Published in final edited form as:

Neurobiol Aging. 2018 November ; 71: 266.e11–266.e24. doi:10.1016/j.neurobiolaging.2018.06.021.

An alternative transcript of the Alzheimer's disease risk gene *SORL1* encodes a truncated receptor

Jenny Blechingberg^a, Annemarie Svane Aavild Poulsen^a, Mads Kjølby^{a,b}, Giulia Monti^a, Mariet Allen^c, Anne Kathrine Ivarsen^a, Sarah J. Lincoln^c, Gangadaar Thotakura^c, Christian B. Vægter^a, Nilüfer Ertekin-Taner^c, Anders Nykjær^{a,d,e,f}, Olav M. Andersen^{a,*}

^aDanish Research Institute of Translational Neuroscience (DANDRITE) Nordic-EMBL Partnership, Aarhus, Denmark

^bDanish Diabetes Academy, Novo Nordisk Foundation, Aarhus University, Aarhus, Denmark

^cDepartment of Neuroscience, Mayo Clinic, FL, USA

^dThe Lundbeck Foundation Research Center MIND, Aarhus, Denmark

^eThe Danish Research Foundation Center PROMEMO, Department of Biomedicine, Aarhus, Denmark

^fDepartment of Neurosurgery, Aarhus University Hospital, Aarhus, Denmark

Abstract

SORL1 encodes a 250-kDa protein named sorLA, a functional sorting receptor for the amyloid precursor protein (APP). Several single nucleotide polymorphisms of the gene *SORL1*, encoding sorLA, are genetically associated with Alzheimer's disease (AD). In the existing literature, *SORL1* is insufficiently described at the transcriptional level, and there is very limited amount of functional data defining different transcripts. We have characterized a *SORL1* transcript containing a novel exon 30B. The transcript is expressed in most brain regions with highest expression in the temporal lobe and hippocampus. Exon 30B is spliced to exon 31, leading to a mature transcript that encodes an 829 amino acid sorLA receptor. This receptor variant lacks the binding site for APP and is unlikely to function in APP sorting. This transcript is expressed in equal amounts in the cerebellum from AD and non-AD individuals. Our data describe a transcript that encodes a truncated sorLA receptor, suggesting novel neuronal functions for sorLA and that alternative transcription provides a mechanism for *SORL1* activity regulation.

*Corresponding author at: Department of Biomedicine, Aarhus University, Aarhus, Denmark. Tel.: (+45) 871 6778; fax: (+45) 8613-1160. o.andersen@biomed.au.dk (O.M. Andersen).

Authors' contributions:

JB, GM, MK, AKI, and ASAP initiated the project by database searches, RT-PCR analysis, and mapping of the TSS by 5' RACE. GT prepared the expression construct for sorLA-30B and did functional studies of the truncated receptor. MA and SJL measured transcripts levels in AD patients. AN, CBV, NE-T, and OMA supervised the overall project, and JB and OMA wrote the manuscript. All authors have read and approved the final manuscript.

Disclosure

N.E.-T is an unpaid consultant for Cytox.

Keywords

SorLA; *SORL1*; Alzheimer's disease risk gene; Alternative transcription; Exon 30B

1. Introduction

SorLA (sorting protein-related receptor with A-type repeats) is a multifunctional type-1 membrane receptor and member of the vacuole protein sorting 10 protein (VPS10p) family of trafficking receptors, also known as sortilins, that play important functions in neurodegenerative and psychiatric disorders (Lane et al., 2012; Willnow et al., 2008; Wilson et al., 2014). SorLA stands out from the other family members as in addition to the VPS10p domain it contains a Tyr-Trp-Thr-Asp (YWTD)-repeated 6-bladed β -propeller with an adjacent EGF domain, a cluster with 11 complement-type repeat (CR) domains, and 6 fibronectin (Fn)-type III domains (Fig. 1A). So far, several of these protein regions have been shown to engage in protein-protein interactions. Hence, the VPS10p domain can bind glia cell line-derived neurotrophic factor (GDNF) to regulate dopaminergic survival and differentiation (Glerup et al., 2013). However, best described is the interaction between the CR domains and amyloid precursor protein (APP). SorLA is a sorting receptor for APP, and many studies have demonstrated that sorLA decreases amyloid- β (A β) peptide production and deposition in vitro and in vivo (Andersen et al., 2005; Dodson et al., 2008; Fjorback et al., 2012; Offe et al., 2006; Rohe et al., 2008). It does so by forming a complex with the precursor by its central CR domains 5–8 thereby retaining APP in neuronal Golgi compartments and decreasing its processing into A β (Andersen et al., 2006; Mehmedbasic et al., 2015; Schmidt et al., 2007).

SorLA is encoded by the Alzheimer's disease (AD) risk gene *SORL1*, and several gene variants predisposing to both sporadic and familial forms have been identified (Lambert et al., 2013; Pottier et al., 2012; Rogaeva et al., 2007; Thonberg et al., 2017). In particular, rare truncating variants of *SORL1* contribute a strong disease penetrance in early-onset AD (EOAD) patients (Verheijen et al., 2016), which recently made researchers suggest that *SORL1* variants should be considered next to variants in *APOE*, *PSEN1/2*, and *APP* for personalized risk assessments in clinical practice (Holstege et al., 2017).

The mechanism of how intronic single-nucleotide polymorphisms (SNPs) influence *SORL1* gene function in the sporadic version of AD is still unknown, but as sorLA protein levels are decreased in brains of AD patients (Sager et al., 2007; Scherzer et al., 2004; Zhao et al., 2007), it is possible that disease-associated SNPs are responsible for the downregulation of *SORL1* gene expression. These effects could be tissue specific and only decrease receptor expression in certain (vulnerable) neuronal populations, or they could act globally to regulate transcriptional gene regulation or RNA processing. Another proposed mechanism suggests that certain SNPs may promote alternative splicing leading to inactive receptor mutants, as shown for a *SORL1* variant lacking exon 2 (sorLA- Ex2), which encodes part of the VPS10p domain (Grear et al., 2009; McCarthy et al., 2012). Recently, whole-exome sequencing on individuals from a family with history of EOAD reported a variant causing exon 22 skipping to segregate with disease in 3 affected and absent in 1 unaffected family

member, providing further evidence that *SORL1* splicing may indeed underlie certain AD cases (Thonberg et al., 2017).

The gene for human *SORL1* shares a rather complicated structure with the other VPS10p genes as they all contain many short exons separated by often very large introns (Hampe et al., 2001). For *SORL1*, the known gene structure is described as 48 exons encoding a protein with 2214 amino acids of which 28 residues constitute the signal sequence that is released upon entry to the secretory pathway (Hampe et al., 2001; Jacobsen et al., 1996) (cf. Fig. 1B).

As suggested for the murine gene (Hirayama et al., 2000), transcription of *SORL1* is believed to take place from a proximal promoter just upstream exon 1 (Guha Thakurta and Andersen, 2015). However, many proteins involved in the development of the central nervous system are transcribed from several different promoters thus enabling multiple regulatory mechanisms for temporal and spatial expression (Deaton and Bird, 2011; Kimura et al., 2006; Ma et al., 2009). Besides the use of different promoters, the genes may also generate several different variants by alternative splicing, adding further options for a complex and highly regulated expression pattern. This is exemplified by the brain-derived neurotrophic factor gene that has as many as 9 functional promoters and produces at least 17 different transcripts (Liu et al., 2006; Pruunsild et al., 2007; Timmusk et al., 1993).

We hypothesize that transcription and splicing of *SORL1* may be more complex than previously anticipated. This assumption is supported by several observations. When comparing published results, Northern blot analyses have shown multiple mRNA bands for both the human and murine genes (Hermans-Borgmeyer et al., 1998; Jacobsen et al., 1996; Kanaki et al., 1999). Moreover, Western blot analysis for sorLA protein has revealed several bands migrating faster than the 250-kDa full-length receptor, suggesting that some of these transcripts may be translated into shorter sorLA variants (Ciarlo et al., 2013; Hampe et al., 2000) (Andersen unpublished).

We here report the existence of a *SORL1* transcript with a novel exon (which we have named exon 30B) that includes a long 5' UTR. Translation initiates within exon 30B and leads to the expression of a truncated sorLA protein, which we refer to as sorLA-30B.

2. Materials and methods

2.1. Generation of the sorLA-30B expression vector

To obtain the sorLA-30B coding fragment, forward primer 5'-gtggcgccgcgatgattctagattagccgctgctcagagctgtgccagtcactggcctccctcagcagattcacatattcttccttc-3' and reverse primer 5'-gactcgagtcaggctatcaccatgggacgtcatctgaaaatccagttatcatag-3' were used in PCR with full-length *SORL1* as a template. The pcDNA3.1/ZEO(+) expression vector and the amplified sorLA-30B fragment were digested with *NotI* and *XhoI* restriction enzymes, ligated with T4 DNA ligase (M0202, New England Biolabs) and transformed in heat-shock competent DB5α *E. coli* (Max efficiency, Invitrogen). Transformed clones were validated by restriction digestion and DNA sequencing.

2.2. Cultivation of cells

ARPE-19 and SH-SY5Y cells were grown in Dulbecco's Modified Eagle's Medium Ham's F-12 (DMEM-F12, Lonza), and NT2, HeLa, HBT, and HEK293 cells in Dulbecco's Modified Eagle's Medium (DMEM, Lonza), at 37°C and 5% CO₂. Media were supplemented with fetal bovine serum (FBS) and penicillin/streptomycin to a final concentration of 10% and 1%, respectively.

HEK293 cells stably expressing sorLA-30B was obtained by selecting a single colony after selection by 300 µg/mL zeocin.

2.3. Reverse transcriptase–polymerase chain reaction (RT-PCR)

Human RNA from the total brain, cerebellum, temporal lobe, entorhinal cortex, frontal cortex, hippocampus, testis, and fetal kidney were from Ambion. Total RNA from cell lines was extracted using the RNeasy Mini Kit (Qiagen cat. no. 74,104) according to manufacturer's instructions. Total RNA from human cerebellum samples was extracted using TRIzol Reagent (Ambion Life Technology) followed by DNase RNA cleanup step using RNeasy (Qiagen). The quantity and quality of RNA samples were determined by the Agilent 2100 Bioanalyzer using Agilent RNA 6000 Nano Chip.

cDNA synthesis was performed with 1–2 µg of RNA using the High Capacity RNA-to-cDNA kit (Applied Biosystems cat. no. 4387406) according to the manufacturer's instructions.

