ABCA7 rare variants and Alzheimer disease risk

Kilan Le Guennec, MSc
Gaël Nicolas, MD, PhD
Olivier Quenez, MSc
Camille Charbonnier, PhD
David Wallon, MD, PhD
Céline Bellenguez, PhD
Benjamin Grenier-Boley, MSc
Stéphane Rousseau, MSc
Anne-Claire Richard, MSc
Anne Rovelet-Lecrux, PhD
Delphine Bacq, MSc
Jean-Guillaume Garnier, MSc
Robert Olaso, PhD
Anne Boland, PharmD, PhD
Vincent Meyer, PhD
Jean-François Deleuze, PhD
Philippe Amouyel, MD, PhD
Hans Markus Munter, PhD
Guillaume Bourque, MSc
Mark Lathrop, PhD
Thierry Frebourg, MD, PhD
Richard Redon, MD, PhD
Luc Letenneur, PhD
Jean-François Dartigues, MD, PhD
Florence Pasquier, MD, PhD
Adeline Rollin-Sillaire, MD
Emmanuelle Génin, PhD
Jean-Charles Lambert, PhD
Didier Hannequin, MD, PhD

Author list continued on next page

ABSTRACT

Objective: To study the association between *ABCA7* rare coding variants and Alzheimer disease (AD) in a case-control setting.

Methods: We conducted a whole exome analysis among 484 French patients with early-onset AD and 590 ethnically matched controls.

Results: After collapsing rare variants (minor allele frequency ≤1%), we detected an enrichment of *ABCA7* loss of function (LOF) and predicted damaging missense variants in cases (odds ratio [OR] 3.40, 95% confidence interval [CI] 1.68–7.35, $p = 0.0002$). Performing a meta-analysis with previously published data, we found that in a combined sample of 1,256 patients and 1,347 controls from France and Belgium, the OR was 2.81 (95% CI 1.89–4.20, $p = 3.60 \times 10^{-7}$).

Conclusions: These results confirm that *ABCA7* LOF variants are enriched in patients with AD and extend this finding to predicted damaging missense variants. ***Neurology*® 2016;86:2134–2137**

GLOSSARY

AD = Alzheimer disease; **CI** = confidence interval; **EOAD** = early-onset Alzheimer disease; **LOF** = loss of function; **OR** = odds ratio; **MAF** = minor allele frequency; **WES** = whole-exome sequencing.

In Alzheimer disease (AD), shifting from a common disease–common variant hypothesis to a common disease–rare variant paradigm has allowed to identify rare variants in several genes, including *TREM2*, *APP*, *UNC5C*, and *SORL1*, that affect AD risk.[1–7] An association with *PLD3* rare variants was also suggested[3] but not replicated in subsequent studies.[8–12] Recently, the role of loss of function (LOF) variants in *ABCA7*, a gene encoding an ATP-binding cassette transporter involved in lipid transport, has been highlighted in 2 reports.[13,14] First, using imputation of the whole-genome sequences of ~2,600 Icelanders into ~3,400 patients with AD and ~151,000 population controls, Steinberg et al.[13] reported an association between 8 rare LOF variants and AD risk. They replicated this finding by genotyping additional patients and controls in datasets from Europe and the United states (combined odds ratio [OR] 2.3, $p = 6.8 \times 10^{-15}$). Second, using targeted resequencing of a panel of candidate genes, Cuyvers et al.[14] identified an increase of *ABCA7* LOF variants in 772 Flanders-Belgian patients compared to 757 geographically matched controls (relative risk 4.03, $p = 0.0002$). Analyzing our whole-exome sequencing (WES) data in 484 unrelated French patients with early-onset AD (EOAD) and 590 ethnically matched controls, we found that besides *SORL1*,[7] *ABCA7* was among the top hits for enrichment in rare (minor allele frequency [MAF] <1%) predicted damaging variants. We present the results of the *ABCA7* study.

