

SHORT REPORT

SORL1 variants across Alzheimer's disease European American cohorts

Maria Victoria Fernández^{1,2}, Kathleen Black^{1,2}, David Carrell¹, Ben Saef^{1,2}, John Budde^{1,2}, Yuetiva Deming^{1,2}, Bill Howells^{1,2}, Jorge L Del-Aguila^{1,2}, Shengmei Ma^{1,2}, Catherine Bi^{1,2}, Joanne Norton^{1,2}, Rachel Chasse^{1,2}, John Morris^{2,3,4}, Alison Goate⁵ and Carlos Cruchaga^{*,1,2} NIA-LOAD family study group, NCRAD

The accumulation of the toxic A β peptide in Alzheimer's disease (AD) largely relies upon an efficient recycling of amyloid precursor protein (APP). Recent genetic association studies have described rare variants in *SORL1* with putative pathogenic consequences in the recycling of APP. In this work, we examine the presence of rare coding variants in *SORL1* in three different European American cohorts: early-onset, late-onset AD (LOAD) and familial LOAD.

European Journal of Human Genetics (2016) 24, 1828–1830; doi:10.1038/ejhg.2016.122; published online 21 September 2016

INTRODUCTION

The main pathogenic feature of Alzheimer's disease (AD) is the accumulation of amyloid beta (A β) peptide in brain intracellular compartments.¹ Elements involved in the transport and recycling of the amyloid precursor protein (APP) have been considered targets for genetic studies to find AD risk factors. By performing association studies between genetic variants in genes of the endocytic pathway and AD, Rogaeva *et al.*² first identified 19 common SNPs and two haplotypes of the neuronal sortilin-related receptor *SORL1* to be associated with late-onset Alzheimer's disease (LOAD). *SORL1* is a member of the retromer complex that directly binds to APP and differentially regulates its sorting into endocytic or recycling pathways. Despite some initial mixed results,^{3,4} recent studies provide supporting evidence for the association of *SORL1* with AD risk.^{5–8} Nonetheless, clear pathogenic variants that cause *SORL1* malfunctioning have not yet been identified. Recently, Vardarajan *et al.*⁹ suggested that coding variants in *SORL1* may be involved in LOAD pathology. They described up to 17 exonic variants significantly associated with the disease ($P_{\text{adjusted}}=0.008$) in a family-based Caribbean-Hispanic data set. Nicolas *et al.*¹⁰ listed 24 rare variants in *SORL1* with a cumulative effect of an odds ratio (OR) = 5.03 ($P=7.49 \times 10^{-5}$) in a French cohort of early-onset Alzheimer's disease (EOAD) patients. However, these findings have not been replicated on independent data sets. Considering that genetic variants, and particularly rare variants, in risk factors are prone to be cohort specific, it is imperative to analyze this gene across different populations. The goal of this study was to evaluate and replicate the presence of non-synonymous variants in *SORL1* gene that may increase the risk for AD in three different samples of the European American population: sporadic early-onset Alzheimer's disease (sEOAD), sporadic late-onset Alzheimer's disease (sLOAD) and familial LOAD (fLOAD; Table 1).

MATERIALS AND METHODS

All samples included in this analysis were recruited by the Charles F. and Joanne Knight Alzheimer's Disease Research Center (Knight-ADRC) and the National Institute on Aging Genetics Initiative for Late-Onset Alzheimer's Disease (NIA-LOAD). sEOAD samples came from the Memory and Aging Project (MAP), part of Washington University School of Medicine's (WUSM) Knight-Alzheimer's Disease Research Consortium (ADRC). The Institutional Review Board at the WUSM in Saint Louis approved the study. Research was carried out in accordance with the approved protocol. Written informed consent was obtained from participants and their family members by the Clinical and Genetics Core of the Knight-ADRC. The approval number for the Knight-ADRC Genetics Core family studies is 201104178.