2.4. Polymerase chain reaction

Exon 30B and exon 2–3 RT-PCR: Reverse transcription of total RNA was performed with OneStep RT-PCR (Qiagen Cat. No. 210212) using 4 µL One-Step Enzyme Mix, 4 µL OneStep RT-PCR buffer 5×, 4 µL Q-Solution 5×, 10 mM of each dNTP, 6 pmol of each primer, 100 ng RNA, and final volume 20 µL. PCR conditions were 30 minutes 50°C, 15 minutes 95°C (30 seconds 94°C, 30 seconds 58°C, and 1 minute 72°C) × (35 cycles for 30B) and (30 cycles for exon 2–3), and 10 minutes 72°C, held at 4°C.

Exon 30B RT-PCR on the temporal lobe using 2.5 U AmpliTaq DNA Polymerase with PCR Buffer I (Life Technologies Cat. No. N8080152), 25 mM of each dNTP, 2.5 pmol of each primer, 1 µL cDNA, final volume 50 µL. PCR conditions were 2 minutes 95°C (20 seconds 95°C, 20 seconds 56°C, and 30 seconds 72°C) × 30 cycles, 3 minutes 72°C, held at 4°C.

Exon 30B RT-PCR on human cerebellum samples using 0.5 U Apex Taq (Genesee Scientific ID No. 5101600) with 2.5 µL 10X buffer (Qiagen), 5 µL Q-Solution (Qiagen), 4 mM dNTPs, 0.8 µM of each primer, and 1 µL cDNA to a final volume of 25 µL. PCR conditions were 95°C 2 minutes, (95°C 30 seconds, 60°C 30 seconds, and 72°C 45 seconds) × 7 cycles, (95°C 30 seconds, 60°C–50°C (–0.5°C each cycle) 30 seconds, 72°C 45 seconds) × 20 cycles, (95°C 30 seconds, 50°C 30 seconds, 72°C 45 seconds) × 6 cycles, 72°C 3 minutes, held at 4°C.

SORL1 exon 2–3 PCR on human cerebellum samples: 0.5 U Apex Taq (Genesee Scientific ID No. 5101600) with 2.5 µL 10X buffer (Qiagen), 5 µL Q-Solution (Qiagen), 4 mM

dNTPs, 0.8 μ M of each primer, and 1 μ L cDNA to a final volume of 25 μ L. PCR conditions were 95°C 2 minutes, (95°C 30 seconds, 60°C 30 seconds, 72°C 45 seconds) $\times$ 7 cycles, (95°C 30 seconds, 60°C–50°C (–0.5°C each cycle) 30 seconds, 72°C 45 seconds) $\times$ 20 cycles, (95°C 30 seconds, 50°C 30 seconds, 72°C 45 seconds) $\times$ 6 cycles, 72°C 3 minutes, held at 4°C.

2.5. Human cohort

All tissue samples were collected from the Mayo Clinic Brain Bank. AD patients had a Braak score of 4.0 and a definite diagnosis according to the NINCDS-ADRDA criteria (McKhann et al., 1984). Control subjects had a Braak score of 3.0 or less and lacked any of the following pathologic diagnoses: AD, Parkinson's disease (PD), DLB, PSP, motor neuron disease (MND), CBD, Pick's disease (PiD), Huntington's disease (HD), FTLD, hippocampal sclerosis (HipScl), or dementia lacking distinctive histology (DLDH). All individuals were older than 60 years of age at time of death and were age matched across diagnosis groups ($\pm$ 1 year); 12 individuals (48%) from each diagnosis group were male. RNA was isolated from the cerebellum, and all samples had an RNA integrity number (RIN) of 6.6, with an equivalent mean RIN for both diagnosis groups (mean RIN AD patients = 8.26, mean RIN controls = 8.24). This study was approved by the appropriate institutional review board.

2.6. Primer design and sequences

The primers were designed using the web-based tool Primer3 (Koressaar and Remm, 2007; Untergasser et al., 2012), and the specificity was verified by search using the Basic Local Alignment Search Tool (BLAST, <http://blast.st-vi.ncbi.nlm.nih.gov/Blast.cgi>) and BLAT (<https://genome.ucsc.edu/cgi-bin/hgBlat?command=start>).

Primer sequences listed in 5'→3' orientation—exon 2 fwd: cactgggctggagagaaaag; exon 3 rev: ggcgataacagcttcactcc; exon 30B fwd: agtcactggcctccctcag; exon 32 rev: tcgaactcaaatcggtcaca; sorLA-GSP rev_a: ggggcaagtctgcattccacatgagt; and sorLA-GSP rev_b(In-Fusion): gattacgccaagcttgggcaagtctgcattccacatgagt.

2.7. 5'-RACE

The ClonTech SMARTer RACE cDNA Amplification Kit (Cat. No. 634923) was used according to the manufacturer's instructions, but with the following alterations—cDNA synthesis: 1 μ g RNA (Human Brain Total, Ambion) was used. 5'-RACE touchdown PCR: 1 μ L 50X Herculase polymerase and 10 μ L 5X Herculase buffer (Agilent technologies), 1 μ L 10 mM dNTP mix, 2.5 μ L cDNA, 5 μ L 10X UPM, 29.5 μ L ddH₂O, and 1 μ L gene-specific primer (10 μ M) per reaction. Gene-specific primer sorLA-GSP rev was used. PCR conditions were 95°C 1 minute, (95°C 20 seconds, 72°C 2 minutes) $\times$ 5 cycles, followed (95°C 20 seconds, 70°C 20 seconds, 68°C 2 minutes) $\times$ 5 cycles, next (95°C 20 seconds, 62°C 20 seconds, 68°C 2 minutes) $\times$ 27 cycles, 68°C 4 minutes, held at 4°C.

2.8. In-Fusion cloning

For first-strand cDNA synthesis, 1 μ g/ μ L RNA (Human Brain Cerebellum), 4 μ L 5X first-strand buffer, 0.5 μ L DTT (100 mM), 1 μ L dNTPs (20 mM), 1 μ L 5' CDS primer, 9 μ L

RNAse-free H₂O, 0.5 µL RNAse inhibitor (40 U/µL), and 20 µL SMARTScribe Reverse Transcriptase (100U) were used. 5'-RACE touchdown PCR: 1 µL SeqAmp DNA polymerase, 25 µL 2X SeqAmp Buffer, 15.5 µL ddH₂O, 2.5 µL cDNA, 5 µL 10X UPM and 1 µL gene-specific primer. PCR conditions were (94°C 30 seconds, 72°C 3 minutes) × 5 cycles, followed (94°C 30 seconds, 70°C 30 seconds, 72°C 3 minutes) × 5 cycles, and (94°C 30 seconds, 68°C 30 seconds, 72°C 3 minutes) × 25 cycles, held at 4°C.

For cloning step, 1 µL linearized pRACE vector (ClonTech), 7 µL purified PCR product, and 2 µL In-Fusion HD Master Mix were used. This mixture was used for the transformation of Stellar Competent Cells, plated on LB/Amp (1:1000) plates and incubated overnight at 37°C. Eight independent clones have been chosen, and then the plasmids were isolated using QIAprep Spin Miniprep Kit (250) according to manufacturer's instructions.

2.9. Agarose gel electrophoresis and purification of PCR products

PCR products were added loading buffer (50 % glycerol, 1× TAE-buffer), loaded on an agarose gel (ethidium bromide; 5'-RACE 1.2% and RT-PCR 2% Ultrapure agarose; Invitrogen, 1× TAE-buffer) and ran in 1× TAE-buffer (200 mA, 200V).

DNA fragments were cut out of the gel, and the DNA was purified using the QIAquick Gel Extraction Kit (Qiagen cat. no. 28,706) according to manufacturer's instructions. The DNA products were sequenced using Sanger sequencing at Eurofins Genomics. The obtained sequences were compared with already existing sequences in databases through searching using BLAST and BLAT.

2.10. Immunocytochemistry

HEK293 cells stably expressing sorLA-wt or sorLA-30B, grown on poly-lysine D pretreated coverslips, were fixed with 4% paraformaldehyde for 10 minutes and rinsed with phosphate buffered saline (PBS) 3 times. After fixation, cells were permeabilized by using 0.1% Triton X-100 in PBS pH 7.4 for 2 × 15 minutes and blocked with 10 % FBS in PBS for 30 minutes at room temperature. Blocked cells were incubated overnight with primary antibody rabbit anti-tail sorLA (polyclonal, 1:300–1:30, in house antibody 641731) in blocking solution at 4°C. For subcellular localization studies, the following antibodies were used: mouse anti-EEA1 for early endosomes (monoclonal, 1:100; BD Transduction Laboratories 610456, San Diego, CA), mouse anti-GM130 for Golgi (monoclonal, 1:100; BD Transduction Laboratories 610822), and mouse anti-TfR for the recycling endosomal compartments (monoclonal, 1:500; Thermo Fisher 13–6800, Fremont, CA). Cells were rinsed with 0.1% Triton-100 PBS for 3 times and incubated with secondary antibodies Alexa Fluor 488 goat anti-rabbit and Alexa Fluor 568 goat anti-mouse in blocking solution for 1 hour at room temperature. After secondary antibody treatment, cells were washed with 0.1% Triton-100 PBS. Hoechst (1:50,000, Thermo Fisher, cat. Number 62249) was used for nuclei staining. Dako fluorescent mounting medium was used for mounting, and analysis was performed by fluorescence microscopy (Zeiss LSM 510 META confocal microscope).

2.11. Western blot

Cells were collected, lysed in RIPA buffer (Pierce 89900), and centrifuged (10,000 rpm, 4°C) to collect the supernatant. Protein lysates were separated using 4%–16% Tris-Glycine gel followed by transfer on nitrocellulose membrane (Amersham) for 2.5 hours in blotting buffer 1X (250 mM Tris-base, 1.92 M glycine). The membrane was blocked for unspecific binding in blocking buffer (20% TST buffer [0.25 M Tris-base, 2.5 M NaCl, 0.5% Tween-20, pH 9], 2% skimmed milk powder, 2% Tween-20) for 1 hour at room temperature and incubated overnight with primary antibody rabbit anti-tail sorLA (1:1000) at 4°C, followed by incubation with polyclonal swine anti-rabbit HRP conjugated secondary antibody. SuperSignal West Femto Maximum Sensitivity Substrate was used to visualize blots.

