

ARTICLE

Characterization of pathogenic *SORL1* genetic variants for association with Alzheimer's disease: a clinical interpretation strategy

Henne Holstege^{x,1,2,3}, Sven J van der Lee^{1,4}, Marc Hulsman^{2,3}, Tsz Hang Wong⁵, Jeroen GJ van Rooij^{5,6}, Marjan Weiss², Eva Louwersheimer¹, Frank J Wolters⁴, Najaf Amin⁴, André G Uitterlinden^{4,6,7}, Albert Hofman^{4,8}, M Arfan Ikram^{4,5,9}, John C van Swieten^{1,5}, Hanne Meijers-Heijboer², Wiesje M van der Flier^{1,10}, Marcel JT Reinders³, Cornelia M van Duijn^{4,11} and Philip Scheltens¹

Accumulating evidence suggests that genetic variants in the *SORL1* gene are associated with Alzheimer disease (AD), but a strategy to identify which variants are pathogenic is lacking. In a discovery sample of 115 *SORL1* variants detected in 1908 Dutch AD cases and controls, we identified the variant characteristics associated with *SORL1* variant pathogenicity. Findings were replicated in an independent sample of 103 *SORL1* variants detected in 3193 AD cases and controls. In a combined sample of the discovery and replication samples, comprising 181 unique *SORL1* variants, we developed a strategy to classify *SORL1* variants into five subtypes ranging from pathogenic to benign. We tested this pathogenicity screen in *SORL1* variants reported in two independent published studies. *SORL1* variant pathogenicity is defined by the Combined Annotation Dependent Depletion (CADD) score and the minor allele frequency (MAF) reported by the Exome Aggregation Consortium (ExAC) database. Variants predicted strongly damaging (CADD score > 30), which are extremely rare (ExAC-MAF < 1×10^{-5}) increased AD risk by 12-fold (95% CI 4.2–34.3; $P = 5 \times 10^{-9}$). Protein-truncating *SORL1* mutations were all unknown to ExAC and occurred exclusively in AD cases. More common *SORL1* variants (ExAC-MAF $\geq 1 \times 10^{-5}$) were not associated with increased AD risk, even when predicted strongly damaging. Findings were independent of gender and the *APOE-ε4* allele. High-risk *SORL1* variants were observed in a substantial proportion of the AD cases analyzed (2%). Based on their effect size, we propose to consider high-risk *SORL1* variants next to variants in *APOE*, *PSEN1*, *PSEN2* and *APP* for personalized risk assessments in clinical practice.

European Journal of Human Genetics (2017) 25, 973–981; doi:10.1038/ejhg.2017.87; published online 24 May 2017

INTRODUCTION

Approximately one-third of the population older than 85 years has Alzheimer's disease (AD).¹ Despite intensive research, the pathophysiology underlying AD is still poorly understood. The risk to develop (non-Mendelian) AD is estimated to be 60–80% heritable,² suggesting that the identification of genetic determinants of AD will provide further insights into underlying molecular mechanisms of AD.

Rapidly accumulating genetic and biological evidence suggests that disturbed function of the sortilin-related receptor 1 (*SORL1*) is associated with AD.^{3–7} Functional *SORL1* reduces the amyloid- β levels in the brain, thereby reducing the load of neurotoxic amyloid- β plaques, a neuropathological hallmark of AD.⁸ *SORL1* reduces amyloid- β levels (i) by binding the amyloid precursor protein (APP), preventing its processing into amyloid- β and (ii) by binding amyloid- β and directing it to the lysosome for degradation.

To exert its function, the *SORL1* protein includes domains from the low-density lipoprotein receptor-like family, including complement-type repeats that interact in a 1:1 stoichiometric complex with APP.^{9,10}

The *SORL1* protein also includes a VPS10 domain from the vacuolar protein sorting-10 receptor family, which binds soluble A β for endosomal inclusion and sorting for lysosomal degradation.^{11,12} Therefore, impaired *SORL1* function is associated both with disturbed APP processing and disturbed A β degradation, two central events underlying the pathophysiology of AD.^{13–16}

Since 2007, a multitude of studies associated both common and rare variants in the *SORL1* gene with AD,^{17–21} but different variants were associated across studies and risk-effects ranged from small modifying effects to causal effects.^{22,23} In genome-wide association studies (GWAS) including thousands of AD cases and controls, common genetic variants near and in the *SORL1* gene were found to associate significantly with AD.²⁴ However, each of these variants confer only a small increase in AD risk (OR = ~1.2), which is comparable to the small changes in AD risk conferred by the other ~20 genetic loci that were also identified in these GWAS studies (reviewed by Van Cauwenberghé *et al.*²⁵). In contrast, recent studies reported that rare *SORL1* variants are associated with a five-fold increased risk for

¹Alzheimer Center, Department of Neurology, VU University Medical Center, Neuroscience Campus Amsterdam, Amsterdam, The Netherlands; ²Department of Clinical Genetics, VU University Medical Center, Neuroscience Campus Amsterdam, Amsterdam, The Netherlands; ³Delft Bioinformatics Lab, Delft University of Technology, Delft, The Netherlands; ⁴Department of Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands; ⁵Department of Neurology, Erasmus Medical Center, Rotterdam, The Netherlands; ⁶Department of Internal Medicine, Erasmus Medical Center, Rotterdam, The Netherlands; ⁷Netherlands Consortium on Health Aging and National Genomics Initiative, Leiden, The Netherlands; ⁸Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA; ⁹Department of Radiology, Erasmus MC University Medical Center, Rotterdam, The Netherlands; ¹⁰Department of Epidemiology & Biostatistics, VU University Medical Center, Neuroscience Campus Amsterdam, Amsterdam, The Netherlands; ¹¹Netherlands Metabolomics Centre, Leiden University, Leiden, The Netherlands

*Correspondence: Dr H Holstege, VUmc Alzheimer Center, VU University Medical Center, de Boelelaan 1118, Amsterdam 1081 HZ, The Netherlands. Tel: +31 20 4440816; Fax: +31 20 4448529; E-mail: h.holstege@vumc.nl

Received 2 April 2017; revised 12 April 2017; accepted 13 April 2017; published online 24 May 2017

early onset AD.^{5,6} These studies suggest that the effect on AD risk of these rare *SORL1* variants is comparable to that of carrying the *ε4* allele of Apolipoprotein-E gene (*APOE*), the most important common risk factor for AD; homozygous and heterozygous *APOE-ε4* carriers are respectively exposed to a 3 to 5-fold increased AD risk and 10 to 15-fold increased AD risk compared with non-*APOE-ε4* carriers.²⁶ Furthermore, recent targeted sequencing studies identified rare pathogenic *SORL1* mutations that segregated with disease in families with familial AD and late onset AD.^{3,4} These findings suggest that specific *SORL1* variants are causal for AD, with effects comparable to single mutations in the *APP* gene,²⁷ or the presenilin-1 and presenilin-2 genes (*PSEN1*, *PSEN2*),^{28,29} that are associated with an autosomal dominant inheritance pattern of AD. However, it is currently unclear which specific *SORL1* variants are major risk factors for AD and which can be considered benign. This raises the need for a strategy to determine *SORL1* variant pathogenicity.

