Exome sequencing identifies novel AD-associated genes.

Authors:

Henne Holstege[1,2,3*#], Marc Hulsman[1,2,3*#], Camille Charbonnier[4*], Benjamin Grenier-Boley[5], Olivier Quenez[4], Detelina Grozeva[6], Jeroen G.J. van Rooij[7], Rebecca Sims[6], Shahzad Ahmad[8], Najaf Amin[8,57], Penny J. Norsworthy[9], Oriol Dols-Icardo[10], Holger Hummerich[9], Amit Kawalia[11], Philippe Amouyel[5], Gary W. Beecham[12], Claudine Berr[13], Joshua C. Bis[14], Anne Boland[15], Paola Bossù[16], Femke Bouwman[1], Dominique Campion[4], Antonio Daniele[17,18], Jean-François Dartigues[19], Stéphanie Debette[19], Jean-François – Deleuze[20], Nicola Denning[21], Anita L DeStefano[22,23,24], Lindsay A. Farrer[22,25,26,27,28], Nick C. Fox[29], Daniela Galimberti[30], Emmanuelle Genin[31], Jonathan L. Haines[32], Clive Holmes[33], M. Arfan Ikram[7,8,34], M. Kamran Ikram[7,8], Iris Jansen[1,35], Robert Kraaij[36], Marc Lathrop[37], Evelien Lemstra[1], Alberto Lleó[10,38], Lauren Luckcuck[6], Rachel Marshall[6], Eden R Martin[12,39], Carlo Masullo[40], Richard Mayeux[41], Patrizia Mecocci[42], Alun Meggy[21], Merel O. Mol[7], Kevin Morgan[43], Benedetta Nacmia[44], Adam C Naj[45,46], Pau Pastor[47], Margaret A. Pericak-Vance[12], Rachel Raybould[21], Richard Redon[48], Anne-Claire Richard[4], Steffi G Riedel-Heller[49], Fernando Rivadeneira[36], Stéphane Rousseau[4], Natalie S. Ryan[29], Salha Saad[6], Pascual Sanchez-Juan[50], Gerard D. Schellenberg[46], Philip Scheltens[1], Jonathan M. Schott[29], Davide Seripa[51], Gianfranco Spalleta[52], Betty Tijms[1], André G Uitterlinden[8,36], Sven J. van der Lee[1,2,3], Michael Wagner[53,54], David Wallon[4], Li-San Wang[46], Aline Zarea[4], Marcel J.T. Reinders[2], Jordi Clarimon[10], John C. van Swieten[7], John J. Hardy[55,29], Alfredo Ramirez[11,53], Simon Mead[9], Wiesje M. van der Flier[1,56], Cornelia M van Duijn[8,57], Julie Williams[21], Gaël Nicolas[4*#,] Céline Bellenguez[5*], Jean-Charles Lambert[5*#]

*Authors contributed equally to this work

[#] To whom correspondence should be addressed

— Henne Holstege: h.holstege@amsterdamumc.nl
— Marc Hulsman: m.hulsman@amsterdamumc.nl
— Gael Nicolas: gaelnicolas@hotmail.com
— Jean-Charles Lambert: jean-charles.lambert@pasteur-lille.fr

Affiliations:

(1) Alzheimer Center, Department of Neurology, VU University Medical Center, Neuroscience Campus Amsterdam, Amsterdam, The Netherlands; **(2)** Department of Clinical Genetics, VU University Medical Center, Neuroscience Campus Amsterdam, Amsterdam, The Netherlands; **(3)** Delft Bioinformatics Lab, Delft University of Technology, Delft, The Netherlands; **(4)** Normandie Univ, UNIROUEN, Inserm U1245 and Rouen University Hospital, Department of Genetics and CNR-MAJ, F 76000, Normandy Center for Genomic and Personalized Medicine, Rouen, France; **(5)** Univ Lille, Inserm, CHU Lille, Institute Pasteur de Lille, U1167 - RID-AGE - Risk factors and molecular determinants of age-related diseases; Institute Pasteur de Lille, University of Lille, Lille Cedex, France; **(6)** Division of Psychological Medicine and Clinical Neuroscience, School of Medicine, Cardiff University, Cardiff, UK; **(7)** Department of Neurology, Erasmus Medical Centre, Rotterdam, The Netherlands; **(8)** Department of Epidemiology, Erasmus MC University Medical Center Rotterdam, Rotterdam, Netherlands; **(9)** MRC Prion Unit at UCL; (**10)** Sant Pau Biomedical Research Institute, Hospital de la Santa Creu i Sant Pau, Universitat Autònoma de Barcelona, Barcelona, Spain; **(11)** Division of Neurogenetics and Molecular Psychiatry, Department of Psychiatry and Psychotherapy, University of Cologne, Medical Faculty, 50937 Cologne, Germany; **(12)** John P Hussman Institute for Human Genomics, Miller School of Medicine, University of Miami; **(13)** Univ. Montpellier, Inserm U1061, Neuropsychiatry: epidemiological and clinical research, PSNREC; **(14)** Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA (USA); (**15)** Université Paris-Saclay, CEA, Centre National de Recherche en Génomique Humaine, 91057, Evry, France; **(16)** IRCCS Fondazione Santa Lucia, Department of Clinical and Behavioral Neurology, Experimental Neuro-psychobiology Lab Via Ardeatina, 306, I-00179 Roma, Italy; **(17)** Department of Neuroscience, Università Cattolica del Sacro Cuore, Rome, Italy; **(18)** Neurology Unit, IRCCS Fondazione Policlinico Universitario A. Gemelli, Rome, Italy; (**19)** Univ. Bordeaux, Inserm U1219, Bordeaux Population Health Research Center, Bordeaux, France; CHU