Genetic data

sEOAD samples were genotyped for 54 variants: 24 variants reported by Nicolas *et al.*¹⁰ as well as 27 non-synonymous variants with MAF < 5% based on the EVS database (<http://evs.gs.washington.edu/EVS/>) (Supplementary Table S1), MassARRAY (Agena Biosciences) or KASP Assay (LGC Genomics, Teddington, UK).

sLOAD samples were genotyped using the Human Exome Bead-Chip v1.0 (Illumina, San Diego, CA, USA) technology. Stringent quality controls for exome array calling were performed. Genotype calling was carried out using Illumina's GenTrain version 1.0 clustering algorithm in GenomeStudio version 2011.1. Cluster boundaries were determined using study samples for only the calls with an intensity signal > 0.3. A minimum call rate of 98% was used to exclude SNPs and individuals. SNPs that were not in Hardy-Weinberg equilibrium ($P < 10^{-6}$) were dropped. Pairwise genome-wide estimates of proportion identity-by-descent was used to test for unanticipated duplicates and cryptic relatedness.

fLOAD samples were sequenced using either whole-exome sequencing (WES, $n=1177$) or whole-genome sequencing (WGS, $n=59$). Exome

¹Department of Psychiatry, Washington University School of Medicine, St. Louis, MO, USA; ²Hope Center for Neurological Disorders, Washington University School of Medicine, St. Louis, MO, USA; ³Knight Alzheimer's Disease Research Center, Washington University School of Medicine, St. Louis, MO, USA; ⁴Department of Neurology, Washington University School of Medicine, St. Louis, MO, USA; ⁵Department of Neuroscience, Icahn School of Medicine at Mount Sinai, NY, USA

*Correspondence: Dr C Cruchaga, Department of Psychiatry, Washington University School of Medicine, 660 South Euclid Avenue B8134, St. Louis 63110, MO, USA. Tel: +1 314 286 0546; Fax: +1 314 362 2244; E-mail: ccruchaga@wustl.edu

Received 17 March 2016; revised 27 June 2016; accepted 5 August 2016; published online 21 September 2016

Table 1 Clinical data of the three demographic groups studied

	<i>n</i>	<i>M</i> (%)	<i>AAO</i> ($x \pm SD$)	<i>ALA</i> ($x \pm SD$)	<i>APOE-ε4</i> (%)
<i>sEOAD</i>					
CO	169	42.60		75.95 ± 9.97	23.08
CA	217	51.15	59.34 ± 9.17	68.93 ± 13.68	62.67
<i>sLOAD</i>					
CO	266	41.73		72.51 ± 6.24	29.70
CA	134	45.52	70.64 ± 7.44	75.23 ± 6.62	28.96
<i>fLOAD</i>					
CO	324	41.36		86.57 ± 8.25	51.58
CA	866	35.80	73.34 ± 7.42	91.71 ± 8.17	73.44

Abbreviations: AAO, age at onset; ALA, age at last assessment; APOE-ε4 (%), percentage of APOE allele 4 carriers; fLOAD, familial late-onset Alzheimer's disease; CA, case; CO, control; M (%), percentage of males; sEOAD, sporadic early-onset Alzheimer's disease; sLOAD, sporadic late-onset Alzheimer's disease.

