thank the CEREBRO cluster at the University of Rouen Normandie, the high-performance computing service at the University of Lille and the Cartesius supercomputer, which is embedded in the Dutch national e-infrastructure with the support of SURF Cooperative. Full ADES consortium acknowledgements and funding sources are listed in Holstege al., 2022. 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. This work involves a CHU Rouen – CNRGH collaboration.

AUTHOR CONTRIBUTIONS

GN conceptualized and designed the study. All authors had a major role in the acquisition of data. GN, KC, FL, OQ, CC, CS had a major role in data analysis and statistics. GN, AZ, KC, FL, OQ, CC, CS, and DW interpreted the data. GN drafted the manuscript, GN, AZ, KC, FL, OQ, CC, CS, ARL, MLe and DW contributed to writing the manuscript. GN, AZ, KC, FL, OQ, CC, CS, ML, and DW revised the manuscript.

FUNDING

This study received funding from Fondation Alzheimer (ECASCAD study, to G.N.), CNRMAJ, the Dutch Research Council (NWO, 918-15-667 to J.A.V.), and benefited from support of the France Génomique National infrastructure, funded as part of the «Investissements d'Avenir» program managed by the Agence Nationale pour la Recherche (contract ANR-10-INBS-09). This research also used 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. Cartesius computing hours were granted to Henne Holstege by the Dutch Research Council ('100plus': project# vuh15226, 15318, 17232, 2020.030; 'Role of VNTRs in AD'; project# 2022.028, 'Alzheimer's Genetics Hub' project# 2022.031 and 2024:036).The CEREBRO computing cluster received fundings from Région Normandie, European Regional Devlopement Fund and the French Ministry of Higher Education, Research and Industry (MESRI).

COMPETING INTERESTS

The author declare no competing interests.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

The analysis of sequencing data was approved by the CERDE ethics committee from the Rouen University Hospital (notification 2019-055). Participants all provided written consent for genetic analyses. Participants with genome sequencing data provided written consent to participate to the RBM02-59 study, approved by the CPP Ile de France II ethics committee.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41380-026-03665-6.

Correspondence and requests for materials should be addressed to Gaël. Nicolas.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2026

[1]Univ Rouen Normandie, Normandie Univ, Inserm U1245 and CHU Rouen, Department of Neurology and CNRMAJ, F-76000 Rouen, France. [2]Univ Rouen Normandie, Normandie Univ, Inserm U1245 and CHU Rouen, Department of Genetics and CNRMAJ, F-76000 Rouen, France. [3]Univ Rouen Normandie, Normandie Univ, Inserm U1245 and CHU Rouen, Department of Biostatistics and CNRMAJ, F-76000 Rouen, France. [4]Univ Rouen Normandie, Normandie Univ, Inserm U1245, F-76000 Rouen, France. [5]Université Paris-Saclay, CEA, Centre National de Recherche en Génomique Humaine (CNRGH), 91057 Evry, France. [6]Department of Human Genetics, Radboud University Medical Center, PO Box 9101, 6500 HB Nijmegen, The Netherlands. [7]Radboud Institute for Molecular Life Sciences, Radboud University Medical Center, Nijmegen, The Netherlands. [8]Institute of Genetics and Cancer, College of Medicine and Veterinary Medicine, University of Edinburgh, Edinburgh, UK. [9]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. [10]Department of Neurology, II B Sant Pau, Hospital de la Santa Creu i Sant Pau, Universitat Autònoma de Barcelona, Barcelona, Spain. [11]Biomedical Research Networking Center on Neurodegenerative Diseases, National Institute of Health Carlos III, Madrid, Spain. [12]Reta Lila Weston Research Laboratories, Department of Molecular Neuroscience, University College London Institute of Neurology, London, UK. [13]Genomics of Neurodegenerative Diseases and Aging, Human Genetics, Vrije Universiteit Amsterdam, Amsterdam UMC location VUmc, Amsterdam, the Netherlands. [14]Alzheimer Center Amsterdam, Neurology, Vrije Universiteit Amsterdam, Amsterdam UMC location VUmc, Amsterdam, the Netherlands. [15]Amsterdam Neuroscience, Neurodegeneration, Amsterdam, the Netherlands. [16]Delft Bioinformatics Lab, Delft University of Technology, Delft, the Netherlands. [17]Medical Research Council Prion Unit at University College London, University College London Institute of Prion Diseases, London, UK. [18]German Center for Neurodegenerative Diseases, Bonn, Germany. [19]Division of Neurogenetics and Molecular Psychiatry, Department of Psychiatry and Psychotherapy, Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany. [20]Department of Psychiatry and Glenn Biggs Institute for Alzheimer's and Neurodegenerative Diseases, San Antonio, TX, USA. [21]Cologne Excellence Cluster on Cellular Stress Responses in Aging-Associated Disease (CECAD), University of Cologne, Cologne, Germany. [22]Centre for Neuropsychiatric Genetics and Genomics, Division of Psychological Medicine and Clinical Neuroscience, School of Medicine, Cardiff University, Cardiff, UK. [23]Department of Neurology, Erasmus Medical Centre,