de Bordeaux, Department of Neurology, Bordeaux, France; **(20)** Université Paris-Saclay, CEA, Centre National de Recherche en Génomique Humaine, 91057, Evry, France; **(21)** UKDRI@ Cardiff, School of Medicine, Cardiff University, Cardiff, UK; **(22)** Department of Biostatistics, Boston University School of Public Health; **(23)** Department of Neurology, Boston University School of Medicine; **(24)** Framingham Heart Study; **(25)** Dept. of Medicine (Biomedical Genetics), Boston Univ. School of Med; **(26)** Department of Neurology, Boston University School of Medicine; **(27)** Department of Ophthalmology, Boston Univ. School of Medicine; **(28)** Department of Epidemiology, Boston Univ. School of Public Health; **(29)** Dementia Research Centre, UCL Queen Square Institute of Neurology, UK Dementia Research Institute; **(30)** University of Milan, Centro Dino Ferrari, CRC Molecular basis of Neuro-Psycho-Geriatrics diseases, Milan, Italy; **(31)** Univ Brest, Inserm, EFS, CHU Brest, UMR 1078, GGB, F-29200, Brest, France; **(32)** Population & Quantitative Health Sciences and Cleveland Institute for Computational Biology, School of Medicine, Case Western Reserve University, Cleveland, Ohio USA; **(33)** Clinical and Experimental Science, Faculty of Medicine, University of Southampton, Southampton, UK; **(34)** Department of Radiology, Erasmus MC University Medical Center, Rotterdam, The Netherlands; **(35)** Complex Trait Genetics Lab, CNCR, VU University, Amsterdam; **(36)** Department of Internal Medicine, Erasmus MC University Medical Center Rotterdam, Rotterdam, Netherlands; **(37)** McGill University and Genome Quebec Innovation Centre, 740 Doctor Penfield Avenue, Montreal, QC, H3A 0G1, Canada; **(38)** Network Center for Biomedical Research in Neurodegenerative Diseases (CIBERNED), Madrid, Spain; **(39)** Department of Human Genetics , University of Miami Leonard M. Miller School of Medicine; **(40)** Istituto di Neurologia Policlinico Universitario A. Gemelli, 00168, Rome, Italy; **(41)** Columbia University; **(42)** Institute of Gerontology and Geriatrics, Department of Medicine and Surgery, University of Perugia, Italy; **(43)** Human Genetics, School of Life Sciences, University of Nottingham, UK NG7 2UH; **(44)** Department of Neuroscience, Psychology, Drug Research and Child Health , University of Florence, Italy; **(45)** Department of Biostatistics, Epidemiology, and Informatics; Perelman School of Medicine, University of Pennsylvania; **(46)** Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania; **(47)** Memory Disorders Unit, Department of Neurology, Hospital Universitari Mutua de Terrassa, Terrassa, Barcelona, Spain; **(48)** Université de Nantes, CHU Nantes, CNRS, INSERM, l'institut du thorax, F-44000, Nantes, France; **(49)** Institute of Social Medicine, Occupational Health and Public Health, Medical Faculty, University of Leipzig, Leipzig, Germany; **(50)** Neurology Service and Centro de Investigación en Red de Enfermedades Neurodegenerativas (CIBERNED), Marques de Valdecilla University Hospital (University of Cantabria and IDIVAL), Santander, Spain; **(51)** Laboratory of Gene Therapy, IRCCS Casa Sollievo della Sofferenza, San Giovanni Rotondo, FG, Italy; **(52)** Laboratory of Neuropsychiatry, IRCCS Santa Lucia Foundation, Rome, Italy; **(53)** Department of Neurodegenerative Diseases and Geriatric Psychiatry, University Hospital Bonn, Bonn, Germany; **(54)** DZNE, German Center of Neurodegenerative Diseases, Bonn, Germany; **(55)** Department of Neurodegenerative Disease, Reta Lila Weston Laboratories, Queen Square Genomics, UCL Dementia Research Institute, Wing 1.2 Cruciform Building, Gower Street, London WC1E 6BT; **(56)** Department of Epidemiology &Biostatistics, VU University Medical Center, Neuroscience Campus Amsterdam, Amsterdam, The Netherlands; **(57)** Nuffield Department of Population Health, University of Oxford, Oxford, United Kingdom.