Ideally, one would like to determine penetrance for each detected *SORL1* variant in multiple large and informative families. However, the rareness of the *SORL1* variants that were thus far associated with AD complicates such efforts.^{3–6} Therefore, a more feasible approach might be to distinguish between pathogenic and non-pathogenic variants using *independent* variant characteristics that might be associated with *SORL1* variant pathogenicity. In this work we explored the contribution on disease outcome of (i) the functional protein domain affected by the *SORL1* variant; (ii) the minor allele frequency (MAF) of the *SORL1* variant in the population; and (iii) predicted damagingness of the *SORL1* variant on the basis of sequence context.

To predict the damagingness of *SORL1* variants we annotated them with the CADD score,³⁰ a novel functional annotation tool that allows for an unbiased annotation of almost all variants in the human genome. The CADD score reflects the difference between the characteristics of genetic variation that is tolerated (fixed) in the human genome and the characteristics of pathogenic variants (randomly simulated variants enriched with pathogenic variants). Scores are based on >60 functional prediction tools that include functional annotations, allelic conservation and regulatory effects.

The MAF of a variant can be derived from the (large) sample in which it was discovered, but also from publically available databases such as the ExAC database,³¹ which includes variants detected in a set of 60 706 exomes. Importantly, the MAF per variant is *not* included in determining its CADD score, such that the CADD score and MAF can be used as independent determinants of variant pathogenicity.

We determined the characteristics of *SORL1* variant pathogenicity in a discovery sample consisting of 640 Dutch early- and late onset AD patients and in 1268 older Dutch cognitively healthy controls. We tested whether these characteristics replicated in an independent data set reported by Verheijen *et al.*⁶ and we performed a combined analysis including data from both studies. Based on these results, we suggest five *SORL1* variant subtypes according to the five-class system of variant pathogenicity supported by the American College of Medical Genetics and Genomics (ACMG)³² and Association for Clinical Genetic Science (ACGS).³³ Lastly, we validated our classification strategy by applying it to *SORL1* variants described in two additional independent publications.^{4,5}

MATERIAL AND METHODS

Detailed methods are described in the Supplementary Methods.

Samples

For a schematic overview of the analysis setup see Figure 1.

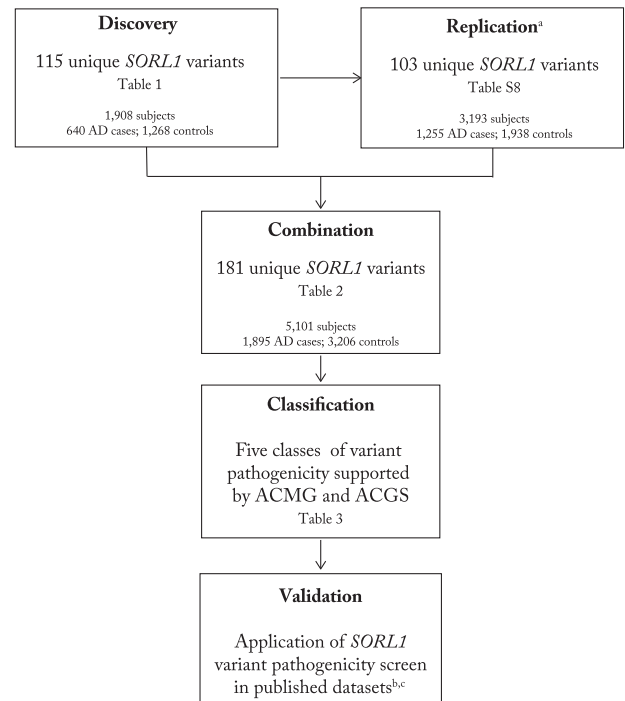


Figure 1 Flowchart of *SORL1* variant pathogenicity analysis. ^a*SORL1* variants in the independent replication data set were reported by Verheijen *et al.*⁶ Pathogenicity screen was applied to *SORL1* variants reported by ^bVardarajan *et al.*⁴ and ^cNicolas *et al.*⁵

Discovery sample. The exome collection was assembled from four Dutch studies: (i) the Rotterdam Study³⁴ contributed 250 AD cases (median age at disease onset 84.5 ± 6.62 , 71.2% females) and 1204 controls (median age at last visit of 82.4 ± 6.8 , 54.3% females), (ii) The Amsterdam Dementia Cohort (ADC-VUmc) contributed 320 AD (median age of 58.4 ± 6.5 , 51.9% females),³⁵ (iii) the Alzheimer Centrum Zuidwest Nederland (ACZN) cohort contributed 80 AD cases (median age 59.2 ± 7.2 , 57.1% females) and (iv) the 100-plus Study (www.100plus.nl) contributed 64 controls (median age of 101.1 years ± 3.5 , 79.7% females).

In total, 1908 exomes passed quality control: 640 cases (median age at onset of 64.8 years, IQR: 57.3–82.2, 60% females) and 1268 controls (median age at last screening of 82.7 years, IQR 78.3–87.6, 55.6% females) (for cohort characteristics see Supplementary Table S1; for age distribution of cases and controls see Supplementary Figure S1). No known AD-causing mutations were detected in the *APP*, *PSEN1* and *PSEN2* genes. Combined data included (i) AD status, (ii) *APOE* status, (iii) gender and (iv) age at onset for AD and (v) age at last screening for controls.

Replication sample. We performed a replication analysis in an independent data set recently published by Verheijen *et al.*⁶ They reported 103 *SORL1* variants in 1255 European early onset AD cases and 1938 age-matched controls. Rare *SORL1* variants that occurred in either cases or controls were given per subject (including gender and age, but not *APOE* genotype), and no two variants occurred in the same subject; the number of case- and control-carriers were given for rare *SORL1* variants that occurred in both cases and controls; common *SORL1* variants (MAF > 0.01 in the sample) were available as sample-MAFs. For an in-depth analysis of independence between the discovery and the replication samples, see Supplementary Data and Supplementary Table S2.

Exome sequencing and variant detection in discovery sample

Exomes from the Rotterdam Study and the ACZN cohort were captured with the Nimblegen v2 Seqcap EZ Exome capture kit. The exomes from the ADC-VUmc cohort and the 100-plus Study cohort were captured with the Nimblegen SeqCap EZ Exome capture kit v3. For *SORL1* variant calling, we used the intersection between these capture kits in the *SORL1* gene: 93.2% of

exons 1–47, 2.7% of exon 48. DNA from all samples was prepared with the Illumina TruSeq Paired-End Library Preparation Kit and 100 bp paired-end reads were acquired by sequencing the libraries on a HiSeq 2000 or 2500. We sequenced to at least 40× mean coverage to ensure sufficient read depth for variant calling.

We removed population outliers based on the first two PCA components and those with an identity by descent value >0.1. Technical differences between data acquisition commonly introduces ‘differential missingness’ (ie, loci may be genotyped in one exome but not in another), which may ultimately result in unwanted bias towards either cases or controls. To overcome this bias we implemented additional quality control (see Supplementary Methods) to obtain a set consisting of 115 true positive variant calls with negligible missingness across the sample. Variants are listed in Supplementary Table S3, and they are submitted to the LOVD database (<https://databases.lovd.nl/shared/references/DOI:10.1038/ejhg.2017.87>). The subset of variants that were detected only once in the sample (singleton variants) were validated by Sanger sequencing (Supplementary Table S4).