3 Results

3.1. Multiple SORL1 transcripts

We initiated our quest for new sorLA variants by a web-based search for truncated transcripts in the UCSC Genome Browser database. We found a total of 5 distinct putative protein encoding transcripts, including the transcript from which the full-length sorLA with 2214 amino acids is expressed. The 4 additional transcripts AK295878, AK295090, AK302589, and AK294404 could all be further validated by searches in the ENSEMBL and NCBI databases and suggested to encode proteins with 1158, 1060, 1124, and 829 amino acids, respectively (Table 1). Interestingly, all the described proteins correspond to truncation of the amino-terminal end deleting the VPS10p and the YWTD-repeated β -propeller domains. Calculation of the molecular weight for the predicted open reading frames (ORFs) of the 4 novel transcripts led to prediction of polypeptides weighing 129.7 kDa, 118.5 kDa, 125.8 kDa, and 92.3 kDa, and corresponding to UniProt entries E9PPB3, E9PS32, E9PP43, and E9PKB0, respectively (<http://proteomicsatlas.org/ENSG00000137642-SORL1/gene>).

3.2. Analysis of a novel SORL1 transcript including exon 30B

We decided to further investigate the shortest of these transcripts (AK294404) as our initial studies using transcript-specific primers for 5'-RACE with brain RNA as a template were not able to verify the presence of the other transcripts. According to the available sequence found in the UCSC database, the transcription of AK294404 would start in intron 30 (between exon 30 and exon 31 of the *SORL1* gene) with a 421-bp 5' UTR and include a unique sequence not present in the mature transcript for the full-length receptor. To confirm the existence of this transcript and to determine the exact sequence including the part of the novel exon 30B predicted by the database, we did a 35-cycle reverse transcriptase polymerase chain reaction (RT-PCR) using a primer pair located in the putative exon 30B (forward primer) and exon 32 (reverse primer) and cDNA prepared from RNA isolated from the total human brain as a template (Fig. 2A). This pair of primers will not detect the full-length transcript (BC137171) but is specific for the novel transcript (AK294404) as the forward primer is annealing to a region corresponding to intron 30 of the full-length gene. In a parallel experiment, we also tested cDNA prepared from RNA of the human temporal lobe by a 30-cycle RT-PCR (Fig. 2A). For both experiments, we found that the exon 30B/32

primer pair produced a product migrating between 200 and 300 bp in accordance with the expected fragment of 239 bp from AK294404 where the introns 30B and 31 have been removed. In comparison, a fragment corresponding to amplification of genomic DNA using the exon 30B/32 primer pair would have produced a band containing 4809 bp, clearly showing that the obtained band is not related to genomic DNA. However, we did observe a weaker product approximately 100 bp longer than the 239-bp products for both reactions (marked by an asterisk, Fig. 2A).

Sanger sequencing confirmed that the fragment (arrow, Fig. 2A) contains the expected 239 bp corresponding to the 3' end of exon 30B, the full exon 31, and the 5' end of exon 32 (Fig. 2B). The band above 342 bp (asterisk, Fig. 2A) corresponds to a product where the intron between exon 30B and exon 31 with 103 base pairs is not removed. In silico studies showed that the translation of the transcript containing the 103-bp intron would lead to a frameshift and a 97 amino acid peptide with no relation to sorLA before termination by a stop codon. From these findings, we conclude that a part of intron 30 from the full-length *SORL1* gene also can function as a new exon that may encode the N-terminal end of a truncated sorLA receptor.

To determine in which brain regions exon 30B containing transcripts are expressed, we applied an alternative protocol using OneStep RT-PCR with RNA as a template. Using this method, we found that the transcript is most highly expressed in the hippocampus and temporal lobe but also present in the cerebellum, entorhinal cortex, and frontal cortex. We did not detect the 239-bp product using RNA from neither the testis nor fetal kidney (Fig. 2C). We also included RNA obtained from several human cell lines, including ARPE-19, NT2, HeLa, HEK293, HBT-186, and SH-SY5Y, but only NT2 showed the presence of spliced exon 30B. To document the presence of the full-length gene product (transcript BC137171), we included a control primer pair located in exon 2 (forward primer) and exon 3 (reverse primer), which as expected generated a 174-bp-long amplification product of the 250-kDa receptor transcript in all the reactions (Fig. 2D).

3.3. Identification of transcription start sites

Besides the suggested start site by the database entries leading to a 421-bp 5' UTR, we had no experimental evidence for the exact transcriptional start site (TSS) of the sorLA-30B transcript, that could in principle be longer than anticipated in the 5' end. We therefore performed 5'-RACE using total human brain RNA as starting material. First, RNA was converted to SMARTer cDNA, and then 5'-RACE was made using a *SORL1*-specific reverse primer (sorLA-GSP rev_a) located in the unique sequence in the transcript upstream of the coding part of exon 30B, and a forward 5' coding strand (CDS fw) primer specifically designed to amplify SMARTer cDNA from the very start of transcripts. The resulting PCR fragments from the 5'-RACE were analyzed by agarose gel electrophoresis. A single fragment of approximately 1500 bp was visible from the reaction using total human brain RNA as a template (Fig. 3A). These data therefore confirm that AK294404 is present as a stable transcript in the human brain.

The DNA fragment obtained by PCR using brain RNA was then excised from the gel, extracted, and sequenced by Sanger sequencing. To identify the sequence, we searched the

human genome hg19 using BLAT. The sequence obtained by the 5'-RACE (Fig. 3A, dark gray) did span from the reverse primer positioned upstream from exon 30B into the rest of intron 30, exon 30, intron 29, exon 29, and partly into intron 28 of the *SORL1* gene (Fig. 4). Interestingly, the experimentally identified TSS maps with high accuracy to an annotated TSS listed in the database of transcriptional start sites (www.dbtss.gnb.jp) (DBTSS; marked with ↓ in Fig. 4).

However, as we were surprised by the very long (1.5 kb) 5' UTR as opposed to the expected 421 bp, we applied a second round of 5'-RACE using human RNA from the cerebellum for mapping of TSS for exon 30B-specific transcripts. For this experiment, we used the SMARTer RACE 5' kit allowing for In-Fusion cloning of the obtained products by the 5'-RACE reaction into the pRACE vector according to the supplier's manual. We sequenced 8 clones (Fig. 3B, a–h) and again all containing an approximately 1.5 kb insert, and their individual start sites for exon 30B transcripts (marked with ↓ in Fig. 4) all mapped very close to the TSS already obtained by the 5'-RACE protocol and RNA from the total brain. Accordingly, we conclude that the 5' UTR for *SORL1-30B* is much longer than the anticipated 5' UTR based on the available sequence of transcript AK294404 in the database.

3.4. SorLA-30B open reading frame characterization

Inspection of the entire sequence upstream and including exon 30B revealed several potential tri-nucleotide ATG translational initiation sites, but only usage of a single ATG results in an 20 amino acid ORF without stop codons and in-frame with the sequence of exon 31 that encodes the ninth CR domain of full-length sorLA (Fig. 4).

This predicted translational initiation site (TACATGA) corresponds to an “adequate” Kozak consensus sequence with a conserved purine at the +4 position of the consensus, whereas the –3 position is not a purine (Kozak, 1986, 1987). For comparison, the Kozak consensus around the initiation site of exon 1 corresponds to a strong consensus with purines at both these important positions (Fig. 5A).

These 20 residues encoded by exon 30B could potentially function as a signal sequence enabling the protein to enter the secretory pathway and insert into the plasma membrane similar to full-length sorLA. To determine if the 20-residue peptide corresponds to a signal sequence, we analyzed the peptide using the SignalP 4.1 server software (www.cbs.dtu.dk), but the program was unable to identify any signal sequence. However, a signal sequence typically is 16-to 30-residue long and contains 1 or more positively charged amino acids followed by a continuous stretch of 6–12 residues that carry hydrophobic side chains. Manual inspection of the 20 amino acids encoded by the ORF of exon 30B, revealed that residues 4–8 comprising Arg-Leu-Ala-Ala-Ala fulfill these criteria, suggesting that the first 20 amino acid sequence in sorLA-30B may constitute a signal sequence. In support of this notion, the DegraBase (Crawford et al., 2013) further revealed that serine-21 has been identified with a free NH₂ group. The DegraBase describes N-termini from 3 general categories of proteolysis: translational N-terminal methionine processing, sites close to the translational N-terminus that likely represent removal of transit or signal peptides, and other endoproteolytic events (Crawford et al., 2013). Accordingly, this finding is in agreement with our hypothesis that residues 1–20 is a functional signal peptide that becomes cleaved

off between Asp-20 and Ser-21, although we cannot formally exclude, this might alternatively be accounted for by proteolytic cleavage of full-length sorLA. We named the resulting truncated receptor as sorLA-30B (Fig. 5C).

From the aforementioned analysis, we concluded that the AK294404 transcript could lead to translation of a 98-kDa protein that uses the ATG start codon in exon 30B.

3.5. Characterization of intron 30B

The sequencing of the 239-bp fragment obtained by RT-PCR revealed that also a fragment of 103 bp corresponding to a novel intron 30B has been removed to produce the mature and spliced messenger. This gene constellation leads to 1 novel exon-intron boundary corresponding to the acceptor site between exon 30B and intron 30B.

To confirm the presence of splicing-directing sequences in exon 30B, we analyzed the complete sequence shown in Fig. 5 using the web-based splice finder software Genie (http://www.fruitfly.org/seq_tools/splice.html) (Reese et al., 1997). This software searches DNA sequences for motifs known to define splice sites and scores the likelihood that splicing may take place. The predicted splice site between exon 30B and intron 30B received a score of 0.80 (Table 2). Maximum score by Genie is 1.0 and cutoff is 0.4. Notably, this value is comparable to those obtained for the acceptor and donor sites for the flanking exons 29, 30, and 31 in full-length sorLA. This analysis supports that splicing from exon 30B to exon 31 is likely to occur, as already documented by the sequencing of the PCR product.