Abstract

Background: With the development of next-generation sequencing technologies, it is possible to identify rare genetic variants that influence the risk of complex disorders. To date, whole exome sequencing (WES) strategies have shown that specific clusters of damaging rare variants in the *TREM2*, *SORL1* and *ABCA7* genes are associated with an increased risk of developing Alzheimer's Disease (AD), reaching odds ratios comparable with the *APOE-ε4* allele, the main common AD genetic risk factor. Here, we set out to identify additional AD-associated genes by an exome-wide investigation of the burden of rare damaging variants in the genomes of AD cases and cognitively healthy controls.

Method: We integrated the data from 25,982 samples from the European ADES consortium and the American ADSP consortium. We developed new techniques to homogenize and analyze these data. Carriers of pathogenic variants in genes associated with Mendelian inheritance of dementia were excluded. After quality control, we used 12,652 AD cases and 8,693 controls for analysis. Genes were analyzed using a burden analysis, including both non-synonymous and loss-of-function rare variants, the impact of which was prioritized using REVEL.

Result: We confirmed that carrying rare protein-damaging genetic variants in *TREM2*, *SORL1* or *ABCA7* is associated with increased AD-risk. Moreover, we found that carrying rare damaging variants in the microglial *ATP8B4* gene was significantly associated with AD, and we found suggestive evidence that rare variants in *ADAM10*, *ABCA1*, *ORC6*, *B3GNT4* and *SRC* genes associated with increased AD risk. High-impact variants in these genes were mostly extremely rare and enriched in AD patients with earlier ages at onset. Additionally, we identified two suggestive protective associations in *CBX3* and *PRSS3*. We are currently replicating these associations in independent datasets.

Conclusion: With our newly developed homogenization methods, we identified novel genetic determinants of AD which provide further evidence for a pivotal role of APP processing, lipid metabolism, and microglia and neuroinflammatory processes in AD pathophysiology.

Introduction

Alzheimer's disease is the leading cause of dementia and its impact will continue to grow due to the increase in life expectancy (1) Beyond rare autosomal dominant forms of early onset AD (less than 1% of all AD cases), the common complex form of AD has an estimated heritability of ~70% (2) This heritability can be explained by the aggregated effect of many genes associated with AD risk. Deciphering this genetic component to the gene or even to the variant level offers a unique window of opportunity to (i) better define the aetiology underlying the disease; and (ii) to develop polygenic risk scores that may predict *who* will develop AD *before* clinical symptoms occur. Comprehensive knowledge of disease etiology is thus essential for the future development of treatment strategies, which will likely be most effective when administered to those with relevant genetic risk, before irreparable damage to brain cells has occurred.

With such ambitious objectives, important efforts have been made to characterize the comprehensive genetic landscape of AD. With the advent of genome wide association studies (GWAS) based on DNA chips, numerous common genetic risk factors/loci have been associated with the risk of AD over the 10 last years (3, 4). However, our knowledge of the genetic component underlying AD is far from complete. While further efforts are underway to capture additional genetic information using GWASs, this approach is not really designed to efficiently capture the effect of rare (and even more singleton) variants on disease risk. However, rare variants are expected to explain at least part of the missing heritability of most complex diseases, including AD.

With the development of the next-generation sequencing technologies, it is possible to identify rare variants in genetic sequences. To date, whole exome sequencing (WES) strategies have shown that rare missense or loss-of-function variants in the *TREM2*, *SORL1* and *ABCA7* genes are associated with an increased risk of developing AD with a moderate to high effect (5-9). For the *SORL1* gene, loss of function variants were associated with an increased risk of AD with an odds ratio in ranges that were not observed since the identification of the main AD genetic risk factor, the common *APOE-ε4* allele (9-12).

