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

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters

| Published Version | doi:10.1038/ejhg.2017.87 |
| Citable link    | http://nrs.harvard.edu/urn-3:HUL.InstRepos:34375312 |
| Terms of Use    | This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA |
Characterization of pathogenic SORL1 genetic variants for association with Alzheimer’s disease: a clinical interpretation strategy

Henne Holstege1,2,3, Sven J van der Lee1,4, Marc Hulsmann3, Tsz Hang Wong5, Jeroen GJ van Rooij3,6, Marjan Weiss7, Eva Louwersheimer1, Frank J Wolters4, Najaf Amin4, André G Uitterlinden4,6, Albert Hofman3,8, M Arfan Ikram1,5,9, John C van Swieten1,5, Hanne Meijers-Heijboer2, Wiesje M van der Flier1,10, Marcel JT Reinders3, Cornelia M van Duijn4,11 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 = 5×10−9). Protein-truncating SORL1 mutations were all unknown to ExAC and occurred exclusively in AD cases. More common SORL1 variants (ExAC-MAF ≥ 1×10−5) 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 Aβ for endosomal inclusion and sorting for lysosomal degradation.11,12 Therefore, impaired SORL1 function is associated both with disturbed APP processing and disturbed Aβ 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 Cauwenberge et al.25). In contrast, recent studies reported that rare SORL1 variants are associated with a five-fold increased risk for...
early onset AD.\textsuperscript{5,6} These studies suggest that the effect on AD risk of these rare \textit{SORL1} variants is comparable to that of carrying the e4 allele of Apolipoprotein-E gene (\textit{APOE}), the most important common risk factor for AD; homozygous and heterozygous \textit{APOE-e4} 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-e4 carriers.\textsuperscript{25} Furthermore, recent targeted sequencing studies identified rare pathogenic \textit{SORL1} mutations that segregated with disease in families with familial AD and late onset AD.\textsuperscript{3,4} These findings suggest that specific \textit{SORL1} variants are causal for AD, with effects comparable to single mutations in the \textit{APP} gene,\textsuperscript{27} or the presenilin-1 and presenilin-2 genes (\textit{PSEN1}, \textit{PSEN2}),\textsuperscript{28,29} that are associated with an autosomal dominant inheritance pattern of AD. However, it is currently unclear which specific \textit{SORL1} variants are major risk factors for AD and which can be considered benign. This raises the need for a strategy to determine \textit{SORL1} variant pathogenicity.

Ideally, one would like to determine penetrance for each detected \textit{SORL1} variant in multiple large and informative families. However, the rareness of the \textit{SORL1} variants that were thus far associated with AD complicates such efforts.\textsuperscript{3,5} Therefore, a more feasible approach might be to distinguish between pathogenic and non-pathogenic variants using independent variant characteristics that might be associated with \textit{SORL1} variant pathogenicity. In this work we explored the contribution on disease outcome of (i) the functional protein domain affected by the \textit{SORL1} variant; (ii) the minor allele frequency (MAF) of the \textit{SORL1} variant in the population; and (iii) predicted damagingness of the \textit{SORL1} variant on the basis of sequence context.

To predict the damagingness of \textit{SORL1} variants we annotated them with the CADD score,\textsuperscript{30} a novel functional annotation tool that allows determining its CADD score, such that the CADD score and MAF can be used as independent determinants of variant pathogenicity. For a schematic overview of the analysis setup see Figure 1.

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,\textsuperscript{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 \textit{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.\textsuperscript{6} and we performed a combined analysis including data from both studies. Based on these results, we suggest five \textit{SORL1} variant subtypes according to the five-class system of variant pathogenicity supported by the American College of Medical Genetics and Genomics (ACMG)\textsuperscript{32} and Association for Clinical Genetic Science (ACGS).\textsuperscript{33} Lastly, we validated our classification strategy by applying it to \textit{SORL1} variants described in two additional independent publications.\textsuperscript{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.