Statistical analysis

Variant annotation. Variants were annotated with the CADD score version 1.3³⁰ and using the Variant Effect Predictor (VEP) tool in the Ensembl database.³⁶ Variants were annotated with SIFT v.5.2.2/PolyPhen v.2.2.2 prediction scores (Supplementary Table S3).

Discovery analysis. Since the rarity of most detected variants does not allow a per-variant calculation of disease association, we tested the burden of all *SORL1* variants that adhered to a specified set of characteristics in AD cases and controls.³⁷ For this, variant characteristics were based on combinations of the MAF and CADD scores. The MAF categories included MAF>0.01, 0.001<MAF<0.01, 0.0005<MAF<0.001 and MAF<0.0005 (singletons). The CADD score categories included CADD 0–20 (predicted not or mildly damaging), 20–30 (predicted moderately damaging) and >30 (predicted strongly damaging). In addition, variants were stratified according to *SORL1* protein domains. We then performed burden tests using an additive genetic model and logistic regression score test with the *burdenMeta* function in the ‘seqMeta’ package v.1.6.0³⁸ in R (v3.2.2), while including gender as a covariate. *APOE* genotypes were missing for 36 controls and 6 cases; therefore, we performed separate burden tests using both gender and *APOE-ε4* genotype as covariates. Furthermore, we tested for an interaction effect between the burden of *SORL1* variants and the presence of the *APOE-ε4* allele.

Replication analyses were performed with a one-tailed Fisher’s exact test due to data availability. As rare variants did not overlap between subjects, this approach is the same as a burden test. Correction for *APOE* and gender was not possible because *APOE* genotypes were not publically available in the replication sample, and gender only for a sample-subset.

Finally, for a combined analysis of the discovery sample and the replication sample, we annotated all variants with their MAFs reported in the publically available ExAC database v.0.3.1³¹ (Supplementary Table S5). To more closely distinguish between benign and (possibly) pathogenic variants, we stratified the CADD tranches into CADD 0–10, 10–20, 20–30 and >30. Combined analyses were performed with a one-tailed Fisher’s exact test.

Multiple testing correction was applied to correct for 70 tests performed in discovery, replication and combined analysis. *P*-values lower than 0.05 after Bonferroni were considered statistically significant ($P<7.1\times 10^{-4}$). We present unadjusted *P*-values.

RESULTS

We detected 115 *SORL1* variants in the exomes of 640 AD cases and 1268 controls: 4 frameshift, 1 stop-gain, 54 missense, 29 synonymous, 4 regulatory, 4 splice site and 19 intronic variants. Details of the filtering steps and Sanger validation are in the Supplementary Results. Of these 115 variants, 15 were common (MAF>0.01) and none was significantly associated with AD (Supplementary Table S3). The remaining 100 variants were rare with an MAF<0.01. These variants did not occur more often in cases than in controls (OR=1.2; 95% CI

0.9–1.6; $P=0.23$) (Table 1A). There were 36 variants predicted to be deleterious by SIFT and Polyphen, and these variants were associated with a two-fold increased AD risk (OR=1.9; 95% CI 1.2–2.9; $P=7.2\times 10^{-3}$) (Table 1A).

Singleton variants with high CADD score are associated with AD

Of the 100 rare variants (MAF<0.01), 26 variants were predicted moderately damaging (CADD 20–30) and carrying such a variant is not associated with a significantly increased AD risk (OR=1.3; 95% CI 0.78–2.1; $P=0.34$) (Table 1A). In contrast, the 19 variants that were predicted strongly damaging (CADD>30) were seen in 15 cases and 8 controls, such that carrying a variant with these characteristics is associated with a four-fold increased risk for AD: OR=4.0; 95% CI 1.7–9.0; $P=9.9\times 10^{-4}$ (Table 1A).

Interestingly, 16 of these 19 predicted strongly damaging variants with MAF<0.01 were seen only once in our sample: singletons. Among the 16 carriers of strongly damaging singletons (CADD>30), 14 had developed AD, such that strongly damaging singletons were associated with a >10-fold increase of AD risk (OR=11.3; 95% CI 4.0–32.1; $P=4.9\times 10^{-6}$). Notably, all five truncating mutations (stop-gain/frameshift) were singletons in our sample, and their carriers all developed AD ($P=1.6\times 10^{-3}$). In sharp contrast, variants with CADD >30 that occur more than once in this sample were not associated with AD (Table 1B). Gender and carrying the *APOE-ε4* genotype did not influence these findings (Supplementary Table S6) and we detected no evidence for an interaction of a predicted damaging *SORL1* variant (CADD 20–30 or CADD>30) and carrying the *APOE-ε4* genotype (Supplementary Table S7 and Supplementary Results). Of note, no subject carried more than one singleton variant, with the exception of one control subject who carried two intronic singleton variants.

The association with AD of singleton variants that were predicted damaging was further illustrated by the finding that the median CADD score for the 30 singletons detected in cases (28.3; IQR 17.6–32.8) was significantly higher than the median CADD score for the 40 singletons in controls (14.9; IQR 8.1–23.1; $P=5.2\times 10^{-5}$). The median CADD score of only slightly more common variants did not significantly differ between cases and controls ($P=0.23$) (Supplementary Figure S2).

The median age at onset for carriers of strongly damaging *SORL1* singletons (CADD >30, $n=14$) was 58.9 years, compared with 65.1 years for cases without these singletons ($P=0.08$, one-tailed Mann–Whitney *U*-test). The five cases with carriers of stopcodon/frameshift mutations had a median age at onset of 57.7 (Supplementary Results).

Next, we analyzed whether there was a differential burden of damaging singleton variants in individual *SORL1* protein domains (Table 1C; for affected amino-acid positions in protein see Supplementary Figure S3). All six subjects who carried a moderately damaging singleton (CADD 20–30) in the VPS10 domain developed AD, which was associated with almost 20-fold increased AD risk (OR=19.3; 95% CI 3.6–105.2; $P=6.1\times 10^{-4}$), and three out of four subjects with a strongly damaging variant in the VPS10 domain (CADD>30) developed AD (OR=6.6; 95% CI 0.82–53.1; $P=7.6\times 10^{-2}$). In contrast, only 1 out of 12 subjects who carried such a moderately damaging variant in one of the other domains developed AD.

Replication in published data

We replicated our findings in 103 *SORL1* variants reported in an independent published data set.⁶ In this data set, singleton variants