From INSERM (K.L.G., G.N., O.Q., C.C., D.W., S.R., A.C.R., A.R.-L., T.F., D.H., D.C.), U1079, IRIB, University of Rouen, Normandy University; Normandy Centre for Genomic Medicine and Personalized Medicine (K.L.G., G.N., O.Q., C.C., D.W., S.R., A.-C.R., A.R.-L., T.F., D.H., D.C.), Rouen; Department of Genetics (G.N., T.F., D.H.), CNR-MAJ (G.N., O.Q., C.C., D.W., S.R., A.-C.R., F.P., A.R.-S., D.H., D.C.), and Department of Neurology (D.W., D.H.), Rouen University Hospital; INSERM (C.B., B.G.-B., P.A., J.-C.L.), U1167, Lille; Institut Pasteur de Lille (C.B., B.G.-B., P.A., J.-C.L.); Université Lille-Nord de France (C.B., B.G.-B., P.A., J.-C.L.); Centre National de Génotypage (D.B., J.-G.G., R.O., A.B., V.M., J.-F.Deleuze.), Institut de Génomique, CEA, Evry; Fondation Jean Dausset (J.-F.Deleuze.), Centre d'Etudes du Polymorphisme Humain, Paris, France; McGill University and Génome Québec Innovation Centre (H.M.M., G.B., M.L.), Montréal, Canada; INSERM (R.R.), UMR 1087, l'Institut du Thorax, CHU Nantes; CNRS (R.R.), UMR 6291, Université de Nantes; INSERM (L.L., J.-F.Dartigues.), U897, Bordeaux; University of Bordeaux (L.L., J.-F.Dartigues.); Department of Neurology (F.P., A.R.S.), Lille University Hospital; INSERM (E.G.), UMR1078, CHU Brest, Université Bretagne Occidentale, Brest; and Department of Research (D.C.), Rouvray Psychiatric Hospital, Sotteville-lès-Rouen, France.

Coinvestigators are listed on the *Neurology*® Web site at Neurology.org.

Go to Neurology.org for full disclosures. Funding information and disclosures deemed relevant by the authors, if any, are provided at the end of the article.

Dominique Campion, MD, PhD
On behalf of the CNR-MAJ collaborators

Correspondence to Dr. Campion: dominique.campion@univ-rouen.fr

Editorial, page 2118

See page 2126

Supplemental data at Neurology.org

METHODS **Patients.** Patient ascertainment is described in detail in reference 7. Briefly, we included 484 unrelated patients with EOAD (age at onset ≤65 years) from French ancestry recruited by the French National CNR-MAJ consortium. All diagnoses fulfilled the National Institute of Neurological and Communicative Disorders and Stroke–Alzheimer's Disease and Related Disorders Association criteria.[15] The clinical examination included personal medical and family history assessment, neurologic examination, neuropsychological assessment, and neuroimaging. In addition, CSF biomarkers indicative of AD were available for 249 patients.[16] Patients with CSF biomarkers not consistent with an AD diagnostic (n = 14) were excluded.

Patients with *PSEN1*, *PSEN2*, or *APP* mutation (detected either by Sanger sequencing or by WES), *APP* duplication (detected by quantitative multiplex PCR of short fluorescent fragments),[17] or *C9ORF72* expansion (detected following the protocol described by DeJesus-Hernandez et al.[18]) were excluded. The number of *APOE* ε4–positive patients was 256 (53%).

Controls. A total of 595 controls of French ancestry were recruited. A total of 396 had an age at inclusion of ≥55 years and had normal Mini-Mental State Examination score, according to age and education level. The remaining controls were healthy blood donors. The mean age at inclusion (±SD) was 66.4 ± 17.5 years. The number of *APOE* ε4–positive controls was 138 (23%).

Standard protocol approvals, registrations, and patient consents. All patients and controls gave informed, written consent for genetic analyses. This study was approved by our ethics committee.

WES. As previously described,[7] exomes were captured using Agilent SureSelect Human All Exon kits (Agilent Technologies, Santa Clara, CA). Final libraries were sequenced on a HiSeq2000 or 2500 (Illumina, San Diego, CA) with paired ends, 76 or 100 bp reads. We processed all exome samples following GATK 3.3–0 Best Practices recommendations.[19] We used BWA 0.7.5a to map reads to the 1,000 Genomes GRCh37 build.[20] Duplicate reads were flagged by using Picard Tools 1.101 (http://picard.sourceforge.net). We applied GATK for short insertion and deletions (indels) realignment, base quality score recalibration, and single nucleotide variants and indels discovery using the Haplotype Caller across all samples simultaneously. The joint variant calling file was annotated with RefGene gene regions, variant effects, and functional effect prediction tools using Annovar.[21] We then extracted high-quality (defined as variants with a variant quality score VQSLOD above −2) exonic and canonical splice site variants (located ±2 bp around each coding exon) with a MAF of less than 1% in our whole dataset within each RefGene gene region.