3.6. Expression of exon 30B transcripts in patients with Alzheimer's disease

The expression of the full-length *SORL1* transcript is down-regulated in the hippocampus and temporal lobe of several patients with AD (Gear et al., 2009; Zhao et al., 2007) (and reviewed in (Andersen et al., 2016)). The level of sorLA protein is also reduced in hippocampal and cortical neurons from AD patients, whereas the *SORL1* level in the cerebellum seems unaffected by the disease (Offe et al., 2006; Scherzer et al., 2004). However, we have recently identified a new splice variant that showed a very robust (>2-fold) decrease in the cerebellum of AD patients compared with nondemented controls. We therefore decided to test the level also of exon 30B containing transcripts in the cerebellum of a cohort of 25 AD patients and 25 nondemented controls by RT-PCR. We applied the primer pair for exon 30B forward and exon 32 reverse as well as the primer pair for exon 2 and exon 3. We observed both the short (spliced) and the longer (containing intron 30B) PCR products for all the samples using the 30B and 32 primer pair (Fig. 6A) as well as a single band for the exon 2 and 3 primer pair (not shown). However, there was no difference in the expression level of neither the short (lower) band nor the long (upper) band between AD and non-AD individuals. Quantification of band intensities showed that the mean level of the lower band was almost identical between the 2 cohorts (100.7 % in AD relative to 100% in controls) (Fig. 6B). We also found no significant difference between AD and controls when comparing the intensity of the upper bands (97.8% in AD relative to 100% in controls). The band corresponding to amplification by primer pair 2 and 3 also showed no difference (Fig. 6B). Furthermore, there was also no difference when testing the ratio between the 2 bands for the 30B and 32 primer pair in the 2 groups as well as the ratio

between either bands and the exon 2–3 band were not changed (Fig. 6C). Accordingly, we conclude that in the cerebellum of this small cohort, there was no effect of the disease on *SORL1–30B* expression similar to the situation for full-length *SORL1*. Future studies should investigate whether there is any change in the expression of this variant in other brain regions.

3.7. Expression of sorLA-30B in HEK293 and NT2 cells

To study expression and cellular localization of the novel receptor variant, we generated an expression construct encoding the entire sorLA-30B protein, which we subsequently used to generate HEK293 cells that stably express sorLA-30B. Expression of sorLA-30B was demonstrated by Western blot analysis, where transfected cells showed a clear band above the 98-kDa marker recognized by an antibody against the cytoplasmic tail of sorLA that was not present in the nontransfected cells (Fig. 7A).

As we observed the spliced AK294404 transcript to be present in NT2 cells, we decided to see if we could identify the endogenous sorLA-30B translation product in this cell line. Expression of endogenous sorLA-30B in NT2 cells was similarly documented by Western blot analysis, where a band showing similar migration as the band representing sorLA-30B in the stable HEK293 cells were identified (Fig. 7A).

The finding of endogenous expression of sorLA-30B in NT2 cells is in agreement with previous reports that described a receptor variant with this size in human NT2 and SK-N-BE2 neuroblastoma cells (Ciarlo et al., 2013; Hampe et al., 2000).

Next, we compared the cellular localization of sorLA-30B and the full-length receptor by immunofluorescence cytochemistry using HEK293 cells stably expressing the 2 receptor variants and localized sorLA-wt and sorLA-30B by confocal microscopy, together with markers for Golgi (GM130), early endosomes (EEA1), or recycling endosomal compartments (TfR) (Fig. 7B and C). The HEK293 cell line stably expressing sorLA-wt has been described before (Gustafsen et al., 2013).

In agreement with previous cellular localization studies for sorLA-wt in HEK293 cells (Herskowitz et al., 2012), we found strong colocalization between sorLA and TfR, but only weak colocalization for sorLA and EEA1, and virtually no colocalization for sorLA and GM130. Most importantly, the degree of colocalization with the different tested markers was indistinguishable for sorLA-wt and sorLA-30B receptor proteins (Herskowitz et al., 2012), strongly suggesting that the shorter receptor variant is sorted along the same trafficking pathways as the full-length receptor.

4. Discussion

We have characterized a new *SORL1* transcript with a long 5' UTR and TSS located in intron 28 of the gene encoding sorLA. The transcript was apparent for all the brain areas that we tested as well as NT2 cells, suggesting a brain-specific expression profile of functional sorLA-30B. The transcript contains a novel exon 30B that is spliced in-frame to exon 31, and the ORF suggests the translation of a protein identical to full-length sorLA from residue

1405 (numbering according to full-length sorLA). We did not experimentally determine the 3' end of the transcript, but all the database entries describing this transcript agree on a common gene structure from exon 31 to exon 48 including its stop codon, which fits the expected size of 98 kDa (i.e., the expressed sorLA-30B protein).

As the expression of a ~ 100-kDa sorLA variant has also been reported for mice (Hampe et al., 2000), we looked for conservation of exon 30B in the mouse genome. However, an alignment of intron 30 from the human and mouse genomes revealed little conservation for the region of exon 30B, suggesting that this exon is not part of murine *SORL1*. We also looked for exon 30B by searching among expressed sequence tags and annotated transcripts for *Bos taurus*, *Sus scrofa*, and *Pan troglodytes*, but we identified no hits for neither the cow, nor pig, nor chimpanzee, suggesting that the described exon 30B is human specific. Previous studies for the VPS10p-domain receptor sortilin, similarly, showed species-specific transcripts between mouse and human genes (Prudencio et al., 2012). However, this would mean that mouse *SORL1* has another way to generate a truncated ~ 100-kDa sorLA receptor, but we did not address this further in the present study.

Induction of a 98-kDa receptor variant was recently described in the presence of a long noncoding RNA molecule named 51A (genomic location indicated in Fig. 1B), and the suggested mechanism was regulation of alternative splicing of the *SORL1* gene (Ciarlo et al., 2013). Our data suggest that 51A may function by repression of transcription from exon 1, and activation of transcription starting with exon 30B to increase the level of the 98-kDa sorLA variant. This mechanism is in line with results from other studies reporting that to increase the size of the transcriptome, the most commonly used mechanisms are alternative splicing and alternative promoter usage. Especially in the brain, the latter mechanism is used. Sixty percent of genes in the brain transcriptome use alternative transcriptional start sites (Pardo et al., 2013). In fact, the use of alternative transcriptional start sites is more common than alternative splicing during brain development (Pal et al., 2011).

The protein encoded by this novel gene constellation contains 829 amino acids lacking all known ligand-binding domains of full-length sorLA (Fig. 5C). The function of sorLA-30B is thus still speculative, but given that all structural domains in sorLA have their own specific ligand(s)—the VPS10p domain binds GDNF (Glerup et al., 2013), the YWTD-repeated β -propeller binds MESD (Andersen et al., 2006), and the cluster of CR domains binds APP (Andersen et al., 2006)—it is likely that the FnIII repeats encoded by sorLA-30B will also bind unique ligand(s). Future studies should aim to find those proteins that are likely to define the function of sorLA-30B.

Of notice, the sorLA-30B ectodomain contains only 3 CR domains (CR9–11) in common with full-length sorLA, and sorLA-30B therefore does not contain the binding site for APP residing in CR5–8 (Mehmedbasic et al., 2015). Thus, it is highly unlikely that sorLA-30B is involved in sorting of APP, although this is considered a prime function for full-length sorLA (Willnow and Andersen, 2013).

However, the receptor still contains the entire cluster of the 6 FnIII domains, which are commonly found in proteins engaged in protein-protein interactions at the plasma

membrane, e.g. neural cell adhesion molecule, thus suggesting a specific function of sorLA-30B in, for example, cell-cell adhesion (Maness and Schachner, 2007). FnIII domains are known to engage in ligand interactions with a plethora of different proteins. Accordingly, the functional consequence of sorLA-30B being expressed but lacking all known ligand-binding domains is that this novel receptor variant is very likely to have a unique function in hitherto not yet recognized roles identified for the full-length receptor.

Intracellular trafficking pathways of transmembrane receptors are most often dictated by short-sequence motifs binding to sorting regulating adaptor proteins at the cytoplasmic site. This has also been reported for sorLA that contains an acidic motif required for AP1- and AP2-dependent trafficking, a FANSHY motif for interaction with the retromer, and a motif for sorting by the GGA adaptor proteins (Fjorback et al., 2012; Herskowitz et al., 2012; Jacobsen et al., 2002; Klinger et al., 2016; Nielsen et al., 2007; Schmidt et al., 2007). As sorLA-30B and the full-length sorLA receptors share the same cytoplasmic tail sequence, it is therefore very likely that they also share an identical sorting machinery, in line with our indistinguishable subcellular localization in the analyzed HEK293 cells. Accordingly, if the 2 variants are expressed in the same cell, it is reasonable that they will locate to the same cellular compartments. The functional consequence of this is currently unknown, but as all other sorLA-related proteins, that is, sortilin, SorCS1, SorCS2, and SorCS3 all have been shown to dimerize (Janulienė et al., 2017; Leloup et al., 2017), it is tempting to speculate that sorLA-30B may influence the putative dimer formation of full-length sorLA if present in the same cell. On the other hand, transcriptional regulation of *SORL1-30B* is determined by another promoter than *SORL1* variant containing exon 1, and the 2 variants may thus be expressed in different cellular populations and/or at distinct time points. Future studies should aim to clarify if the 2 variants display distinct or overlapping expression profiles.

Although never previously reported for sorLA, it is not surprising that sorLA-30B expresses only some of the CR domain. This seems to be a common feature of many CR domain-containing receptors and is possible because each CR domain is encoded by single individual exons thus promoting domain shuffling (Hampe et al., 2001). Accordingly, these receptors are able to combine their CR domains differently by alternative splicing and transcription, whereby a single gene generates receptor domains with different ligand binding properties. A well-described example is the apolipoprotein E receptor 2 where splicing of specific exons can fine-tune ligand binding in different tissues (Brandes et al., 2001; Hibi et al., 2009; Koch et al., 2002; Korschneck et al., 2001) and alter neuronal receptor functions (Beffert et al., 2005; Wasser et al., 2014). This seems similar to the situation for sorLA, where the receptor with all 11 CR domains binds APP with high affinity, but sorLA-30B is unlikely to do so. Alternative splicing has also been reported for the related sorCS1 receptor. In this case, alternative splicing that leads to receptor isoforms with different cytoplasmic domains are mediating diverse neuronal localization due to the presence of distinct trafficking motifs (Hermey et al., 2003; Nielsen et al., 2008).