---

**Discovery**

115 unique \textit{SORL1} variants

\begin{table}[h]
\centering
\begin{tabular}{|c|c|}
\hline
Case & Control \hline
640 AD cases & 1,268 controls \hline
\end{tabular}
\end{table}

**Replication**

103 unique \textit{SORL1} variants

\begin{table}[h]
\centering
\begin{tabular}{|c|c|}
\hline
Case & Control \hline
5,193 subjects & 5,397 AD cases & 5,294 controls \hline
\end{tabular}
\end{table}

**Classification**

Five classes of variant pathogenicity supported by ACMG and ACGS

\begin{table}[h]
\centering
\begin{tabular}{|c|}
\hline
Case \hline
5,101 subjects \hline
5,097 AD cases & 5,094 controls \hline
\end{tabular}
\end{table}

**Validation**

Application of \textit{SORL1} variant pathogenicity screen in published datasets\textsuperscript{4,6}

---

**Figure 1** Flowchart of \textit{SORL1} variant pathogenicity analysis. *\textit{SORL1} variants in the independent replication data set were reported by Verheijen et al.\textsuperscript{6} Pathogenicity screen was applied to \textit{SORL1} variants reported by Vardarajan et al.\textsuperscript{4} and Nicolas et al.\textsuperscript{5}

**Discovery sample.** The exome collection was assembled from four Dutch studies: (i) the Rotterdam Study\textsuperscript{34} contributed 250 AD cases (median age at disease onset 84.5 ± 6.6, 71.2% females) and 1204 controls (median age at last visit of 82.4 ± 6.8, 54.3% females), (ii) The Amsterdam Dementia Cohort (ADCV-Umc) contributed 320 AD (median age of 58.4 ± 6.5, 51.9% females),\textsuperscript{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, \textit{PSEN1} and \textit{PSEN2} genes. Combined data included (i) AD status, (ii) \textit{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.\textsuperscript{6} They reported 103 \textit{SORL1} variants in 1255 European early onset AD cases and 1938 age-matched controls. Rare \textit{SORL1} variants that occurred in either cases or controls were given per subject (including gender and age, but not \textit{APOE} genotype), and no two variants occurred in the same subject; the number of case- and control-carriers were given for rare \textit{SORL1} variants that occurred in both cases and controls; common \textit{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 ADCV-Umc cohort and the 100-plus Study cohort were captured with the Nimblegen SeqCap EZ Exome capture kit v3. For \textit{SORL1} variant calling, we used the intersection between these capture kits in the \textit{SORL1} gene: 93.2% of
exons 1–47, 2.7% of exon 48. DNA from all samples was prepared with the Illumina TrueSeq 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 x 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/DOE10.1038/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 1.3.30 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.

RESULTS

Variant annotation

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/DOE10.1038/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).

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 x 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 x 10⁻⁶). Notably, all five truncating mutations (stop-gain/frameshift) were singletons in our sample, and their carriers all developed AD (P = 1.6 x 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 x 10⁻⁴). 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 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 x 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 x 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.5 In this data set, singleton variants...
with CADD > 30 associated with a 14-fold increased AD risk (OR = 14.1; 95% CI 3.3–60.8; \( P = 3.5 \times 10^{-3} \)) and all eight stop-gain/frame-shift 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 1 Discovery analysis: 115 SORL1 variants in 640 cases and 1268 controls

<table>
<thead>
<tr>
<th>CADD score</th>
<th>N-variants in burden test</th>
<th>With at least one variant</th>
<th>With at least one variant</th>
<th>OR (95% CI)</th>
<th>P-value a</th>
</tr>
</thead>
<tbody>
<tr>
<td>cases</td>
<td>controls</td>
<td>Total 640</td>
<td>Total 1268</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All variants</td>
<td>100</td>
<td>83</td>
<td>140</td>
<td>1.2 (0.9–1.6)</td>
<td>0.23</td>
</tr>
<tr>
<td>SIFT/Polyphen damaging</td>
<td>36</td>
<td>40</td>
<td>44</td>
<td>1.9 (1.2–2.9)</td>
<td>7.2 \times 10^{-3}</td>
</tr>
<tr>
<td>CADD 0–20</td>
<td>55</td>
<td>42</td>
<td>94</td>
<td>0.9 (0.6–1.3)</td>
<td>0.59</td>
</tr>
<tr>
<td>CADD 20–30</td>
<td>26</td>
<td>28</td>
<td>43</td>
<td>1.3 (0.78–2.1)</td>
<td>0.34</td>
</tr>
<tr>
<td>CADD &gt; 30</td>
<td>19</td>
<td>15</td>
<td>8</td>
<td>4.0 (1.7–9.0)</td>
<td>9.9 \times 10^{-4}</td>
</tr>
</tbody>
</table>