We filtered genotypes according to the following criteria: genotype read depth had to be above 6 as well as genotype quality above 70. Nonsense, frameshift indels, and canonical splice site variants were classified as LOF. Missense variants were considered as strictly damaging if they were simultaneously predicted damaging by the following 3 software programs: Polyphen2 HumDiv, Mutation Taster, and SIFT.[21]

Sanger sequencing. *ABCA7* rare variants included in this study were confirmed by Sanger sequencing.

Quality checks. As previously described,[7] we used recommended quality control steps for genetic case-control association studies to analyze our data.[22] Most checks were carried out with PLINK 1.9 (https://www.cog-genomics.org/plink2). We processed all individuals in the sample through the following steps: (1) verifying concordant sex information using Plink sex check,[23] (2) discarding contaminated samples identified as such by significantly high heterozygosity rates and freemix contamination scores provided by the VerifyBamID software,[24] and (3) discarding samples when Plink pi_hat relatedness estimation exceeded 18.5%. Out of 485 initially included cases and 595 controls, 484 cases and 590 controls passed these quality checks. We did not detect any individual of divergent ancestry with Plink neighbor function. To confirm ethnic match between cases and controls, we used principal component analysis on common variants (MAF >5%) after exclusion of long-range linkage disequilibrium regions and variant pruning on linkage disequilibrium ($r^2 > 0.2$).

We filtered out variants that (1) were missing in more than 5% of individuals, (2) showed a significant deviation from Hardy-Weinberg equilibrium, or (3) presented significantly different missing call rates between cases and controls, as confirmed by Plink test missing at a threshold of 1.10^{-6}. None of these filters discarded any of the *ABCA7* variants with a MAF below 1%.

Finally, the following 2 filters were applied locally to each RefGene coding region: (1) samples that missed information for more than 50% of variants in the given region were excluded from analysis and (2) variants that remained missing in more than 5% of remaining individuals were excluded from analysis. None of these filters affected the number of samples and variants under study within the *ABCA7* gene.

Statistics. We imported all rare (MAF <1% on the whole case-control sample) LOF and strictly damaging missense *ABCA7* into Variant Association Tools.[25] We computed unadjusted gene-level collapsing tests using the CFisher function grouped by RefGene name, which corresponds here to a Fisher exact test comparing the proportions of variant carriers between cases and controls. We computed corresponding OR and confidence intervals (CIs) with R statistical software (http://www.R-project.org/). Single analysis results were obtained via the Fisher test function from package stats, while fixed effect meta-analysis results were obtained via the metabin function from package meta. We applied the same quality checks and gene-level association tests to every RefGene coding region in order to extract all possible gene-level p values and compute the genomic inflation factor. Using the estlamba function of the R statistical package genABEL, we found a value of $\lambda = 0.88$ for this genomic inflation factor, thereby excluding possible confounding from fine-scale population stratification. Finally, we checked that the observed association signal was independent from *APOE* status in a logistic regression model using the glm function from the R statistical software, adjusting for *APOE* ε4–positive status.

RESULTS We first analyzed rare LOF variants (nonsense, frameshift indels, and splice site) and identified in the entire dataset 12 variants (3 nonsense, 6 frameshift indels, and 3 splice site) present in 20 patients and 5 variants (3 frameshift and 2 splice site) present in 8 controls (OR 3.13 [1.30–8.30], $p = 0.006$) (table e-1 on the *Neurology*® Web site at Neurology.org). Besides canonical splice site variants, we included in this count the c.5570+5G>C variant, which has previously been shown to disrupt the splicing of exon 41.[13] Of note, 4 other intronic mutations, located close to splice sites (c.2067+4A>T, c.2684+5G>A, c.3472+5G>C, c.5279+4A>G),

were exclusively found in patients. Although they are predicted in silico as probably affecting splicing, we conservatively decided not to include them in the count. Overall, 4 of the 14 LOF mutations found in the whole dataset were already reported in the Icelandic study,[13] 6 were found in the Flanders-Belgian cohort,[14] and 8 were novel (table e-1). Four variants were simultaneously present in the 3 datasets from Iceland, Belgium, and France.