Rotterdam, the Netherlands. [24]Department of Cognitive Disorders and Old Age Psychiatry, University Hospital Bonn, Medical Faculty, Bonn, Germany. [25]Moondance Dementia Research Laboratory, Cardiff University, Cardiff, UK. [26]AP-HP Sorbonne Université, Hôpital Pitié-Salpêtrière, Department of Neurology, Institute of Memory and Alzheimer's Disease, Paris, France. [27]Nantes Université, CHU Nantes, INSERM, Centre Mémoire Ressource et Recherche (CMRR), Department of Neurology, CIC 1413, F-44000 Nantes, France. [28]McGill University Research Center for Studies on aging, Douglas Mental Health University Institute, McGill University Department of Neurology and Neurosurgery, Faculty of Medicine, McGill University, Montreal, Canada. [29]INSERM, U1077, CHU de Caen, GIP Cyceron, Neuropsychologie et Imagerie de la Mémoire Humaine, and Department of Neurology, CHU Caen-Normandie, F-14000 Caen, France. [30]Univ Lille, CHU Lille, Inserm 1172, Memory center, CNRMAJ, LiCEND, Labex DistAlz, 59000 Lille, France. [31]CMRR, CRMR Neurogénétique, Service de Neurologie, CHU d'ANGERS, Angers, France. [32]Memory Ressources Research Center, Department of Neurology, University Hospital of Montpellier, Montpellier, France. [33]Institute of Neurophysiopathology UMR 7051 Aix Marseille Université & Assistance Publique de Marseille, Marseille, France. [34]Service de Génétique, CHU Poitiers, Poitiers Cedex, France. [35]Sorbonne Université, INSERM U1127, CNRS 7235, Institut du Cerveau - ICM, Paris, France; AP-HP Sorbonne Université, Pitié-Salpêtrière Hospital, Department of Neurology, Institute of Memory and Alzheimer's Disease, Paris, France. [36]Department of Neurology, Caen University Hospital, Caen, France; Normandie UNIV, UNICAEN, PSL Research University, EPHE, INSERM, CHU de Caen, Neuropsychologie et Imagerie de la Mémoire Humaine, Caen, France. [37]Département des Maladies du Système Nerveux, Institut de la mémoire et de la maladie d'Alzheimer (IM2A), Hôpital de la Pitié-Salpêtrière, AP-HP, Paris, France. [38]Service de génétique clinique, CLAD Ouest, CHU Rennes, Hôpital Sud, Rennes, France. [39]Department of Neurology of Memory and Language, CMRR-Paris Sainte Anne, GHU Paris Psychiatrie & Neurosciences, Hôpital Sainte Anne, F-75014, Paris, France; Université Paris-Cité, F-75006 Paris, France. [40]Université Paris-Saclay, BioMaps, Service Hospitalier Frédéric Joliot CEA, CNRS, Inserm, F-91401 Orsay, France. [41]Department of Neurology, CHU Nantes, 44093 Nantes, France. [42]Université de Bourgogne Europe, CHU Dijon Bourgogne, Centre de Génétique, Centre de Référence Maladies Rares Neurogène, INSERM, CTM UMR 1231, GAD, 21000 Dijon, France. [43]These authors contributed equally: Aline Zarea, Kevin Cassinari, François Lecoquierre, Olivier Quenez, Camille Charbonnier.
✉email: gaelnicolas@hotmail.com