Table 1 Discovery analysis: 115 *SORL1* variants in 640 cases and 1268 controls

CADD score	N-variants in burden test	Cases	Controls	OR (95% CI)	P-value ^a (unadjusted)
		With at least one variant Total 640	With at least one variant Total 1268		
A. Variant stratification according to sample MAF <0.01 and SIFT/Polyphen and CADD					
MAF <0.01					
All variants	100	83	140	1.2 (0.9–1.6)	0.23
SIFT/Polyphen damaging	36	40	44	1.9 (1.2–2.9)	7.2 × 10 ⁻³
CADD 0–20	55	42	94	0.9 (0.6–1.3)	0.59
CADD 20–30	26	28	43	1.3 (0.78–2.1)	0.34
CADD >30	19	15	8	4.0 (1.7–9.0)	9.9 × 10 ⁻⁴
B. Variant stratification according to sample MAF and CADD					
MAF >0.01 (variants detected in >38 subjects in this sample)					
0–20	13	640	1268	1.0 (1.0–1.1)	0.54
20–30	1	48	103	1.0 (0.7–1.3)	0.83
>30	1	17	44	0.8 (0.5–1.3)	0.29
0.001 < MAF < 0.01 (variants detected in 3–38 subjects in this sample)					
0–20	7	27	47	1.2 (0.7–1.9)	0.54
20–30	3	18	24	1.5 (0.8–2.9)	0.21
>30	1	1	3	0.7 (0.1–5.5)	0.73
0.0005 < MAF < 0.001 (variants detected in 2–3 subjects in this sample)					
0–20	12	6	21	0.7 (0.3–1.4)	0.33
20–30	5	3	9	0.7 (0.2–2.4)	0.57
>30	2	1	3	0.7 (0.1–5.2)	0.70
MAF <0.0005 (singletons in this sample)					
0–20	36	9	26*	0.7 (0.4–1.4)	0.31
20–30	18	7	11	1.3 (0.5–3.3)	0.65
>30	16	14	2	11.3 (4.0–32.1)	4.9 × 10⁻⁶
>30 missense	11	9	2	8.7 (2.5–30.4)	7.2 × 10 ⁻⁴
>30 stop/frameshift	5	5	0	19.8 (3.1–126.8) ^b	1.7 × 10 ⁻³
C. Variant stratification of singleton variants according to protein domain and CADD					
VPS10					
0–20	5	1	4	0.6 (0.1–3.6)	0.54
20–30	6	6	0	19.3 (3.6–105.2)^b	6.1 × 10⁻⁴
>30	4	3	1	6.6 (0.8–53.1)	7.6 × 10 ⁻²
LDL A and B					
0–20	10	3	7	0.9 (0.2–3.2)	0.81
20–30	4	0	4	0.22 (0.0–1.7) ^b	0.14
>30	8	7	1	10.8 (2.5–46.7)	1.5 × 10 ⁻³
Fibronectin					
0–20	3	0	3	0.2 (0.0–2.5) ^b	0.23
20–30	5	0	5	0.2 (0.0–1.4) ^b	0.11
>30	4	4	0	20.5 (2.6–164.5) ^b	4.5 × 10 ⁻³

Variants were categorized according to the associated minor allele frequency in the sample and CADD score. For each category, multiple variants per individual were collapsed into one variant such that an odds ratio and an associated P-value could be calculated using the score-based SeqMeta burden test, while using gender as a covariate. LDL-Receptor, low-density lipoprotein receptor A and BOR; odds ratio; VPS10, Vacuolar Protein Sorting domain 10. No subject carried more than one singleton, with the exception of one control subject who carried two intronic singletons with CADD 0–20 (*).

^aAssociations that are significant after multiple testing correction are shown in bold ($P < 7.1 \times 10^{-4}$).

^bResults for the odds ratios can be considered a one-step approximation to the maximum likelihood result, such that ORs of infinite have lower values when controls contribute no variants to the association.

with CADD > 30 associated with a 14-fold increased AD risk (OR = 14.1; 95% CI 3.3–60.8; $P = 3.5 \times 10^{-6}$) and all eight stop-gain/frameshift variants were singletons observed exclusively in cases ($P = 5.6 \times 10^{-4}$) (Supplementary Table S8). Singleton missense variants with CADD > 30 associated with an eight-fold increased AD

risk (OR = 7.8; 95% CI 1.7–35.5; $P = 2.4 \times 10^{-3}$). We were not able to replicate our findings in the VPS10 domain: moderately damaging variants in the VPS10 domain were carried only by five young control subjects (aged 55, 60 and 62 years and two with unknown ages).

Table 2 Combined analysis: 181 unique *SORL1* variants in 1895 cases and 3206 controls

CADD score	N-variants in burden test ^a	Cases	Controls	OR (95% CI)	P-value ^b (unadjusted)
		With at least one variant Total 1895	With at least one variant Total 3206		
<i>ExAC-MAF > 0.01 (each variant detected in > 1200 subjects in the ExAC database)</i>					
0–10	8 ^c	1,895	3,206	1.0	1.0
10–20	4 ^c	1,895	3,206	1.0	1.0
20–30	1	203	300	1.2 (1.0–1.4)	7.3 × 10 ⁻²
>30	1	67	162	0.7 (0.5–0.9)	1.0
<i>0.001 < ExAC-MAF < 0.01 (each variant detected in 121–1200 subjects in ExAC database)</i>					
0–10	5 ^c	93	158	1.0 (0.8–1.3)	0.52
10–20	4 ^c	68	151	0.8 (0.6–1.0)	0.98
20–30	3	37	40	1.6 (1.0–2.5)	3.2 × 10 ⁻²
>30	1	4	7	1.0 (0.3–3.3)	0.63
<i>0.0001 < ExAC-MAF < 0.001 (each variant detected in 13–120 subjects in ExAC database)</i>					
0–10	5	10	12	1.4 (0.6–3.3)	0.27
10–20	5	9	11	1.4 (0.6–3.4)	0.30
20–30	11	15	21	1.2 (0.6–2.4)	0.34
>30	4	5	4	2.1 (0.6–7.9)	0.21
<i>0.00001 < ExAC-MAF < 0.0001 (each variant detected in 2–12 subjects in ExAC database)</i>					
0–10	5	7	10	1.2 (0.5–3.1)	0.45
10–20	20	9	21	0.7 (0.3–1.6)	0.84
20–30	22	9	23	0.7 (0.3–1.4)	0.90
>30	9	10	2	8.5 (1.9–38.8)	1.4 × 10 ⁻³
<i>ExAC-MAF < 0.00001 (unknown or singletons in ExAC database)</i>					
0–10	7	2	8	0.4 (0.1–2.0)	0.93
10–20	15	9	7	2.2 (0.8–5.9)	9.4 × 10 ⁻²
20–30	22	13	11	2.0 (0.9–4.5)	6.6 × 10 ⁻²
>30	29	28	4	12.0 (4.2–34.3)	5.0 × 10⁻⁹

Variants were categorized according to the associated minor allele frequency reported in the ExAC database (ExAC-MAF) and CADD score.

^aIn the replication sample, we were not able to account for the overlap when one subject carried multiple variants with the same MAF.

^bAssociations that are significant after multiple testing correction are shown in bold ($P < 7.1 \times 10^{-4}$).

^cFor common variants derived from the replication sample, the variant MAF reported in cases and controls was used to estimate the number of case-carriers and control-carriers.

Combined analysis

We substantiated the characterization of *SORL1* variant pathogenicity in the 181 unique *SORL1* variants in the 1895 cases and 3206 controls of the combined the discovery and replication samples (Table 2); all variants are listed in Supplementary Table S5.

SORL1 variants that were (i) novel or listed only once in the ExAC database (MAF $< 1 \times 10^{-5}$) and (ii) high predicted variant damagingness (CADD > 30) had the largest effect on AD risk (OR = 12.0; 95% CI 4.2–34.3; $P = 5 \times 10^{-9}$) (Table 2 and Figure 2). Slightly more common variants observed $2\text{--}12 \times$ in the ExAC database (ExAC-MAF between 1×10^{-5} and 1×10^{-4}) with CADD > 30 associated with an 8.5-fold increased AD risk (95% CI 1.9–38.8; $P = 1.4 \times 10^{-3}$); more common variants with ExAC-MAF $> 1 \times 10^{-4}$ and CADD > 30 do not associate with a significantly increased AD risk (Table 2 and Figure 2). Together, the

maximum statistical evidence for an effect on AD risk is obtained for *SORL1* variants with CADD > 30 and ExAC-MAF $< 1 \times 10^{-4}$ (OR = 10.9; 95% CI 4.6–25.7; $P = 1.8 \times 10^{-11}$) (Table 3). *SORL1* variants with these characteristics occur in 38 from 1895 cases (2%) and in 6 from 3206 controls (0.19%).