**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) \( P = 0.23 \)
  - SIFT/Polyphen damaging: 36, 40, 44, 1.9 (1.2–2.9) \( P = 7.2 \times 10^{-3} \)
  - CADD 0–20: 55, 42, 94, 0.9 (0.6–1.3) \( P = 0.59 \)
  - CADD 20–30: 26, 28, 43, 1.3 (0.78–2.1) \( P = 0.34 \)
  - CADD > 30: 19, 15, 8, 4.0 (1.7–9.0) \( P = 9.9 \times 10^{-4} \)

**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) \( P = 0.54 \)
  - 20–30: 1, 48, 103, 1.0 (0.7–1.3) \( P = 0.83 \)
  - 430: 1, 17, 44, 0.8 (0.5–1.3) \( P = 0.29 \)

- **MAF < 0.001** (variants detected in 3–38 subjects in this sample)
  - 0–20: 7, 27, 47, 1.2 (0.7–1.9) \( P = 0.54 \)
  - 20–30: 3, 18, 24, 1.5 (0.8–2.9) \( P = 0.21 \)
  - 430: 1, 1, 3, 0.7 (0.1–5.5) \( P = 0.73 \)

- **MAF < 0.0005** (variants detected in 2–3 subjects in this sample)
  - 0–20: 12, 6, 21, 0.7 (0.3–1.4) \( P = 0.33 \)
  - 20–30: 5, 3, 9, 0.7 (0.2–2.4) \( P = 0.57 \)
  - 430: 1, 1, 3, 0.7 (0.1–5.2) \( P = 0.70 \)

- **MAF < 0.0005 (singletons in this sample)**
  - 0–20: 36, 9, 26*, 0.7 (0.4–1.4) \( P = 0.31 \)
  - 20–30: 18, 7, 11, 1.3 (0.5–3.3) \( P = 0.65 \)
  - 430: 4, 3, 1, 19.8 (4.0–32.1) \( P = 4.9 \times 10^{-6} \)

**C. Variant stratification of singleton variants according to protein domain and CADD**

- **VPS10**
  - 0–20: 5, 1, 4, 0.6 (0.1–3.6) \( P = 0.54 \)
  - 20–30: 6, 6, 0, 19.3 (3.6–105.2) \( P = 6.1 \times 10^{-4} \)
  - 430: 4, 3, 1, 6.6 (0.8–53.1) \( P = 7.6 \times 10^{-2} \)

- **LDL A and B**
  - 0–20: 10, 3, 7, 0.9 (0.2–3.2) \( P = 0.81 \)
  - 20–30: 4, 0, 4, 0.22 (0.0–1.7) \( P = 0.14 \)
  - 430: 8, 7, 1, 10.8 (2.5–46.7) \( P = 1.5 \times 10^{-3} \)

- **Fibronectin**
  - 0–20: 3, 0, 3, 0.2 (0.0–2.5) \( P = 0.23 \)
  - 20–30: 5, 0, 5, 0.2 (0.0–1.4) \( P = 0.11 \)
  - 430: 4, 4, 0, 20.5 (2.6–164.5) \( P = 4.5 \times 10^{-3} \)

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 B0R; 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.
Table 2 Combined analysis: 181 unique SORL1 variants in 1895 cases and 3206 controls

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

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

\( a \)In the replication sample, we were not able to account for the overlap when one subject carried multiple variants with the same MAF.

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

\( c \)For common variants derived form 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 × 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 2–12× 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 × 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 × 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 × 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; \( P = 0.66 \)).
### 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 4.6–25.9; P = 1.8 × 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 × 10⁻⁶). This suggests that protein-truncating SORL1 variants may be pathogenic (Table 3, classification: 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 × 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 likely not pathogenic).