We then focused on rare strictly damaging missense variants, namely variants with a MAF below 1% in our entire dataset, which were predicted as damaging by 3 software programs: Polyphen2 HumDiv, Mutation Taster, and SIFT. We identified 10 rare strictly damaging missense variants present in 13 cases and 5 in 5 controls, resulting in an OR of 3.23 (95% CI [1.07–11.63], $p = 0.03$) (table e-1). In total, 32 patients (6.6%) vs 12 controls (2.0%) carried at least one LOF or strictly damaging variant. Missense mutations were located throughout the sequence of *ABCA7*. No particular exon or functional domain was hit. The cumulative effect of LOF and strictly damaging missense variants in patients and controls resulted in an OR of 3.40 (95% CI [1.68–7.35], $p = 0.0002$). Moreover, after adjustment for *APOE4*+ status, the effect of *ABCA7* variants remained largely significant (OR 3.1, 95% CI [1.57–6.48], $p = 0.0016$). Therefore, the observed signal of association between EOAD and *ABCA7* variants can be considered independent from *APOE4* status. However, no interaction between these 2 risk factors could be detected ($p = 0.666$). Of note, 4 patients also carried a *SORL1* risk allele (table e-1).

We next performed a meta-analysis including the *ABCA7* published data. Due to differences in the study design, it was not possible to recover the full load of *ABCA7* rare variants in patients and controls of the Icelandic sample. Therefore, this analysis was restricted to the Flanders-Belgian sample (table e-2). In the combined study group of 1,256 patients and 1,347 controls from France and Belgium, the OR resulting from a fixed-effect meta-analysis was 2.89 (95% CI [1.78–4.70], $p = 1.83 \times 10^{-5}$) for LOF variants and 2.81 (95% CI [1.89–4.20], $p = 3.60 \times 10^{-7}$) for combined LOF and strictly damaging variants.

Finally, after the report that a frequent variant (p.Glu188Gly) had a nominally significant increased frequency in controls in the Flanders-Belgian sample,[14] we checked the allelic frequency of this variant in our dataset. The frequency of this variant was 0.41 in patients compared to 0.43 in controls (NS).

DISCUSSION In this study, we explored the common disease–rare variants paradigm in EOAD. We performed WES with high depth of coverage (mean ~120×) on 484 well-characterized French patients with EOAD and 590 ethnically matched controls. There was a clear excess of cases carrying at least one LOF or a predicted damaging missense variant in the *ABCA7* gene compared to controls (6.6% vs 2.0%). Of note, about half of the controls were still in the age range for developing EOAD, and although this condition is rare we cannot exclude that a small subset of them will develop it. Therefore the association could have been slightly underestimated. Performing a meta-analysis of these data with the previously published Belgian series allowed reaching an exome-wide significant association signal, strongly supporting the role of *ABCA7* in AD determinism. It had initially been suggested that *ABCA7* played a role in AD by favoring amyloid clearance.[26] Recently, it has been shown that *ABCA7* also interferes with APP processing and reduces Aβ production, possibly by regulating endocytic pathways.[27] In each case, suppression of *ABCA7* is predicted to result in increased Aβ load, which is consistent with the burden of LOF mutations found in patients with AD. Interestingly, considering the genetic data presented here, it is likely that at least a subset of rare missense strictly damaging *ABCA7* mutations also results in *ABCA7* LOF.

Of note, among the 32 patients with an *ABCA7* LOF or missense variant, 4 additionally carried a *SORL1* risk allele, pointing toward a possible oligogenic model of inheritance. However, considering the rarity of the observed variants, performing an accurate polygenic score analysis including *SORL1*, *ABCA7*, and *TREM2* rare variants together with *APOE* status is still premature in terms of statistical power. Finally, the *ABCA7* frequent variant (p.Glu188Gly) that had a nominally significant increased frequency in controls of the Flanders-Belgian sample was not found to be associated in the present sample, although a same trend toward an increased frequency in controls was noted. Our negative result might reflect a lack of power.

AUTHOR CONTRIBUTIONS

Principal investigator and study supervision: Dominique Campion. Manuscript draft: Kilan Le Guennec, Gaël Nicolas, Camille Charbonnier, Dominique Campion. Acquisition, collection, and interpretation of data: Kilan Le Guennec, 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 the CNR-MAJ collaborators.

STUDY FUNDING

This study was funded by grants from the Clinical Research Hospital Program from the French Ministry of Health (GMAJ, PHRC 2008/067), the CNR-MAJ, the JPND PERADES, and France Génomique.

This work was supported by Labex GENMED ANR-10-LABX-0013, the National Foundation for Alzheimer's Disease and Related Disorders, the Institut Pasteur de Lille, the Centre National de Génotypage, INSERM, FRC (Fondation pour la Recherche sur le Cerveau), Rotary, and LABEX (Laboratory of Excellence Program Investment for the Future) and DISTALZ (Development of Innovative Strategies for a Transdisciplinary Approach to Alzheimer's Disease) grants.