Variants at the *SORL1* locus are associated with AD, but the trait is highly complex, and it is far from clear how individual associated SNPs are functionally linked to disease onset. Haploblock analysis found 2 independent blocks in the 5' and the 3' end of *SORL1* (13), and it has been particularly difficult to understand how the regions in the 3' end of the gene

influences *SORL1* overall gene expression as this region is located very far from the TSS of full-length *SORL1*. However, they are in close proximity for the TSS of *SORL1-30B*. The experimentally identified TSS leads to a surprisingly long 1.5 kb 5'-UTR for the transcript AK294404. This 5'-UTR fragment contains 3 SNPs that are significantly associated with AD, that is, rs2276412 (Reitz et al., 2013), rs12272618 (Reitz et al., 2013), and Asn1358Ser (Pottier et al., 2012) (Fig. 5). For rs2276412 and rs12272618, no functional link to the disease has been established as these SNPs are occurring as a synonymous substitution of a codon in exon 30 and within intron 29, respectively. The variant corresponding to Asn1358Ser leads to substitution of an asparagine to a serine residue within exon 29 of the full-length sorLA, thus potentially leading to translation of a dysfunctional receptor.

Future studies are needed to establish whether any of the described SNPs are related to expression of sorLA-30B.

5. Conclusion

In summary, we have characterized a *SORL1* transcript that starts with an exon 30B and contains a 1.5 kb 5'-UTR. This transcript encodes a truncated receptor and may explain the origin of a 98-kDa receptor variant previously observed in human brain samples. The level of *SORL1-30B* transcripts is unchanged in the cerebellum of AD patients.

Acknowledgments

This work was supported by grants from The Lundbeck Foundation [R248-2017-431, DANDRITE-R248-2016-2518], The Danish National Research Foundation [DNRF133], The Novo Nordisk Foundation, The Augustinus Foundation, The Hartmann Foundation, The Henrik Henriksen Foundation, The Oda and Hans Svenningsen Foundation, and The Danish Council for Independent Research | Medical Sciences DFF [4004-00371 and 7016-00261]. This work was also supported by the National Institute on Aging [U01 AG046139 and RF1 AG051504], National Institute of Neurological Disorders and Stroke [R01 NS080820] and Bright Focus Foundation [27J-01-03] awarded to N.E.-T. The authors are grateful for expert technical assistance by Sandra Bonnesen.

References

- Andersen OM, Reiche R, Schmidt V, Gotthardt M, Spoelgen R, Behlke J, von Arnim CA, Breiderhoff T, Jansen P, Wu X, Bales KR, Cappai R, Masters C, Gliemann J, Mufson EJ, Hyman BT, Paul SM, Nykjær A, Willnow TE, 2005 SorLA/LR11, a neuronal sorting receptor that regulates processing of the amyloid precursor protein. *Proc. Natl. Acad. Sci. U. S. A* 102, 13461–13466. [PubMed: 16174740]
- Andersen OM, Rudolph IM, Willnow TE, 2016 Risk factor SORL1: from genetic association to functional validation in Alzheimer's disease. *Acta Neuropathol* 132, 653–665. [PubMed: 27638701]
- Andersen OM, Schmidt V, Spoelgen R, Gliemann J, Behlke J, Galatis D, Mckinstry WJ, Parker MW, Masters CL, Hyman BT, Cappai R, Willnow TE, 2006 Molecular dissection of the interaction between APP and its neuronal trafficking receptor SorLA/LR11. *Biochemistry* 45, 2618–2628. [PubMed: 16489755]
- Beffert U, Weeber EJ, Durudas A, Qiu S, Masiulis I, Sweatt JD, Li WP, Adelmann G, Frotscher M, Hammer RE, Herz J, 2005 Modulation of synaptic plasticity and memory by reelin involves differential splicing of the lipoprotein receptor apoer2. *Neuron* 47, 567–579. [PubMed: 16102539]
- Brandes C, Kahr L, Stockinger W, Hiesberger T, Schneider WJ, Nimpf J, 2001 Alternative splicing in the ligand binding domain of mouse apoE receptor-2 produces receptor variants binding reelin but not alpha2-macroglobulin. *J. Biol. Chem* 276, 22160–22169. [PubMed: 11294845]
- Ciarlo E, Massone S, Penna I, Nizzari M, Gigoni A, Dieci G, Russo C, Florio T, Cancedda R, Pagano A, 2013 An intronic ncRNA-dependent regulation of SORL1 expression affecting Abeta formation

is upregulated in post-mortem Alzheimer's disease brain samples. *Dis. Model Mech* 6, 424–433. [PubMed: 22996644]

- Crawford ED, Seaman JE, Agard N, Hsu GW, Julien O, Mahrus S, Nguyen H, Shimbo K, Yoshihara HA, Zhuang M, Chalkley RJ, Wells JA, 2013 The DegraBase: a database of proteolysis in healthy and apoptotic human cells. *Mol. Cell. Proteomics* 12, 813–824. [PubMed: 23264352]
- Deaton AM, Bird A, 2011 CpG islands and the regulation of transcription. *Genes Dev* 25, 1010–1022. [PubMed: 21576262]
- Dodson SE, Andersen OM, Karmali V, Fritz JJ, Cheng D, Peng J, Levey AI, Willnow TE, Lah JJ, 2008 Loss of LR11/SORLA enhances early pathology in a mouse model of amyloidosis: evidence for a proximal role in Alzheimer's disease. *J. Neurosci* 28, 12877–12886. [PubMed: 19036982]
- Fjorback AW, Seaman M, Gustafsen C, Mehmedbasic A, Gokool S, Wu C, Militz D, Schmidt V, Madsen P, Nyengaard JR, Willnow TE, Christensen EI, Mobley WB, Nykjær A, Andersen OM, 2012 Retromer binds the FANSHY sorting motif in sorLA to regulate amyloid precursor protein sorting and processing. *J. Neurosci* 32, 1467–1480. [PubMed: 22279231]
- Glerup S, Lume M, Olsen D, Nyengaard JR, Vaegter CB, Gustafsen C, Christensen EI, Kjolby M, Hay-Schmidt A, Bender D, Madsen P, Saarma M, Nykjaer A, Petersen CM, 2013 SorLA controls neurotrophic activity by sorting of GDNF and its receptors GFRalpha1 and RET. *Cell Rep* 3, 186–199. [PubMed: 23333276]
- Grear KE, Ling IF, Simpson JF, Furman JL, Simmons CR, Peterson SL, Schmitt FA, Markesbery WR, Liu Q, Crook JE, Younkin SG, Bu G, Estus S, 2009 Expression of SORL1 and a novel SORL1 splice variant in normal and Alzheimer's disease brain. *Mol. Neurodegener* 4, 46. [PubMed: 19889229]
- Guha Thakurta I, Andersen OM, 2015 Associations of sorLA/SORL1 with Alzheimer's disease. *Receptor Clin. Invest* 2, e700.
- Gustafsen C, Glerup S, Pallesen LT, Olsen D, Andersen OM, Nykjær A, Madsen P, Petersen CM, 2013 Sortilin and SorLA display distinct roles in processing and trafficking of amyloid precursor protein. *J. Neurosci* 33, 64–71. [PubMed: 23283322]
- Hampe W, Rezgaoui M, Hermans-Borgmeyer I, Schaller HC, 2001 The genes for the human VPS10 domain-containing receptors are large and contain many small exons. *Hum. Genet* 108, 529–536. [PubMed: 11499680]
- Hampe W, Riedel IB, Lintzel J, Bader CO, Franke I, Schaller HC, 2000 Ectodomain shedding, translocation and synthesis of SorLA are stimulated by its ligand head activator. *J. Cell Sci* 113, 4475–4485. [PubMed: 11082041]
- Hermans-Borgmeyer I, Hampe W, Schinke B, Methner A, Nykjær A, Süsens U, Fenger U, Herbarth B, Schaller HC, 1998 Unique expression pattern of a novel mosaic receptor in the developing cerebral cortex. *Mech. Dev* 70, 65–76. [PubMed: 9510025]
- Hermey G, Keat SJ, Madsen P, Jacobsen C, Petersen CM, Gliemann J, 2003 Characterization of sorCS1, an alternatively spliced receptor with completely different cytoplasmic domains that mediate different trafficking in cells. *J. Biol. Chem* 278, 7390–7396. [PubMed: 12482870]
- Herskowitz JH, Offe K, Deshpande A, Kahn RA, Levey AI, Lah JJ, 2012 GGA1-mediated endocytic traffic of LR11/SorLA alters APP intracellular distribution and amyloid-beta production. *Mol. Biol. Cell* 23, 2645–265. [PubMed: 22621900]
- Hibi T, Mizutani M, Baba A, Hattori M, 2009 Splicing variations in the ligand-binding domain of ApoER2 results in functional differences in the binding properties to Reelin. *Neurosci. Res* 63, 251–258. [PubMed: 19167437]
- Hirayama S, Bujo H, Yamazaki H, Kanaki T, Takahashi K, Kobayashi J, Schneider WJ, Saito Y, 2000 Differential expression of LR11 during proliferation and differentiation of cultured neuroblastoma cells. *Biochem. Biophys. Res. Commun* 275, 365–373. [PubMed: 10964672]
- Holstege H, van der Lee SJ, Hulsman M, Wong TH, van Rooij JG, Weiss M, Louwersheimer E, Wolters FJ, Amin N, Uitterlinden AG, Hofman A, Ikram MA, van Swieten JC, Meijers-Heijboer H, van der Flier WM, Reinders MJ, van Duijn CM, Scheltens P, 2017 Characterization of pathogenic SORL1 genetic variants for association with Alzheimer's disease: a clinical interpretation strategy. *Eur. J. Hum. Genet* 25, 973–981. [PubMed: 28537274]