The detection of additional AD associated genes by investigating the differential burden of rare damaging variants between AD cases and controls requires very large sample sizes. Variants are often very rare such that many cases and controls are necessary to collect enough evidence for a statistically significant association. In addition, beyond issues of statistical power, WES analyses need to take into account common technical biases leading to strong batch effects that can have important impacts on the generated results with a risk to generate false positives or negatives. Furthermore, all genes have unique features, both functionally and genetically, and this is reflected by the diverse characteristics of variants that drive their association with AD. Using WES, unique variants may be observed in very few or only single carriers which requires alternate interpretation strategies compared to the classical GWAS analyses in which all measured variants are common. For these reasons, genome-wide comparisons of rare variants in AD cases and controls have likely not yet led to the identification of novel AD-associated genes beyond *SORL1*, *ABCA7* and *TREM2*,(12)

Here, to identify an association between the burden of rare coding variants at the gene level, we developed novel analysis methods to study the largest WES dataset available worldwide encompassing 21,345 samples (12,652 AD cases and 8,693 controls). This unique effort led to the identification of 11 genes associated with AD-risk, of which rare variants in eight genes were not previously significantly associated with AD genetic risk. Per gene, we report the effect sizes of the variant burden after a final refinement analysis that takes into account that a uniform exome-wide analysis does not comply with gene-specific idiosyncrasies.

Methods

Sample

We analyzed the exome sequences of 25,982 individuals: sequence data from 15,088 individuals was collected as part of the Alzheimer Disease European Sequencing consortium (ADES) and sequence data from 11,365 individuals was obtained from the Alzheimer's Disease Sequencing Project (ADSP) (12), see **Table S1** for samples contributed per study. The total sample comprised 14,658 AD cases and 10755 controls (569 were N/A). For sample description, see supplemental data. DNA samples were sequenced using a paired-end Illumina platform, whole exome sequences (WES) was generated using different exome capture kits (**Table S2**), a subset of the sample was sequenced using whole genome sequencing (WGS) (**Figure S1, Table S2**).

Data processing, Quality control (QC) and genotype calling

Raw sequencing data from all studies were collected on a single site and processed relative to the GRCh37 reference genome, using a uniform pipeline as described in detail in the supplementary methods. On the merged sample, we performed a sample QC (**Figure 1a**) after which 21,345 samples were available for analysis: 12,652 cases (4,060 EOAD, onset ≤ 65 years) and 8,693 controls. The variant QC was applied as described in **Figure 1b**; variant selection and annotation was performed as described in **Figure 1c**: The burden analysis was performed at the gene level based on protein-coding Ensembl transcripts with a 'Gencode basic' tag. Missense variants were annotated using REVEL (Rare Exome Variant Ensemble Learner) (13, 14) and LOF variants were annotated using LOFTEE (15). We selected variants that were estimated to have at least one carrier, and had a minor allele frequency (MAF) of <1%. We removed variants with >20% genotyping missingness or that did not pass a filter for differential missingness between the EOAD, LOAD and control groups (genotypes with a read depth <6 are considered missing, see supplement).

Gene burden test: Variant impact categories and thresholds

Variants were divided in four ***deleteriousness categories***: a LOF category, and 3 missense categories: REVEL ≥ 75, REVEL 50-75 and REVEL 25-50 (**Figure 1c**). Based on these, we constructed four ***deleteriousness thresholds*** in which we incrementally added variants with lower levels of variant predicted deleteriousness: first only LOF variants, then LOF variants + variants with a REVEL score ≥75, then LOF + REVEL≥50, and last LOF + REVEL≥25. This allows us to concentrate on the test which provides maximum evidence for a differential burden-signal. Multiple testing correction was performed across all performed tests (up to 4 per gene).

Gene burden test: age-at-onset association

Based on previous findings in *SORL1*, *TREM2* and *ABCA7* (16), we expect an enrichment of high impact rare risk variants in early onset cases relative to late onset cases. Therefore, we applied a test based on ordinal logistic regression, in which the genetic risk for AD is considered to increase in the sample categories: i.e. $burden_{EOAD} > burden_{LOAD} > burden_{control}$. This test is optimally suited for picking up differential variant loads between the sample categories, and can also detect regular case-control signals for which genetic risk is equally distributed across EOAD and LOAD cases ($burden_{EOAD} \sim burden_{LOAD} > burden_{control}$) as well as EOAD-specific signals ($burden_{EOAD} > burden_{LOAD} \sim burden_{control}$). We considered an additive model, while correcting for population covariates (see supplement). Genes were only tested if the cumulative minor allele count (cMAC) of predicted damaging variants was ≥10. Genes were considered suggestively associated with AD if the False Discovery Rate (FDR) (Benjamini-Hochberg procedure (17) as <20% (FDR<0.2). Genes were considered significantly associated with AD in our discovery sample when the corrected p was <0.05 after family-wise correction using the Holm-Bonferoni procedure (18).