DISCLOSURE

The authors report no disclosures relevant to the manuscript. Go to Neurology.org for full disclosures.

Received October 26, 2015. Accepted in final form February 29, 2016.

REFERENCES

1. Jonsson T, Stefansson H, Steinberg S, et al. Variant of TREM2 associated with the risk of Alzheimer's disease. N Engl J Med 2013;368:107–116.
2. Guerreiro R, Wojtas A, Bras J, et al. TREM2 variants in Alzheimer's disease. N Engl J Med 2013;368:117–127.
3. Cruchaga C, Karch CM, Jin SC, et al. Rare coding variants in the phospholipase D3 gene confer risk for Alzheimer's disease. Nature 2014;505:550–554.
4. Pottier C, Hannequin D, Coutant S, et al. High frequency of potentially pathogenic SORL1 mutations in autosomal dominant early-onset Alzheimer disease. Mol Psychiatry 2012;17:875–879.
5. Jonsson T, Atwal JK, Steinberg S, et al. A mutation in APP protects against Alzheimer's disease and age-related cognitive decline. Nature 2012;488:96–99.
6. Wetzel-Smith MK, Hunkapiller J, Bhangale TR, et al. A rare mutation in UNC5C predisposes to late-onset Alzheimer's disease and increases neuronal cell death. Nat Med 2014;20:1452–1457.
7. Nicolas G, Charbonnier C, Wallon D, et al. SORL1 rare variants: a major risk factor for familial early-onset Alzheimer's disease. Mol Psychiatry Epub 2015 Aug 25.
8. Heilmann S, Drichel D, Clarimon J, et al. PLD3 in non-familial Alzheimer's disease. Nature 2015;520:E3–E5.
9. Hooli BV, Lill CM, Mullin K, et al. PLD3 gene variants and Alzheimer's disease. Nature 2015;520:E7–E8.
10. Lambert JC, Grenier-Boley B, Bellenguez C, et al. PLD3 and sporadic Alzheimer's disease risk. Nature 2015;520:E1.
11. van der Lee SJ, Holstege H, Wong TH, et al. PLD3 variants in population studies. Nature 2015;520:E2–E3.
12. Cacace R, Van den Bossche T, Engelborghs S, et al. Rare variants in PLD3 do not affect risk for early-onset Alzheimer disease in a European Consortium Cohort. Hum Mutat 2015;36:1226–1235.
13. Steinberg S, Stefansson H, Jonsson T, et al. Loss-of-function variants in ABCA7 confer risk of Alzheimer's disease. Nat Genet 2015;47:445–447.
14. Cuyvers E, De Roeck A, Van den Bossche T, et al. Mutations in ABCA7 in a Belgian cohort of Alzheimer's disease patients: a targeted resequencing study. Lancet Neurol 2015;14:814–822.
15. McKhann GM, Knopman DS, Chertkow H, et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 2011;7:263–269.
16. Nicolas G, Wallon D, Charbonnier C, et al. Screening of dementia genes by whole-exome sequencing in early-onset Alzheimer disease: input and lessons. Eur J Hum Genet Epub 2015 Aug 5.
17. Rovelet-Lecrux A, Hannequin D, Raux G, et al. APP locus duplication causes autosomal dominant early-onset Alzheimer disease with cerebral amyloid angiopathy. Nat Genet 2006;38:24–26.
18. DeJesus-Hernandez M, Mackenzie IR, Boeve BF, et al. Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. Neuron 2011;72:245–256.
19. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 2010;20:1297–1303.
20. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 2010; 26:589–595.
21. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 2010;38:e164.
22. Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT. Data quality control in genetic case-control association studies. Nat Protoc 2010;5: 1564–1573.
23. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 2015;4:7.
24. Jun G, Flickinger M, Hetrick KN, et al. Detecting and estimating contamination of human DNA samples in sequencing and array-based genotype data. Am J Hum Genet 2012;91:839–848.
25. Wang GT, Peng B, Leal SM. Variant association tools for quality control and analysis of large-scale sequence and genotyping array data. Am J Hum Genet 2014;94: 770–783.
26. Kim WS, Li H, Ruberu K, et al. Deletion of Abca7 increases cerebral amyloid-beta accumulation in the J20 mouse model of Alzheimer's disease. J Neurosci 2013; 33:4387–4394.
27. Satoh K, Abe-Dohmae S, Yokoyama S, St George-Hyslop P, Fraser PE. ABCA7 loss of function Alters Alzheimer amyloid processing. J Biol Chem 2015;290: 24152–24165.