- Jacobsen L, Madsen P, Moestrup SK, Lund AH, Tommerup N, Nykjær A, Sottrup-Jensen L, Gliemann J, Petersen CM, 1996 Molecular characterization of a novel human hybrid-type receptor that binds the $\alpha 2$ -macroglobulin receptor-associated protein. *J. Biol. Chem* 271, 31379–31383. [PubMed: 8940146]
- Jacobsen L, Madsen P, Nielsen MS, Geraerts WPM, Gliemann J, Smit AB, Petersen CM, 2002 The sorLA cytoplasmic domain interacts with GGA1 and –2 and defines minimum requirements for GGA binding. *FEBS Lett* 511, 155–158. [PubMed: 11821067]
- Janulienė D, Manavalan A, Ovesen PL, Pedersen KM, Thirup S, Nykjaer A, Moeller A, 2017 Hidden twins: SorCS neuroreceptors form stable dimers. *J. Mol. Biol* 429, 2907–291. [PubMed: 28827148]
- Kanaki T, Bujo H, Hirayama S, Ishii I, Morisaki N, Schneider WJ, Saito Y, 1999 Expression of LR11, a mosaic LDL receptor family member, is markedly increased in atherosclerotic lesions. *Arterioscler. Thromb. Vasc. Biol* 19, 2687–2695. [PubMed: 10559012]
- Kimura K, Wakamatsu A, Suzuki Y, Ota T, Nishikawa T, Yamashita R, Yamamoto J, Sekine M, Tsuritani K, Wakaguri H, Ishii S, Sugiyama T, Saito K, Isono Y, Irie R, Kushida N, Yoneyama T, Otsuka R, Kanda K, Yokoi T, Kondo H, Wagatsuma M, Murakawa K, Ishida S, Ishibashi T, Takahashi-Fujii A, Tanase T, Nagai K, Kikuchi H, Nakai K, Isogai T, Sugano S, 2006 Diversification of transcriptional modulation: large-scale identification and characterization of putative alternative promoters of human genes. *Genome Res* 16, 55–65. [PubMed: 16344560]
- Klinger SC, Hojland A, Jain S, Kjolby M, Madsen P, Svendsen AD, Olivecrona G, Bonifacino JS, Nielsen MS, 2016 Polarized trafficking of the sorting receptor SorLA in neurons and MDCK cells. *FEBS J* 283, 2476–2493. [PubMed: 27192064]
- Koch S, Strasser V, Hauser C, Fasching D, Brandes C, Bajari TM, Schneider WJ, Nimpf J, 2002 A secreted soluble form of ApoE receptor 2 acts as a dominant-negative receptor and inhibits Reelin signaling. *EMBO J* 21, 5996–6004. [PubMed: 12426372]
- Koressaar T, Remm M, 2007 Enhancements and modifications of primer design program Primer3. *Bioinformatics* 23, 1289–1291. [PubMed: 17379693]
- Korschineck I, Ziegler S, Breuss J, Lang I, Lorenz M, Kaun C, Ambros PF, Binder BR, 2001 Identification of a novel exon in apolipoprotein E receptor 2 leading to alternatively spliced mRNAs found in cells of the vascular wall but not in neuronal tissue. *J. Biol. Chem* 276, 13192–13197. [PubMed: 11152697]
- Kozak M, 1986 Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* 44, 283–292. [PubMed: 3943125]
- Kozak M, 1987 An analysis of 5′-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res* 15, 8125–8148. [PubMed: 3313277]
- Lambert JC, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C, Destafano AL, Bis JC, Beecham GW, Grenier-Boley B, Russo G, Thornton-Wells TA, Jones N, Smith AV, Chouraki V, Thomas C, Ikram MA, Zelenika D, Vardarajan BN, Kamatani Y, Lin CF, Gerrish A, Schmidt H, Kunkle B, Dunstan ML, Ruiz A, Bihoreau MT, Choi SH, Reitz C, Pasquier F, Cruchaga C, Craig D, Amin N, Berr C, Lopez OL, de Jager PL, Deramecourt V, Johnston JA, Evans D, Lovestone S, Letenneur L, Moron FJ, Rubinsztein DC, Eiriksdottir G, Sleegers K, Goate AM, Fievet N, Huentelman MW, Gill M, Brown K, Kamboh MI, Keller L, Barberger-Gateau P, McGuinness B, Larson EB, Green R, Myers AJ, Dufouil C, Todd S, Wallon D, Love S, Rogaeva E, Gallacher J, St George-Hyslop P, Clarimon J, Lleó A, Bayer A, Tsuang DW, Yu L, Tsolaki M, Bossu P, Spalletta G, Proitsi P, Collinge J, Sorbi S, Sanchez-Garcia F, Fox NC, Hardy J, Deniz Naranjo MC, Bosco P, Clarke R, Brayne C, Galimberti D, Mancuso M, Matthews F, European Alzheimer's Disease, I., Genetic, Environmental Risk in Alzheimer's, D., Alzheimer's Disease Genetic, C., Cohorts for, H., Aging Research in Genomic, E., Moebus S, Mecocci P, del Zompo M, Maier W, Hampel H, Pilotto A, Bullido M, Panza F, Caffarra P, Nacmias B, Gilbert JR, Mayhaus M, Lannefelt L, Hakonarson H, Pichler S, Carrasquillo MM, Ingelsson M, Beekly D, Alvarez V, Zou F, Valladares O, Younkin SG, Coto E, Hamilton-Nelson KL, Gu W, Razquin C, Pastor P, Mateo I, Owen MJ, Faber KM, Jonsson PV, Combarros O, O'Donovan MC, Cantwell LB, Soininen H, Blacker D, Mead S, Mosley TH Jr, Bennett DA, Harris TB, Fratiglioni L, Holmes C, de Bruijn RF, Passmore P, Montine TJ, Bettens K, Rotter JI, Brice A, Morgan K, Foroud TM, Kukull WA, Hannequin D, Powell JF, Nalls MA, Ritchie K, Lunetta KL, Kauwe JS, Boerwinkle E, Riemenschneider M,

Boada M, Hiltunen M, Martin ER, Schmidt R, Rujescu D, Wang LS, Dartigues JF, Mayeux R, Tzourio C, Hofman A, Nöthen MM, Graff C, Psaty BM, Jones L, Haines JL, Holmans PA, Lathrop M, Pericak-Vance MA, Launer LJ, Farrer LA, van Duijn CM, van Broeckhoven C, Moskvina V, Seshadri S, Williams J, Schellenberg GD, Amouyel P, 2013 Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat. Genet* 45, 1452–1458. [PubMed: 24162737]

Lane RF, St George-Hyslop P, Hempstead BL, Small SA, Strittmatter SM, Gandy S, 2012 Vps10 family proteins and the retromer complex in aging-related neurodegeneration and diabetes. *J. Neurosci* 32, 14080–14086. [PubMed: 23055476]

Leloup N, Lössl P, Meijer DH, Brennich M, Heck AJR, Thies-Weesie DME, Janssen BJC, 2017 Low pH-induced conformational change and dimerization of sortilin triggers endocytosed ligand release. *Nat. Commun* 8, 1708. [PubMed: 29167428]

Liu QR, Lu L, Zhu XG, Gong JP, Shaham Y, Uhl GR, 2006 Rodent BDNF genes, novel promoters, novel splice variants, and regulation by cocaine. *Brain Res* 1067, 1–12. [PubMed: 16376315]

Ma X, Li-Ling J, Huang Q, Chen X, Hou L, Ma F, 2009 Systematic analysis of alternative promoters correlated with alternative splicing in human genes. *Genomics* 93, 420–425. [PubMed: 19442634]

Maness PF, Schachner M, 2007 Neural recognition molecules of the immunoglobulin superfamily: signaling transducers of axon guidance and neuronal migration. *Nat. Neurosci* 10, 19–26. [PubMed: 17189949]

McCarthy JJ, Saith S, Linnertz C, Burke JR, Hulette CM, Welsh-Bohmer KA, Chiba-Falek O, 2012 The Alzheimer's associated 5' region of the SORL1 gene cis regulates SORL1 transcripts expression. *Neurobiol. Aging* 33, e1–e8. [PubMed: 22244089]

McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM, 1984 Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA work group under the auspices of Department of Health and Human Services Task Force on Alzheimer's disease. *Neurology* 34, 939–944. [PubMed: 6610841]

Mehmedbasic A, Christensen SK, Nilsson J, Rüetschi U, Gustafsen C, Poulsen AS, Rasmussen RW, Fjorback AN, Larson G, Andersen OM, 2015 SorLA complement-type repeat domains protect the amyloid precursor protein against processing. *J. Biol. Chem* 290, 3359–3376. [PubMed: 25525276]

Nielsen MS, Gustafsen C, Madsen P, Nyengaard JR, Hermey G, Bakke O, Mari M, Schu P, Pohlmann R, Dennes A, Petersen CM, 2007 Sorting by the cytoplasmic domain of the amyloid precursor protein binding receptor SorLA. *Mol. Cell. Biol* 27, 6842–6851. [PubMed: 17646382]

Nielsen MS, Keat SJ, Hamati JW, Madsen P, Gutzmann JJ, Engelsberg A, Pedersen KM, Gustafsen C, Nykjaer A, Gliemann J, Hermans-Borgmeyer I, Kuhl D, Petersen CM, Hermey G, 2008 Different motifs regulate trafficking of SorCS1 isoforms. *Traffic* 9, 980–994. [PubMed: 18315530]

Offe K, Dodson SE, Shoemaker JT, Fritz JJ, Gearing M, Levey AI, Lah JJ, 2006 The lipoprotein receptor LR11 regulates amyloid β production and amyloid precursor protein traffic in endosomal compartments. *J. Neurosci* 26, 1596–1603. [PubMed: 16452683]

Pal S, Gupta R, Kim H, Wickramasinghe P, Baubet V, Showe LC, Dahmane N, Davuluri RV, 2011 Alternative transcription exceeds alternative splicing in generating the transcriptome diversity of cerebellar development. *Genome Res* 21, 1260–1272. [PubMed: 21712398]

Pardo LM, Rizzu P, Francescato M, Vitezic M, Leday GG, Sanchez JS, Khamis A, Takahashi H, van de Berg WD, Medvedeva YA, van de Wiel MA, Daub CO, Carninci P, Heutink P, 2013 Regional differences in gene expression and promoter usage in aged human brains. *Neurobiol. Aging* 34, 1825–1836. [PubMed: 23428183]

Pottier C, Hannequin D, Coutant S, Rovelet-Lecrux A, Wallon D, Rousseau S, Legallic S, Paquet C, Bombois S, Pariente J, Thomas-Anterion C, Michon A, Croisile B, Etcharry-Bouyx F, Berr C, Dartigues JF, Amouyel P, Dauchel H, Boutoleau-Bretonniere C, Thauvin C, Frebourg T, Lambert JC, Campion D, 2012 High frequency of potentially pathogenic SORL1 mutations in autosomal dominant early-onset Alzheimer disease. *Mol. Psychiatry* 17, 875–879. [PubMed: 22472873]