Extremely rare variants (ExAC-MAF $< 1 \times 10^{-5}$) that are moderately damaging (CADD 20–30) and mildly damaging variants (CADD 10–20) are both suggestively associated with a two-fold increased AD risk (OR = 2.0; 95% CI 0.9–4.5; $P = 6.6 \times 10^{-2}$) and (OR = 2.18, 95% CI 0.8–5.9; $P = 9 \times 10^{-2}$), respectively, whereas variants with CADD 0–10 were not associated with any risk increase (OR = 0.4; 95% CI 0.1–2.0; $P = 0.93$) (Table 2 and Figure 2). In this combined analysis, we found no evidence that moderately damaging variants in the VPS10 domain are associated with AD (OR = 0.91; 95% CI 0.36–2.29; $P = 0.66$).

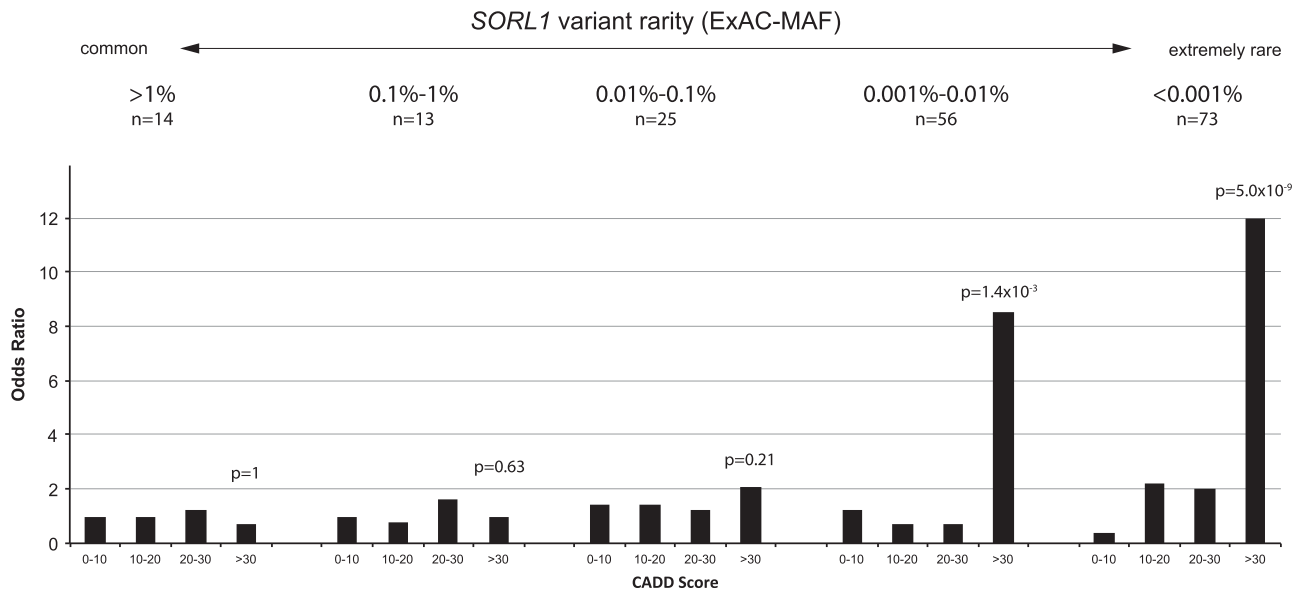


Figure 2 Only the rarest variants with the highest CADD scores are associated with increased AD risk. The 181 *SORL1* variants detected in 5101 AD cases and controls from the combined analysis were first separated by their ExAC-MAF, and then by their CADD values (see also Table 2).

Table 3 Clinical selection criteria of variants

Variant selection criteria	CADD	ExAC-MAF v.0.3.1	N-variants in burden test ^a	Cases		Controls		P-value ^a (unadjusted)
				Total 1895	With at least one variant	Total 3206	With at least one variant	
A. Effect on AD risk								
Maximum effect size	>30	<1 × 10 ⁻⁵	29	28	4	12.0 (4.2–34.3)	5.0 × 10⁻⁹	
Maximum evidence for effect	>30	<1 × 10 ⁻⁴	38	38	6	10.9 (4.6–25.9)	1.8 × 10⁻¹¹	
B. Suggested variant subtypes								
Pathogenic	>30 (truncating)	<1 × 10 ^{-5b}	13	13	0	inf (5.2– inf)	2.5 × 10⁻⁶	
Likely pathogenic	>30 (missense)	<1 × 10 ⁻⁴	25	25	6	7.1 (2.9–17.4)	8.6 × 10⁻⁷	
Uncertain significance								
Possibly pathogenic	10–30	<1 × 10 ⁻⁵	33	21	15	2.4 (1.2–4.6)	7.7 × 10 ⁻³	
Most likely not pathogenic	>30	>1 × 10 ⁻⁴	6	76	173	0.73 (0.6–1.0)	0.99	
Likely benign	10–30	>1 × 10 ⁻⁵	70	1,895	3,206	1.0	1.0	
Benign	0–10	0–1	30	1,895	3,206	1.0	1.0	

^aAssociations that are significant after multiple testing correction are shown in bold ($P < 7.1 \times 10^{-4}$).

^bStop/frameshift mutations were all unknown to the ExAC database v.0.3.1.

Proposed classification of *SORL1* variants

SORL1 variants with CADD > 30 and ExAC-MAF < 1 × 10⁻⁴ are associated with a strong increased AD risk (OR = 10.9; 95% CI 4.6–25.7; $P = 1.8 \times 10^{-11}$) (Table 3), but the subset of protein-truncating *SORL1* variants ($n = 13$) occurred exclusively in AD cases, and they were novel to ExAC (OR = inf; 95% CI 5.2–inf; $P = 2.5 \times 10^{-6}$). This suggests that protein-truncating *SORL1* variants may be pathogenic (Table 3, classification: pathogenic). The subset of non-truncating missense mutations with CADD > 30 and ExAC-MAF < 1 × 10⁻⁴ accounted for a >7-fold increased AD risk (OR = 7.1, 95% CI 2.9–17.4, 8.6×10^{-7}), suggesting that variants with these characteristics are strong risk factors for AD (Table 3, classification: likely pathogenic).

Variants predicted to be mildly-moderately damaging (CADD 10–30) with ExAC-MAF < 1 × 10⁻⁵ were associated with >2-fold increased AD risk (OR = 2.4; 95% CI 1.2–4.6; $P = 7.7 \times 10^{-3}$), suggesting that variants with these characteristics are risk factors for AD and some might be pathogenic (Table 3, classification: possibly pathogenic). *SORL1* variants with ExAC-MAF < 1 × 10⁻⁴, including those classified pathogenic or likely pathogenic, did not concentrate in specific *SORL1* protein domains (Figure 3).