libraries were prepared using Agilent's SureSelect Human All Exon kits V3 and V5 (Agilent, Santa Clara, CA, USA). Both, WES and WGS samples were sequenced on a HiSeq2000 (Illumina) with paired ends reads, with a mean depth of coverage of 50–150× for WES and 30× for WGS. Variant calling was performed following GATK's 3.4 Best Practices (<https://www.broadinstitute.org/gatk/>). Alignment was conducted against UCSC hg19 genome reference. WES and WGS sequences were aligned and variants were called separately, following GATK's recommendations; back calling was performed to ensure that the same variants were called in both WES and WGS samples. Variant calling was restricted to Agilent's Exome capture kit with a padded 100 bp region. Only those variants and indels that fell within the above 99.9 tranche and whose quality was ≥ 30 , read depth ≥ 10 and missingness $\leq 5\%$; and those genotypes satisfying a genotype quality ≥ 20 and a DP ≥ 6 were kept for analysis. Variants with differential missingness between cases and controls, as well as between WES and WGS data sets, out of Hardy-Weinberg equilibrium ($P < 1 \times 10^{-6}$) were removed from analysis. In addition, individuals with discordant sex from that reported in the clinical database were removed from data set. Finally, individual and familial relatedness was corroborated using PLINK1.9 (<https://www.cog-genomics.org/plink2/ibd>) and an existing GWAS data set for these individuals. Functional annotation and population frequencies were annotated with SnpEff.¹¹ All *SORL1* variant annotations refer to sequence with Accession Number NM_003105.5. The data and phenotypes used in this study have been submitted to NIAGADS – The NIA Genetics of Alzheimer's Disease Storage Site '<https://www.niagads.org/>' under accession number NG00051.

Statistical analysis

Single-variant association analysis with risk for AD for all data sets were performed using PLINK1.9, including significant covariates (gender and Principal Component for population stratification); for the family-based data set we used DFAM. For gene-wise analysis we only considered non-synonymous variants (missense, splice site or stop modifier) with MAF < 5% within a data set using the SNP-set Kernel Association Test (SKAT).¹² The fLOAD samples were analyzed in addition via GEE Kernel Machine score test (GSKAT).¹³

RESULTS

Sporadic EOAD

From the 48 successfully genotyped variants in the sEOAD cohort (Supplementary Table S1), only one variant (rs117260922:G > A, hg19

Table 2 Gene-based analysis of non-synonymous and damaging *SORL1* variants in each of the demographic groups studied

<i>Test</i>	<i>Non-synonymous</i>				<i>PoD/PrD</i>			
	<i>n</i>	<i>cMAF-A</i>	<i>cMAF-U</i>	<i>P-value</i>	<i>n</i>	<i>cMAF-A</i>	<i>cMAF-U</i>	<i>P-value</i>
<i>sEOAD</i>								
SKAT	4	0.041	0.016	0.055	4	0.041	0.016	0.055
<i>sLOAD</i>								
SKAT	17	0.104	0.111	0.795	6	0.052	0.060	0.945
<i>fLOAD</i>								
G-SKAT	45			0.097	17			0.362
SKAT ^a	45			0.341	17			0.569

Abbreviations: cMAF-A, cumulative minor allele frequency in affected individuals; cMAF-U, cumulative minor allele frequency in unaffected individuals; fLOAD, familial late-onset Alzheimer's disease; n, number of SNPs included in kernel association test; PoD/PrD, possibly damaging/probably damaging according to Polyphen2; sEOAD, sporadic early-onset Alzheimer's disease; sLOAD, sporadic late-onset Alzheimer's disease.

^aFour principal components and kinship matrix were included as covariates to account for related individuals.

chr11:g.121367627G > A) was found nominally associated with AD status (OR = 3.462, $P_{\text{nominal}} = 0.043$), found in 12 cases ($n = 217$) and three controls ($n = 169$) (Supplementary Table S2). This variant was previously reported associated with LOAD risk in Hispanic families⁹ ($P = 7.68 \times 10^{-7}$), but we did not find a significant association in our fLOAD data set (Supplementary Table S3). Two other variants were more frequent in EOAD cases than in controls (rs140327834:T > A, rs142884576:C > T) although we did not find a significant difference. Nonetheless, gene-based analysis indicated there are more non-synonymous variants in EOAD cases than in controls (collapsed MAF: CA = 4.13%; CO = 1.55%; OR = 2.66; 95%CI = 0.6380–8.9850) almost reaching statistical significance (SKAT $P = 0.055$; Table 2).