Prudencio M, Jansen-West KR, Lee WC, Gendron TF, Zhang YJ, Xu YF, Gass J, Stupati C, Stetler C, Rademakers R, Dickson DW, Buratti E, Petrucelli L, 2012 Misregulation of human sortilin splicing leads to the generation of a nonfunctional progranulin receptor. *Proc. Natl. Acad. Sci. U. S. A* 109, 21510–21515. [PubMed: 23236149]

- Pruunsild P, Kazantseva A, Aid T, Palm K, Timmusk T, 2007 Dissecting the human BDNF locus: bidirectional transcription, complex splicing, and multiple promoters. *Genomics* 90, 397–406. [PubMed: 17629449]
- Reese MG, Eeckman FH, Kulp D, Haussler D, 1997 Improved splice site detection in Genie. *J. Comput. Biol* 4, 311–323. [PubMed: 9278062]
- Reitz C, Tosto G, Vardarajan B, Rogaeva E, Ghani M, Rogers RS, Conrad C, Haines JL, Pericak-Vance MA, Fallin MD, Foroud T, Farrer LA, Schellenberg GD, George-Hyslop PS, Mayeux R, 2013 Independent and epistatic effects of variants in VPS10-d receptors on Alzheimer disease risk and processing of the amyloid precursor protein (APP). *Transl. Psychiatry* 3, e256. [PubMed: 23673467]
- Rogaeva E, Meng Y, Lee JH, Gu Y, Kawarai T, Zou F, Katayama T, Baldwin CT, Cheng R, Hasegawa H, Chen F, Shibata N, Lunetta KL, Pardossi-Piquard R, Bohm C, Wakutani Y, Cupples LA, Cuenco KT, Green RC, Pinessi L, Rainero I, Sorbi S, Bruni A, Duara R, Friedland RP, Inzelberg R, Hampe W, Bujo H, Song YQ, Andersen OM, Willnow TE, Graff-Radford N, Petersen RC, Dickson D, Der SD, Fraser PE, Schmitt-Ulms G, Younkin S, Mayeux R, Farrer LA, St George-Hyslop P, 2007 The neuronal sortilin-related receptor SORL1 is genetically associated with Alzheimer disease. *Nat. Genet* 39, 168–177. [PubMed: 17220890]
- Rohe M, Carlo AS, Breyhan H, Sporbert A, Militz D, Schmidt V, Wozny C, Harmeier A, Erdmann B, Bales KR, Wolf S, Kempermann G, Paul SM, Schmitz D, Bayer TA, Willnow TE, Andersen OM, 2008 Sortilin-related receptor with A-type repeats (SORLA) affects the amyloid precursor protein-dependent stimulation of ERK signaling and adult neurogenesis. *J. Biol. Chem* 283, 14826–14834. [PubMed: 18362153]
- Sager KL, Wu J, Leurgans SE, Rees HD, Gearing M, Mufson EJ, Levey AI, Lah JJ, 2007 Neuronal LR11/sorLA expression is reduced in mild cognitive impairment. *Ann. Neurol* 62, 640–647. [PubMed: 17721864]
- Scherzer CR, Offe K, Gearing M, Rees HD, Fang G, Heilman CJ, Schaller HC, Levey AI, Lah JJ, 2004 ApoE receptor LR11 in Alzheimer's disease: gene profiling of lymphoblasts mirrors changes in the brain. *Arch. Neurol* 61, 1200–1205. [PubMed: 15313836]
- Schmidt V, Sporbert A, Rohe M, Reimer T, Rehm A, Andersen OM, Willnow TE, 2007 SorLA/LR11 regulates processing of amyloid precursor protein via interaction with adaptors GGA and PACS-1. *J. Biol. Chem* 282, 32956–32964. [PubMed: 17855360]
- Thonberg H, Chiang HH, Lilius L, Forsell C, Lindström AK, Johansson C, Björkström J, Thordardottir S, Sleegers K, van Broeckhoven C, Rönneback A, Graff C, 2017 Identification and description of three families with familial Alzheimer disease that segregate variants in the SORL1 gene. *Acta Neuropathol. Commun* 5, 43. [PubMed: 28595629]
- Timmusk T, Palm K, Metsis M, Reintam T, Paalme V, Saarma M, Persson H, 1993 Multiple promoters direct tissue-specific expression of the rat BDNF gene. *Neuron* 10, 475–489. [PubMed: 8461137]
- Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG, 2012 Primer3–new capabilities and interfaces. *Nucleic Acids Res* 40, e115. [PubMed: 22730293]
- Verheijen J, van den Bossche T, van der Zee J, Engelborghs S, Sanchez-Valle R, Llado A, Graff C, Thonberg H, Pastor P, Ortega-Cubero S, Pastor MA, Benussi L, Ghidoni R, Binetti G, Clarimon J, Lleo A, Fortea J, de Mendonca A, Martins M, Grau-Rivera O, Gelpi E, Bettens K, Mateiu L, Dillen L, Cras P, de Deyn PP, van Broeckhoven C, Sleegers K, 2016 A comprehensive study of the genetic impact of rare variants in SORL1 in European early-onset Alzheimer's disease. *Acta Neuropathol* 132, 213–224. [PubMed: 27026413]
- Wasser CR, Masiulis I, Durakoglugil MS, Lane-Donovan C, Xian X, Beffert U, Agarwala A, Hammer RE, Herz J, 2014 Differential splicing and glycosylation of Apoer2 alters synaptic plasticity and fear learning. *Sci. Signal* 7, ra113. [PubMed: 25429077]
- Willnow TE, Andersen OM, 2013 Sorting receptor SORLA - a trafficking path to avoid Alzheimer disease. *J. Cell Sci* 126, 2751–2760. [PubMed: 23813966]
- Willnow TE, Petersen CM, Nykjaer A, 2008 VPS10P-domain receptors - regulators of neuronal viability and function. *Nat. Rev. Neurosci* 9, 899–909. [PubMed: 19002190]
- Wilson CM, Naves T, Saada S, Pinet S, Vincent F, Lalloué F, Jauberteau MO, 2014 The implications of sortilin/vps10p domain receptors in neurological and human diseases. *CNS Neurol. Disord. Drug Targets* 13, 1354–1365. [PubMed: 25345507]

Zhao Y, Cui JG, Lukiw WJ, 2007 Reduction of sortilin-1 in Alzheimer hippocampus and in cytokine-stressed human brain cells. *Neuroreport* 18, 1187–1191. [PubMed: 17589324]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Fig. 1.

SorLA/*SORL1*. (A) Schematic representation of the full-length human sorLA receptor containing a mosaic structure in the extracellular region with an N-terminally VPS10p- β -propeller followed by its 10CC domain, an YWTD β -propeller, an EGF domain, 11 CR domains, and 6 FnIII domains, a transmembrane (TM) region, and a C-terminal cytoplasmic tail. (B) The *SORL1* gene encoding full-length sorLA is located on chromosome 11 and contains 48 known exons. Numbers correspond to genome version hg19, where *SORL1* maps to 121.322.912–121.504.471. The region encoded by each exon is indicated under the corresponding domain in panel A. The position of the long noncoding RNA molecule named 51A is also displayed.

Fig. 2.

Identification of exon 30B in the *SORL1* gene structure. (A) RT-PCR using cDNA produced from human RNA of the total brain (left) or temporal lobe (right) and amplified using either a 35-cycle or 30-cycle PCR protocol, respectively. Arrow indicates a band at 239 bp, and asterisk indicates band at 342 bp. (B) Schematic model for genomic DNA surrounding the novel exon 30B (red) between exon 30 and exon 31 (green). Transcription initiation from the transcriptional start site (TSS) in intron 28 is indicated, and splicing of exon 30B to exon 31 leads to the novel *SORL1* transcript AK294404 as provided in USCS's Genome Browser. The ATG start site used for translation of a polypeptide in-frame with the coding region of exon 31 is indicated within exon 30B. Exonic sequences not included in the spliced transcript are indicated by broken rectangles. Primer positions for exon 30B (forward) and exon 32 (reverse) for use in RT-PCR to detect *SORL1-30B* are shown as black arrows. PCR

with the spliced transcript as a template is predicted to give a 239-bp product, in comparison to genomic DNA that would produce a 4809-bp fragment. The sequence of the indicated 103-bp intron 30B is provided in Fig. 5. (C) OneStep RT-PCR using RNA from human tissues or cell lines as a template as indicated. The primer pair for exon 30B (forward) and exon 32 (reverse) produces a 239-bp product (by a 35-cycle PCR) as indicated by the arrow, but also a product (asterisk) including intron 30B in some of the samples (upper panel). The primer pair for exon 2 (forward) and exon 3 (reverse) produces a 174-bp fragment (by a 30-cycle PCR) representing full-length *SORL1* (lower panel). (D) The genomic DNA for the *SORL1* locus around exon 2 (pink) and exon 3 (violet) including primer positions used to determine the presence of full-length *SORL1*. The PCR product with spliced mRNA as a template is 174 bp long compared with the 8169-bp fragment if intron 2 was not spliced out. The localization of genomic exon 30B is indicated by a broken rectangle, and this exon is not included in the BC137171 transcript encoding full-length sorLA.

Fig. 3.

The 5'-UTR of *SORL1-30B*. (A) Agarose gel electrophoresis showing the results of 5'-RACE. DNA sequencing determined the exact size of the DNA fragment to be 1508 bp. This fragment was identified by the use of human total brain RNA as a template. Negative control using water as a template returned no result. (B) 5'-RACE using SMARTer cDNA prepared from human cerebellum RNA was used for In-Fusion cloning and inserts corresponding to products amplified from the transcriptional start site (TSS) were cut out from the pRACE vector and analyzed by agarose gel electrophoresis. The exact start position of the 8 fragments (a-h) revealed heterogeneous start sites as indicated in Fig. 4. (C) The sequence of the 1508-bp fragment was aligned to the human genome hg19 using BLAT, leading to identification of a long 5'UTR spanning a region of the *SORL1* gene that normally represents exon 29 (blue rectangle with a broken line) and exon 30 (orange rectangle with a broken line) of the full-length gene. These 2 regions do not undergo splicing of flanking introns in the 30B transcript and are thus noncoding regions and part of the 5'UTR of AK294404. The sequence corresponding to fragment obtained by 5'-RACE is shown in gray, and the position of the identified TSS is located within the normal intron 28. Position of the reverse primer is indicated for exon 30B (sorla-GSP rev_a) and the forward CDS primer.