---

**Table 3 Clinical selection criteria of variants**

<table>
<thead>
<tr>
<th>Variant selection criteria</th>
<th>CADD</th>
<th>ExAC-MAF v.0.3.1</th>
<th>N-variants in burden test</th>
<th>Total 1895</th>
<th>Total 3206</th>
<th>OR (95% CI)</th>
<th>P-value a (unadjusted)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A. Effect on AD risk</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maximum effect size</td>
<td>&gt; 30</td>
<td>&lt; 1 × 10⁻⁵</td>
<td>29</td>
<td>28</td>
<td>4</td>
<td>12.0 (4.2–34.3)</td>
<td>5.0 × 10⁻⁹</td>
</tr>
<tr>
<td>Maximum evidence for effect</td>
<td>&gt; 30</td>
<td>&lt; 1 × 10⁻⁴</td>
<td>38</td>
<td>38</td>
<td>6</td>
<td>10.9 (4.6–25.9)</td>
<td>1.8 × 10⁻¹¹</td>
</tr>
<tr>
<td><strong>B. Suggested variant subtypes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pathogenic</td>
<td>&gt; 30 (truncating)</td>
<td>&lt; 1 × 10⁻⁵</td>
<td>13</td>
<td>13</td>
<td>0</td>
<td>inf (5.2–inf)</td>
<td>2.5 × 10⁻⁶</td>
</tr>
<tr>
<td>Likely pathogenic</td>
<td>&gt; 30 (missense)</td>
<td>&lt; 1 × 10⁻⁴</td>
<td>25</td>
<td>25</td>
<td>6</td>
<td>7.1 (2.9–17.4)</td>
<td>8.6 × 10⁻⁷</td>
</tr>
<tr>
<td>Uncertain significance</td>
<td>&gt; 30</td>
<td>&lt; 1 × 10⁻⁵</td>
<td>33</td>
<td>21</td>
<td>15</td>
<td>2.4 (1.2–4.6)</td>
<td>7.7 × 10⁻³</td>
</tr>
<tr>
<td>Possibly pathogenic</td>
<td>10–30</td>
<td>&lt; 1 × 10⁻⁵</td>
<td>33</td>
<td>21</td>
<td>15</td>
<td>2.4 (1.2–4.6)</td>
<td>7.7 × 10⁻³</td>
</tr>
<tr>
<td>Most likely not pathogenic</td>
<td>&gt; 30</td>
<td>&gt; 1 × 10⁻⁴</td>
<td>6</td>
<td>76</td>
<td>173</td>
<td>0.73 (0.6–1.0)</td>
<td>0.99</td>
</tr>
<tr>
<td>Likely benign</td>
<td>10–30</td>
<td>&gt; 1 × 10⁻⁵</td>
<td>70</td>
<td>1,895</td>
<td>3,206</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Benign</td>
<td>0–10</td>
<td>0–1</td>
<td>30</td>
<td>1,895</td>
<td>3,206</td>
<td>1.0</td>
<td>1.0</td>
</tr>
</tbody>
</table>

**a**Associations that are significant after multiple testing correction are shown in bold (P < 7.1 × 10⁻⁴).

**b**Stop/frameshift mutations were all unknown to the ExAC database v.0.3.1.
Characterization of pathogenic SORL1 genetic variants

H Holstege et al

likely not pathogenic). For SORL1 variants observed more often in the ExAC database than ExAC-MAF $<1\times10^{-5}$, we found no association with AD risk (Table 3, classification: likely benign). Likewise, variants with ExAC-0.01 occurred both in affected families and in unaffected controls; in accordance, our approach classified it to be 'most likely not pathogenic'.

Likely pathogenic. SORL1 variants predicted extremely damaging (CADD $>30$) and extremely rare (MAF $<1\times10^{-4}$) are not associated with increase AD risk (Table 3, classification: benign).

Likely benign. SORL1 variants predicted mildly- moderately damaging (CADD 10–30) which are novel or reported only once in the ExAC database (ExAC-MAF $<1\times10^{-5}$).

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

Applicatin 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 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.

Figure 3 Protein position of SORL1 variants with MAF $<1\times10^{-4}$ in ExAC database. One hundred and twenty-one coding variants with ExAC-MAF $<1\times10^{-4}$ 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 projection: likely pathogenic variants predicted not damaging (CADD 0–4 in ExAC database). One hundred and twenty-one coding variants with ExAC-MAF $<1\times10^{-4}$ 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 projection: likely pathogenic variants predicted not damaging (CADD 0–4 in ExAC database).
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. 

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. 

Thus far, such high impact on AD is observed only for variants in PSEN1/2 and APP, which are associated with familial AD. 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. 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. 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 ACMC and ACGS. When we applied our strategy to SORL1 variants reported by the independent studies of Vardarajan et al. and Nicholas et al., 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. 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. Furthermore, Fu et al. 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.

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 Sørensen-de Koste at the Amsterdam site and Andrea Vermnej-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]; JPND-PERADES (ZonMW 73305102)); 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).


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