When we focus on the more common *SORL1* variants with CADD > 30, with ExAC-MAF > 10⁻⁴, we found that despite their high CADD values, they were not associated with increased AD risk (OR = 0.73; 95% CI 0.6–1.0; $P = 0.99$) (Table 3, classification: most

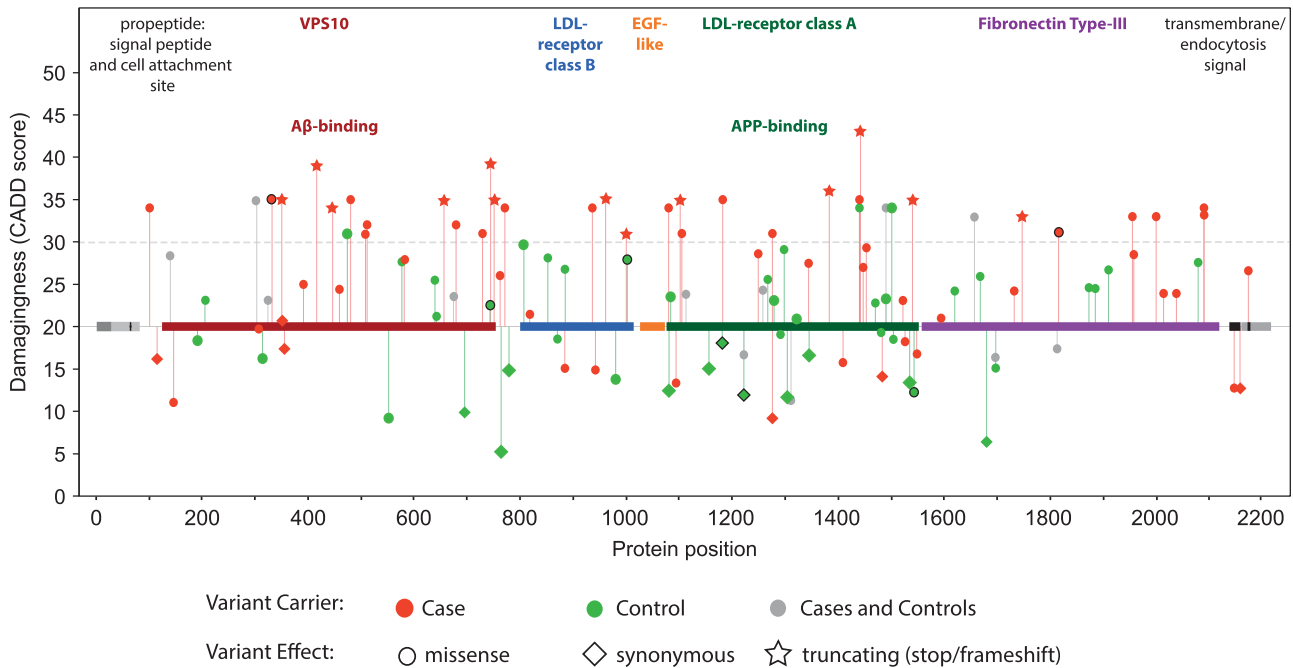


Figure 3 Protein position of *SORL1* variants with $MAF < 1 \times 10^{-4}$ in ExAC database. One hundred and twenty-one coding variants with $ExAC-MAF < 1 \times 10^{-4}$ were detected the combined analysis of 5101 subjects (1895 cases and 3206 controls). Each symbol represents one case carrier (red) or control (green) carrier. Protein domains are depicted on the $CADD=20$ level, variants with $CADD$ scores between 20 and 30 are considered 'moderately damaging' and variants with $CADD$ scores >30 were considered 'strongly damaging'. Markers outlined in black represent variants that were detected in multiple cases or in multiple controls.

likely not pathogenic). For *SORL1* variants observed more often in the ExAC database than $ExAC-MAF > 10^{-5}$ with $CADD$ values <30 , we found no association with AD risk (Table 3, classification: likely benign). Likewise, variants with $CADD$ 0–10, regardless of their rarity, were not associated with increase AD risk (Table 3, classification: benign).

Our findings lead us to propose the following five *SORL1* variant subtypes:

Pathogenic. Truncating *SORL1* variants.

Likely pathogenic. *SORL1* variants predicted extremely damaging ($CADD >30$) and extremely rare ($MAF < 1 \times 10^{-4}$ in the publically available ExAC database v.0.3.1³¹).

Uncertain significance. **Possibly pathogenic:** variants predicted mildly to moderately damaging ($CADD$ 10–30) which are novel or reported only once in the ExAC database ($ExAC-MAF < 1 \times 10^{-5}$).

Most likely not pathogenic: variants predicted extremely damaging ($CADD >30$) that are observed more commonly in the ExAC database ($ExAC-MAF \geq 1 \times 10^{-4}$).

Likely benign. *SORL1* variants predicted mildly- moderately damaging ($CADD$ 10–30) that are reported more than once in the ExAC database ($ExAC-MAF \geq 1 \times 10^{-5}$).

Benign. *SORL1* variants predicted not damaging ($CADD$ 0–10) regardless of their rareness.

Application of *SORL1* variant pathogenicity screen

We applied our classification approach to the 17 *SORL1* prioritized variants detected in a family based analysis.⁴ These variants were enriched in members from 87 Caribbean Hispanic families affected with late onset AD, compared with 498 age-matched controls.

Vardarajan *et al.* identified three truncating deletions unknown to the ExAC database. Our approach to screen for variant pathogenicity classified these variants to be 'pathogenic' and indeed these variants were detected exclusively in families affected with AD. Furthermore, 13 variants are classified 'likely benign', they had $CADD$ scores <30 and occur more than once in ExAC; indeed, these variants were detected both in the affected families and non-affected families. One variant with $CADD$ score 34 and $ExAC-MAF > 0.01$ occurred both in affected families and in unaffected controls; in accordance, our approach classified it to be 'most likely not pathogenic'.

Likewise, we applied our classification strategy to *SORL1* variants reported by Nicolas *et al.*⁵ They studied 24 rare variants (sample $MAF < 0.01$) that were predicted deleterious by SIFT and Polyphen in 484 AD cases, mostly with family history of AD, and 498 controls. Of these variants, 15 were novel to the ExAC database: 8 truncating variants and 7 missense variants with $CADD >30$ that were classified as 'pathogenic' and 'likely pathogenic' variants, respectively; indeed, these were seen exclusively in cases. Another case carried a variant with $CADD$ score 35 with $ExAC-MAF 6 \times 10^{-5}$ classified to be 'likely pathogenic'; indeed, this variant was also detected in two cases in the Verheijen sample. Furthermore, Nicolas *et al.* detected a *SORL1* variant unknown to ExAC with $CADD$ score 32, which was located within the VPS10 domain. This 'likely pathogenic' variant was found to disturb the binding of $A\beta$ for lysosomal degradation¹² and Pottier *et al.*³ found that this variant segregated with disease. Nicolas *et al.* also detected a variant with $ExAC-MAF 0.003$ and $CADD$ score 25.5, which was classified to be 'likely benign': indeed, it occurred equally in cases and controls. Finally, Nicolas *et al.* identified seven 'possibly pathogenic' variants that were novel to ExAC with $CADD$ scores ranging between 26.7 and 29.4, all occurred in AD cases. The evidence for the association with AD of variants with these characteristics is

relatively low, suggesting that further research into the pathogenicity of these variants is necessary.