TCGGTAAAGTTTCATGGAATGGATTAATTAACAAAGTGTGTGCTCTGTTAGATTAGGTTTGGCTTTTGCCTTTCAAGGC
 ↓*a, b* Mapped TSS ↓*d, g* *h*↓ *c*↓ *e*↓*f*↓ *intron 28*
 ATGGTTTAGCCAGACAGTCCTTTGCCAGAACTCACTGCTGTTCCCCCTTGCTTCTGGGCACCACTGCTGCTTCGACCCC
 ↓ DBTSS
 TCAGCATGGAAGTTCCTGGACTTCATGGATGTTTGTTCCTCTGTAG **CCCAAGATCCTGAGTTCACAAAGTATGTGAT** *exon 29*
GAGTTCGGTTTCCAGTGTGAGTGGAGTGTGCATCAGTTTGATTTGGAAGTGCACGGGATGGATGATTGCGGCGATT
 Asn1358Ser
ATTCTGATGAAGCCCACTGCGGTAAGATGTCCAGTCTGCCCTCCACATCCTAATCCTCTAGTTAACAGTGTATGCCTGC
 AGTTACAGGGACACTCACCGCAACCCCACTGCAGACTTGCCCTTCCCGTAACAGCAGCAGTTGCTGCGCAGTTCCTG
 AATACTCATACGAGTTGAGTCTTGAAGGGGGCAGCTGGCACCCAGCAGGTGCTAGGGAAAAGTTTATTGTATCTCCTTT
 rs12272618
 CCTTTCTG **GGTTGGGAAGACCACATTATGTATAAGGAAGTGGTAGATAAGGAATTATGGTAAATATTATGGAGTGAC** *intron 29*
 CTCCCCCTAAAAGTGACAGGCCACAATGTAATAGTTATTGTTCATGATTGGATTACCATCAATAGTTGTTTTATTCTCTTT
 GGTAATAATGTCAGAAATAAAATTCAGGTGATTGTTCTCAGGATATTGTAGTATAGATCAGAGGAACATGGGTTAGGGGC
 TGGGAAGATGGGGTTTGTGTTAGTTTGTAGGTGTATTTTTGTTCCTCCATTGGGTCCATGAAGTGCCTTATCTCTTTA
 TGGGTGGGAAGTGTGGTTGCCCTGGAACCTGGGTCTGGAGGAGGATGCCATAGAGTTGGAGCCCTGTGTTACACATGAT
 GCAGGGAAGCAACTGATGATGTGGTTGATCTCTGCAG **AAAACCCCAAGAGCCCAAACTGCTCCCGCTACTTCCAG** *exon 30*
 rs2276412
TTTCGGTGTGAGAATGGCCACTGCATCCCCAACAGATGGAAATGTGACAGGGAGAACTGACTGTGGGGACTGGTCTGATG
AGAAGGATTGTGGAGGTAAAGAGCCCTGGGGCTGGGTTAGCCCCATAACCAAGACACACAGAATAGTTCCACCAGC
 TATGCAGGATGTGCCCTGGCTGATTCCTGTGTTTGGGGGCATATTTTGTGAGAGGAGTGACTCAAATTTTTTCTGTTC
 CTCTTTCTCATTTTGAATAAGTGTGGTTGGTTTCTCTATTTTCAGGCTATTTATTCCGTATTTGCAGGGCAGAGTGA
 GAAGTCTAGACCTTCTACTCAAAGTATGGTTTGCAACCAGTAGCATCTGCATCTCCTTGGAACCTGTTAGAAATGTA
 GAATCTCTGACCTCACCCAGACCTACAGAGTCAGAAATCTGCATGTCTAATAAGCCCCAGGTCCTGTGTGCTCCACAAT *intron 30*
 GCATTTGAGGAGCACTCA **TGTGGAATGCAGACTTGCCC**CTTTCTTTCTGGCATTCTCATCTCTGAACCAAGGTCGTGGA
 GTAGTTGCGGTTGAATTAAGTGAACACAATGCTTTGCCCTCAGTAGGAATTCCTTAGCCACCTGTGCCCTCCTCAGCC
 ACCTGTTTCTTTACCAGCCCTGCAAAAGCAAATTTATCTGAATATTAACCGAAATCAAATATAGTAAGGGAATCTA
 CTCCCCAAAGGGAAAAGAATATTGCCTCTTCTGACAACCCATTATAC **ATG ATT TCT AGA TTA GCC GCT GCT** *exon 30B*
 Met Ile Ser Arg Leu Ala Ala Ala
 Gln Ser Cys Ala Ser His Trp Pro Pro Ser Ala
CAG AGC TGT GCC AGT CAC TGG CCT CCC TCA GCA G

Fig. 4.

DNA sequence of the 5'-UTR of *SORL1-30B*. The DNA sequence of the 5'-UTR of *SORL1-30B* is shown including positions a-h of the experimentally mapped TSS from the human cerebellum (↓) and from total brain RNA (↓). The position described in the database for transcript start sites (↓, DBTSS, www.dbtss.hgc.jp) is identical to experimentally determined positions a and b. The boxed sequence indicates the primer sorLA-GSP rev_a used in 5'-RACE. Letters shown in white on a black background represent SNPs in the *SORL1* gene previously identified to associate with Alzheimer's disease, and letters in *italics* correspond to the UTR sequence provided for AK294404 in the GenBank database (shown as the boxed area upstream of exon 30B in Fig. 3C).

Fig. 5.

Translated DNA sequence for *SORL1* exon 30B. (A) The Kozak consensus sequence for exon 1 and exon 30B. For a “*strong*” consensus, the nucleotides at positions +4 and –3 must both match the RCCATGR consensus, i.e., be purines (R = A/G). An “*adequate*” consensus has only 1 of these sites conserved, whereas a “*weak*” consensus has neither. (B) DNA sequencing of the 239-bp PCR fragment corresponds to the expected part of exon 30B (red), exon 31 (green), and exon 32 (yellow). The positions of the primers used for RT-PCR are boxed. The sequence for the full intron 30B (103 bp) and parts of the flanking regions of intron 31 (total 4466 bp) are also included to identification of splice site boundaries. The predicted amino acid sequence for sorLA-30B starts at the ATG codon within exon 30B and produces an ORF including a putative 20 amino acid signal peptide (SP) in frame with reading frame of exon 31 and is shown above the DNA sequence. The arrow (↓) points to an identified free N-terminal suggesting that the mature sorLA-30B starts with Ser-21. (C) Schematic presentation of the sorLA-30B receptor comprising only 3 CR domains and 6 FnIII domains in its luminal fragment.

Fig. 6. *SORL1-30B* expression in cerebellum AD and non-AD brain samples. (Panel A) RT-PCR using cDNA produced from RNA of the cerebellum of 25 AD (A) and 25 non-AD (C) individuals. The 2 bands marked with an arrow and with an asterisk are described in Fig. 2. (Panel B) Quantification of band intensities from the gel of PCR products from reaction with 30B/32 (as in panel A) and 2/3 primer pairs (not shown). Data are presented as relative levels of the lower band (left), the upper band (center), or the exon 2-3 band (right) in AD samples relative to the mean of the band intensity of the control bands (set to 100%). (Panel C) The ratio between intensities of the upper and the lower bands (left), between the lower and the exon 2-3 bands (center), and between the upper and the exon 2-3 bands (right). n.s., not significant.

Fig. 7.

SorLA-30B expression. (A) Western blot analysis for HEK293 cells nontransfected or stably expressing sorLA-30B. A band for sorLA-30B above the 98-kDa marker is seen only for transfected HEK293 cells and indicated by a black arrowhead (left). Endogenous sorLA-30B expression was evident by similar Western blot analysis of lysates prepared from nontransfected NT2 cells (right). (B, C) HEK293 cells stably transfected with constructs for full-length sorLA (B) or sorLA-30B (C) were plated on glass cover slides. Cells were fixed, permeabilized, and stained for sorLA and GM130 (top row), sorLA and EEA1 (middle row), and sorLA and TfR (bottom row) using fluorescently labeled secondary antibodies as indicated and examined by confocal microscopy. Representative pictures from 3 independent

experiments are shown, and right hand pictures show merged images, including DAPI nuclear stain.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1

Database summary for predicted *SORL1* transcripts

GenBank ^a	Length (bp)	ENSEMBL ^b	UniProt ^c	Protein (aa)	Molecular mass (kDa)	Predicted protein domain structure
BC137171	6725	ENST00000260197	Q92673	2214	248.4
AK295878	3836	ENST00000525532	E9PPB3	1158	129.7
AK295090	3692	ENST00000532694	E9PS32	1060	118.5
AK302589	3501	ENST00000534286	E9PP43	1124	125.8
AK294404	3023	ENST00000527934	E9PKB0	829	92.3

The predicted protein domain structures follow the schematic presentation in Fig. 1A. The red bars with question marks correspond to amino acid sequences not present in the 2214 residue receptor.

^aGenBank; www.ncbi.nlm.nih.gov.

^bENSEMBL; www.ensembl.org.

^cUniProt; www.uniprot.org.

Table 2

Splice site prediction

Donor site prediction ^a			Acceptor site prediction ^a		
Intron/Exon	Score	Intron/Exon sequence	Exon/Intron	Score	Exon/Intron sequence
28 / 29	0.72	TTCCCTCTGT AG CCCCAAGATCC	29 / 29	0.99	AACTGCC GT AAGATG
29 / 30	0.88	TGATCTCTGCC AG AAAA'CCCCAC	30 / 30	0.97	TGTGGAG GT AAGAGG
			30B / 30B	0.80	TCAGCAG GT ACAGCT
30B / 31	0.98	TGTTTTTCTC AG ATTCACAAT	31 / 31	0.95	TTGCTTG GT GAGTTC
31 / 32	0.46	CTTTTATTTT AG CAAA'CGTCAC			

^aPrediction of splice site probability was performed using the web-based splice finder software Genie (Reese et al., 1997) (http://www.fruitfly.org/seq_tools/splice.html). Maximum score by this software is 1.0, and cutoff is 0.4. Color coding for exon sequences is identical to the one used in Figures 2 and 3, where blue, orange, red, green, and yellow correspond to exons 29, 30, 30B, 31, and 32, respectively. The nucleotides at the splice site are displayed in larger font.