Sporadic LOAD

From the 46 *SORL1* variants genotyped in the LOAD case-control sample (134 cases and 266 controls) 16 were polymorphic (Supplementary Table S3). Four variants were present more often in cases than in controls (rs139794846:A > G, rs62617129:A > G, rs143286467:A > G, rs143615238:G > A), although the association of any of these variants to AD risk was not significant in this data set ($P = 0.2–0.6$; Supplementary Table S3). The gene-based analysis of the missense variants was non-significant (OR = 0.992; $P = 0.979$), and the same output was achieved when the analysis was restricted to those considered as probably or possibly damaging by Polyphen2 (OR = 0.861, $P = 0.999$; Table 2).

Familial LOAD

Within the fLOAD data set (875 cases and 328 controls), we identified 78 polymorphic variants in *SORL1* coding region, 43 of which were considered non-synonymous and among those 17 were classified as probably or possibly damaging by Polyphen2 (Supplementary Table S4). No single-variant test provided a significant association for AD risk. The combined gene-based G-SKAT analysis of the 45 coding non-synonymous variants did not find any significant association with LOAD ($P = 0.337$), nor did the analysis of the 21 variants considered probably damaging ($P = 0.596$; Supplementary Table S4).

DISCUSSION

Gene-based analyses provide more power to detect association than single-variant analyses, especially when these variants present a low frequency (MAF < 1%). This is supported by previous studies in which multiple independent variants have been reported as causative¹⁴ or increase AD risk in *APP*, *PSEN1*, *PSEN2*, *APOE*, *TREM2*, *PLD3* and *ABCA7*.^{15–18}

In this study, we have observed similar effect of rare variants in *SORL1* as in previous studies, both at single-variant level (rs117260922:G > A), and at a gene-based level in the sEOAD cohort, adding support to the role of rare missense variants in *SORL1* as risk factors for AD. Although our sEOAD sample size (217 cases and 169 controls) was smaller than that of Nicolas *et al.*¹⁰ (484 cases and 498 controls), we still had enough statistical power (83.4%) to replicate the original finding (gene-based OR = 5.03).

Our results suggest that the effect size of *SORL1* may be lower than originally reported. However, it is important to note that in this data set we performed genotyping and not sequencing; therefore, we may have missed additional variants that could affect our OR estimation and *P*-value.

The lack of significant findings in the sLOAD data set may possibly be due to a combination of limited power and the fact that we were only looking at exome-chip variants, not sequencing data. Instead, the lack of significant association on our fLOAD data set raises some concerns. We analyzed a very large data set containing sequencing data for familial LOAD (345 families, 1190 individuals), but we were unable to find a significant association at the gene-based level, even though we found some variants that seem to segregate in some small families. On the other hand, the different degree of association of *SORL1* across the different populations (French,¹⁰ Caribbean-Hispanic⁹ and European American), reinforces the idea that the specific effect size for these low-frequency and rare variants are population-specific. Therefore, in order to replicate these studies, single-variant analyses are not optimal. Instead resequencing of the entire genes in well-matched populations are the most idealistic approach to determine whether these genes are really associated with disease status, and to determine the real effect size of the association.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We thank contributors who collected samples used in this study, as well as patients and their families, whose help and participation made this work possible. This work was supported by grants from the National Institutes of Health (R01-AG044546, P01-AG003991, and R01-AG035083), and the Alzheimer Association (NIRG-11-200110, BAND-14-338165 and BFG-15-362540). This research was conducted while CC was a recipient of a New Investigator Award in Alzheimer's disease from the American Federation for Aging Research. CC is a recipient of a BrightFocus Foundation Alzheimer's Disease Research Grant (A2013359S). The recruitment and clinical

characterization of research participants at Washington University were supported by NIH P50 AG05681, P01 AG03991, and P01 AG026276. Samples from the National Cell Repository for Alzheimer's Disease (NCRAD), which receives government support under a cooperative agreement grant (U24 AG21886) awarded by the National Institute on Aging (NIA), were used in this study. NIALOAD samples were collected under a cooperative agreement grant (U24 AG026395) awarded by the National Institute on Aging.