DISCUSSION

Protein truncating and rare pathogenic missense variants in the *SORL1* gene associate with AD

It is clear that nonfunctional *SORL1* associates with AD, but a comprehensive set of characteristics that defines the associated genetic variants and their impact has been lacking. Therefore, the ‘need for pathogenicity assays’ has been raised to aid with the clinical interpretation of *SORL1* variants.⁶ Here, we analyzed 181 unique *SORL1* variants detected in a large sample of 1895 cases and 3206 controls and we propose that *SORL1* variant-pathogenicity can be classified according to the combination of two independent variant characteristics: the predicted level of variant damagingness and the level of variant-rareness.

Our findings indicate that stop-gain and frameshift mutations occurred exclusively in cases, suggesting that variants leading to premature disruption of *SORL1* transcription are highly penetrant. This supports previous findings that loss of one copy of *SORL1* (i.e., haploinsufficiency) is causal to AD.⁶ Thus far, such high impact on AD is observed only for variants in *PSEN1/2* and *APP*, which are associated with familial AD.²⁵ Furthermore, our findings indicate that variants novel to the ExAC database with a CADD score >30 are associated with a significant 12-fold increased AD risk, which is comparable to the effect of *APOE-ε4* homozygosity.²⁵ In line with the increased risk, we found suggestive evidence that pathogenic *SORL1* variants lead to an earlier age at onset.

Although variants are individually rare, 2% of the AD cases (and <0.2% of the controls) in our analysis carried a *SORL1* variant with these characteristics. By comparison, variants in the classical AD genes *PSEN1*, *PSEN2* and *APP* collectively explain <1% of AD cases.²⁵ We propose therefore that in clinical practice, rare pathogenic *SORL1* mutations should be considered next to *PSEN1*, *PSEN2* and *APP*.

Moderately damaging variants in the VPS10 domain might be associated with AD

Pathogenic variants occurred throughout the *SORL1* gene without preference to a specific functional domain. In our discovery analysis, moderately damaging variants in the VPS10 domain were detected only in four cases but not in older controls. In contrast, moderately damaging variants in the LDL-receptor A and fibronectin domains occurred only in control subjects. However, we could not confirm these findings in the replication or combined sample, possibly due to the many young control subjects in the replication sample who might still develop disease at a later age. In the future, larger samples will have to clarify whether or not moderately damaging variants in the Aβ-binding VPS10 domain might dangerously affect *SORL1* function.

Our results indicate that some *SORL1* variants with lower CADD scores may hold some pathogenicity when they are extremely rare, but the effect size is only two-fold and the evidence for this is not as strong. On the other hand, we found no evidence for pathogenicity for common variants, even variants with CADD scores >30. Risk increases were independent of gender and we detected no evidence for synergy between disrupted *SORL1* function and carrying the *APOE-ε4* genotype.

Five *SORL1* variant subtypes

For the clinical interpretation of *SORL1* variant pathogenicity based on ExAC-MAF and CADD scores, we propose five *SORL1* variant subtypes according to the five-class system of variant pathogenicity

supported by the ACMG³² and ACGS.³³ When we applied our strategy to *SORL1* variants reported by the independent studies of Vardarajan *et al.*⁴ and Nicholas *et al.*,⁵ variants were classified according to their occurrence in cases and controls. Even though the classification strategy presented here is based on two large samples, additional research is necessary to determine the exact risk of individual variants. We caution that genetic context might influence variant pathogenicity: for example, one possibly pathogenic *SORL1* variant (CADD score 23.6, ExAC-MAF <1 × 10⁻⁵) was found to segregate with disease and increase AD risk in a family with several generations of *APOE-ε4* homozygosity.³⁹ It is likely that classification will be refined as larger samples with sequencing data become available. Lastly, this classification is based on evidence from populations with European ancestry and should be replicated in populations with other ethnic backgrounds.

Pathogenic *SORL1* variants are rare

We find that truncating *SORL1* variants are pathogenic: all 24 truncating variants collectively reported across the previous studies by Verheijen *et al.*, Nicolas *et al.*, Vardarajan *et al.* and this present study occurred exclusively in AD cases and were unknown to the ExAC database. Likewise, across these studies, >70% of all likely pathogenic variants (missense variants with CADD score >30) were unknown to ExAC. This suggests that the increased pathogenicity of extremely rare variants may explain part of the ‘missing heritability’ that remained undetected in genetic association studies such as GWAS, which test the association of common variants with disease.

In concordance with this, evidence is mounting that especially extremely rare mutations are the major contributors to the development of disease. In a sequencing analysis of 202 drug-targeted genes in 14 002 persons, 74% of the detected mutations occurred only in one or two subjects, indicating that mutations that associated with disease are abundant, but mostly very rare.⁴⁰ Furthermore, Fu *et al.*⁴¹ found that more than half of all SNPs detected in exomes from 6515 individuals were, in fact, singletons. They found that 86% of detected damaging variants arose very recently in the population, which partly explains the restricted propagation of the variant in the population, that is, the variant rarity relative to common (old) variants.⁴¹

The rarity of the variants also suggests that natural selection pressure eliminates pathogenic *SORL1* variants from the population. As AD onset occurs well after the reproductive phase, we might expect that variants associated with AD would not be under influence of selection pressure. Therefore, it is surprising that pathogenic *SORL1* variants are not propagated in the population. The rarity of harmful *SORL1* variants suggests that *SORL1* function may not be restricted to the maintenance of cells in the brain and that disturbed *SORL1* function might affect reproductive success and/or individual health far before the age of AD onset.

CONCLUSION

With the increasing availability of whole-exome sequencing in clinical practice, it is possible to detect highly personal exonic variants in *SORL1*. We characterized *SORL1* variants based on variant frequency and damagingness and we suggest five variant subtypes ranging from pathogenic to benign. Our findings suggest that in the clinic, pathogenic *SORL1* variants should be considered in personalized AD risk assessments alongside *APOE*, *PSEN1*, *PSEN2*, and *APP*.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We thank all study participants and all personnel involved in data collection for the contributing studies. This work was carried out on the Dutch national e-infrastructure with the support of SURF Cooperative. Specific acknowledgments go out to Lisbet Sornsen-de Koste at the Amsterdam site and Andrea Vermeij-Verdoold at the Rotterdam site who were involved in the Sanger validation experiments. We thank all personnel involved with the biobanking participant-biomaterials for each cohort, Mr Pascal Arp, Ms Mila Jhamai, Mr Marijn Verkerk and Dr Robert Kraaij for their help in creating the RS-Exome Sequencing database, Dr Hans Gille, Dr Erik Sistermans and Ms Ingrid Bakker of the VUmc Amsterdam for their help in creating the ADC-VUmc and 100-plus Study cohorts. This work was supported by Stichting Alzheimer Nederland (WE.09-2014-06, WE.05-2010-06); Stichting Dioraphte; Internationale Stichting Alzheimer Onderzoek (#11519); JPN-D-PERADES (ZonMw 733051022); Centralized Facility for Sequence to Phenotype analyses (ZonMW 9111025); Netherlands Consortium for Healthy Aging (NCHA 050-060-810); Biobanking and Biomolecular Research Infrastructure Netherlands (BBMRI-NL CP2010-41); Netherlands Genomics Initiative (NGI)/NWO.