AUTHOR CONTRIBUTIONS

Study design was developed by CC and MVF. Phenotype data was collected by SM, JN and RC. ADRC, NIA-LAOD and AG recruited participants. Sample collection was conducted by JM, AG, JN and RC. DNA extraction was performed by DC, CB and KB. Genotyping was conducted by KB. Statistical analyses were performed by YD, BH, JA and MVF. Manuscript was drafted by MVF and CC. Funds support was received from CC. All authors read and approved the final manuscript.

- 1 Mattson MP: Pathways towards and away from Alzheimer's disease. *Nature* 2004; **430**: 631–639.
- 2 Rogaeva E, Meng Y, Lee JH *et al*: The neuronal sortilin-related receptor SORL1 is genetically associated with Alzheimer disease. *Nat Genet* 2007; **39**: 168–177.
- 3 Webster JA, Myers A, Pearson JV *et al*: SORL1 as an Alzheimer's disease predisposition gene? *Neurodegener Dis* 2008; **5**: 60–64.
- 4 Li H, Wetten S, Li L *et al*: Candidate single-nucleotide polymorphisms from a genomewide association study of Alzheimer disease. *Arch Neurol* 2009; **65**: 45–53.
- 5 Seshadri S, DeStefano A, Au R *et al*: Genetic correlates of brain aging on MRI and cognitive test measures: a genome-wide association and linkage analysis in the Framingham Study. *BMC Med Genet* 2007; **8**(Suppl 1): S15.
- 6 Bettens K, Brouwers N, Engelborghs S *et al*: SORL1 is genetically associated with increased risk for late-onset Alzheimer disease in the Belgian population. *Hum Mutat* 2008; **29**: 769–770.
- 7 Tan EK, Lee J, Chen CP *et al*: SORL1 haplotypes modulate risk of Alzheimer's disease in Chinese. *Neurobiol Aging* 2009; **30**: 1048–1051.
- 8 Meng Y, Lee JH, Cheng R, George-Hyslop PS, Mayeux R, Farrer L: Association between SORL1 and Alzheimer's disease in a genome-wide study. *Neuroreport* 2007; **18**: 1761–1764.
- 9 Vardarajan BN, Zhang Y, Lee JH *et al*: Coding mutations in SORL1 and Alzheimer disease. *Ann Neurol* 2015; **77**: 215–227.
- 10 Nicolas G, Charbonnier C, Wallon D *et al*: SORL1 rare variants: a major risk factor for familial early-onset Alzheimer's disease. *Mol Psychiatry* 2015; **21**: 831–836.
- 11 Cingolani P, Platts A, Wang Le L *et al*: A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)* 2012; **6**: 80–92.
- 12 Wu MC, Seunggeum L, Tianxi C, Li Y, Boehnke M, Lin X: Rare-variant association testing for sequencing data with the sequence kernel association test. *Am J Hum Genet* 2011; **89**: 82–93.
- 13 Wang X, Lee S, Zhu X, Redline S, Lin X: GEE-based SNP set association test for continuous and discrete traits in family-based association studies. *Genet Epidemiol* 2013; **37**: 778–786.
- 14 Cruchaga C, Chakraverty S, Mayo K *et al*: Rare variants in APP, PSEN1 and PSEN2 increase risk for AD in late-onset Alzheimer's disease families. *PLoS One* 2012; **7**: e31039.
- 15 Hollingworth P, Harold D, Sims R *et al*: Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. *Nat Genet* 2011; **43**: 429–435.
- 16 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.
- 17 Jin SC, Sheng C, Benitez BA *et al*: Coding variants in TREM2 increase risk for Alzheimer's disease. *Hum Mol Genet* 2014; **23**: 5838–5346.
- 18 Del-Aguila JL, Koboldt DC, Black K *et al*: Alzheimer's disease: rare variants with large effect sizes. *Curr Opin Genet Dev* 2015; **33**: 49–55.

Supplementary Information accompanies this paper on European Journal of Human Genetics website (<http://www.nature.com/ejhg>)