- 1 Alzheimer's Association: 2016 Alzheimer's disease facts and figures. *Alzheimers Dement* 2016; **12**: 459–509.
- 2 Gatz M, Reynolds CA, Fratiglioni L *et al*: Role of genes and environments for explaining Alzheimer disease. *Arch Gen Psychiatry* 2006; **63**: 168–174.
- 3 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.
- 4 Vardarajan BN, Zhang Y, Lee JH *et al*: Coding mutations in SORL1 and Alzheimer disease. *Ann Neurol* 2015; **77**: 215–227.
- 5 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.
- 6 Verheijen J, Van den Bossche T, van der Zee J *et al*: A comprehensive study of the genetic impact of rare variants in SORL1 in European early-onset Alzheimer's disease. *Acta Neuropathol* 2016; **132**: 213–224.
- 7 Andersen OM, Rudolph IM, Willnow TE: Risk factor SORL1: from genetic association to functional validation in Alzheimer's disease. *Acta Neuropathol* 2016; **132**: 653–665.
- 8 Willnow TE, Andersen OM: Sorting receptor SORLA—a trafficking path to avoid Alzheimer disease. *J Cell Sci* 2013; **126**: 2751–2760.
- 9 Mehmedbasic A, Christensen SK, Nilsson J *et al*: SorLA complement-type repeat domains protect the amyloid precursor protein against processing. *J Biol Chem* 2015; **290**: 3359–3376.
- 10 Andersen OM, Schmidt V, Spoelgen R *et al*: Molecular dissection of the interaction between amyloid precursor protein and its neuronal trafficking receptor SorLA/LR11. *Biochemistry* 2006; **45**: 2618–2628.
- 11 Hermey G: The Vps10p-domain receptor family. *Cell Mol Life Sci* 2009; **66**: 2677–2689.
- 12 Caglayan S, Takagi-Niidome S, Liao F *et al*: Lysosomal sorting of amyloid-beta by the SORLA receptor is impaired by a familial Alzheimer's disease mutation. *Sci Transl Med* 2014; **6**: 223ra220.
- 13 Scherzer CR, Offe K, Gearing M *et al*: Loss of apolipoprotein E receptor LR11 in Alzheimer disease. *Arch Neurol* 2004; **61**: 1200–1205.
- 14 Andersen OM, Reiche J, Schmidt V *et al*: Neuronal sorting protein-related receptor sorLA/LR11 regulates processing of the amyloid precursor protein. *Proc Natl Acad Sci USA* 2005; **102**: 13461–13466.
- 15 Offe K, Dodson SE, Shoemaker JT *et al*: The lipoprotein receptor LR11 regulates amyloid beta production and amyloid precursor protein traffic in endosomal compartments. *J Neurosci* 2006; **26**: 1596–1603.
- 16 Spoelgen R, von Arnim CA, Thomas AV *et al*: Interaction of the cytosolic domains of sorLA/LR11 with the amyloid precursor protein (APP) and beta-secretase beta-site APP-cleaving enzyme. *J Neurosci* 2006; **26**: 418–428.
- 17 Rogava 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.
- 18 Bettens K, Brouwers N, Engelborghs S, De Deyn PP, Van Broeckhoven C, Sleegers K: SORL1 is genetically associated with increased risk for late-onset Alzheimer disease in the Belgian population. *Hum Mutat* 2008; **29**: 769–770.

- 19 Caglayan S, Bauerfeind A, Schmidt V *et al*: Identification of Alzheimer disease risk genotype that predicts efficiency of SORL1 expression in the brain. *Arch Neurol* 2012; **69**: 373–379.
- 20 Tan EK, Lee J, Chen CP, Teo YY, Zhao Y, Lee WL: SORL1 haplotypes modulate risk of Alzheimer's disease in Chinese. *Neurobiol Aging* 2009; **30**: 1048–1051.
- 21 Kimura R, Yamamoto M, Morihara T *et al*: SORL1 is genetically associated with Alzheimer disease in a Japanese population. *Neurosci Lett* 2009; **461**: 177–180.
- 22 Liu F, Ikram MA, Janssens AC *et al*: A study of the SORL1 gene in Alzheimer's disease and cognitive function. *J Alzheimer's Dis* 2009; **18**: 51–64.
- 23 Reitz C, Cheng R, Rogava E *et al*: Meta-analysis of the association between variants in SORL1 and Alzheimer disease. *Arch Neurol* 2011; **68**: 99–106.
- 24 Lambert J, Ibrahim-Verbaas C, Harold D *et al*: Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet* 2013; **45**: 1452–1458.
- 25 Van Cauwenbergh C, Van Broeckhoven C, Sleegers K: The genetic landscape of Alzheimer disease: clinical implications and perspectives. *Genet Med* 2016; **18**: 421–430.
- 26 Corder EH, Saunders AM, Strittmatter WJ *et al*: Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science* 1993; **261**: 921–923.
- 27 Goate A, Chartier-Harlin MC, Mullan M *et al*: Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer's disease. *Nature* 1991; **349**: 704–706.
- 28 Schellenberg GD, Bird TD, Wijsman EM *et al*: Genetic linkage evidence for a familial Alzheimer's disease locus on chromosome 14. *Science* 1992; **258**: 668–671.
- 29 Levy-Lahad E, Wasco W, Poorkaj P *et al*: Candidate gene for the chromosome 1 familial Alzheimer's disease locus. *Science* 1995; **269**: 973–977.
- 30 Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J: A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* 2014; **46**: 310–315.
- 31 Lek M, Karczewski K, Minikel E *et al*: Analysis of protein-coding genetic variation in 60,706 humans. *bioRxiv* 2015; **536**: 285–291.
- 32 Richards S, Aziz N, Bale S *et al*: Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* 2015; **17**: 405–423.
- 33 Wallis Y, Payne S, McNulty C *et al*: Practice Guidelines for the Evaluation of Pathogenicity and the Reporting of Sequence Variants in Clinical Molecular Genetics: Association for Clinical Genetic Science, 2013.
- 34 Hofman A, Brusselle GG, Darwish Murad S *et al*: The Rotterdam Study: 2016 objectives and design update. *Eur J Epidemiol* 2015; **30**: 661–708.
- 35 van der Flier WM, Pijnenburg YA, Prins N *et al*: Optimizing patient care and research: the Amsterdam Dementia Cohort. *J Alzheimers Dis* 2014; **41**: 313–327.
- 36 McLaren W, Gil L, Hunt SE *et al*: The Ensembl variant effect predictor. *Genome Biol* 2016; **17**: 122.
- 37 Lee S, Abecasis GR, Boehnke M, Lin X: Rare-variant association analysis: study designs and statistical tests. *Am J Hum Genet* 2014; **95**: 5–23.
- 38 R Core Team: *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing, 2013.
- 39 Louwersheimer E, Cohn-Hokke PE, Pijnenburg YA *et al*: Rare genetic variant in SORL1 may increase penetrance of Alzheimer's disease in a family with several generations of APOE-varepsilon4 homozygosity. *J Alzheimer's Dis* 2017; **56**: 63–74.
- 40 Nelson MR, Wegmann D, Ehm MG *et al*: An abundance of rare functional variants in 202 drug target genes sequenced in 14,002 people. *Science* 2012; **337**: 100–104.
- 41 Fu W, O'Connor TD, Jun G *et al*: Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. *Nature* 2013; **493**: 216–220.



This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/4.0/>

© The Author(s) 2017

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