Gene burden test: Testing for an age-at-onset or a deleteriousness-category effect

To test whether the burden of damaging variants increased (or decreased for protective variants) towards younger patients, an ordinal regression was performed using only cases (no controls). Cases were grouped in 4 age-at-onset bins: ≤65, (65-70], (70-80] and >80. A significant effect (FDR < 0.05) signaled that there was a difference in enrichment between young and older cases. To determine if there was a significant trend in effect sizes between the different deleteriousness categories (REVEL 25-50, 50-75, 75-100 and LOF), an ordinal logistic regression test was

performed with constrained beta's $|b_{REVEL\ 25-50}| \leq |b_{REVEL\ 50-75}| \leq |b_{REVEL\ 75-100}| \leq |b_{LOF}|$, and compared to a H0-model with a single beta (see supplement).

Carrier frequency and odds ratios

A carrier of a set of variants was defined as a sample for which the summed dosage of those variants was ≥0.5. Carrier frequencies (CFs) were determined as #carriers / #samples. Effect sizes (odds ratios, ORs) of the ordinal logistic regression can be interpreted as weighted averages of the OR of being an AD case versus control, and the OR of being an early-onset AD case or not. Ordinal odds ratios were calculated for each test, as well as separately for the 4 variant categories REVEL 25-50, 50-75, 75-100 and LOF. Next to ordinal ORs, we estimated 'standard' ORs. This was done across all samples (case/control), as well as per age category (EOAD versus controls and LOAD versus controls), as well as for smaller age-at-onset categories: ≤65 (EOAD), (65-70], (70-80] and >80 using multinomial logistic regression, while correcting for 6 PCA covariates.

Sensitivity analysis

A sensitivity analysis was performed to determine if effects were potentially due to age differences between cases and controls. We constructed an age-matched sample, by dividing samples in strata based on age/age-at-onset, with each stratum covering 2.5 years. Case/control ratios in all strata were kept between 0.1 and 10 by down sampling respectively controls or cases. Subsequently, samples were weighted using the propensity weighting within strata method proposed by Posner and Ash (19). Finally, a case-control logistic regression was performed both on the unweighted and weighted case-control labels, and estimated odds ratios and confidence intervals were compared.

Variant-specific analysis

We performed a variant-specific analysis of the genes considered as significantly or suggestively associated with AD, to detect gene-specific idiosyncrasies not covered by our uniform exome-wide analysis. We checked for outlier variants among those that were included in the burden test, determining which ones had a significantly lower or opposite effect size (fisher exact test) compared to other included variants of the same category (missense or LOF). Furthermore, we determined which missense or potential LOF variants did associate with AD (logistic regression test, at least 15 carriers), irrespective of REVEL/LOFTEE or MAF thresholds. We performed

corrections for multiple testing per gene using FDR, reporting only variants with a threshold of FDR < 0.2 **(Table S3).** We manually removed and added these variants to the burden tests, in order to calculate, next to standard odds ratios, also refined odds ratios.

Results

Sample description:

After sample QC (**Figure 1a**), 21,345 participants were included in the main analysis (12,652 cases; 8,693 controls) (**Table 1**). AD cases were separated in EOAD cases with age at onset ≤ 65 (n=4,060) and LOAD cases (N=8,592). All demographic data are available in **Table S1.** As expected, cases were more likely to carry at least one APOE ε4 allele: the fraction of homozygous APOE ε4 carriers was 6.6% of the cases vs. 0.9% of the controls; fraction of heterozygous APOE ε4 carriers was 40.6% of the cases vs 18.4% of the controls (**Table 1**).

Burden tests using different deleteriousness thresholds

We detected a total of 13,522,252 variants in these individuals, and 7,674,898 variants passed quality control (**Figure 1b**). These variants were annotated according to four predicted deleteriousness categories based on LOFTEE score for LOF variants and the REVEL prediction score for missense variants. Finally, we selected 407,032 coding missense and loss of function (LOF) variants with MAF <1% based on criteria as described in the methods (**Figure 1c**). We used four deleteriousness thresholds by incrementally including variants with on lower levels of variant predicted deleteriousness: respectively LOF (n=56,565), LOF + REVEL≥75 (n=109,576), LOF + REVEL≥50 (n=208,720), and LOF + REVEL≥25 (n=407,032).

Among the 19,822 autosomal protein-coding genes considered in our annotation, we tested 13,299 genes with at least 10 minor alleles (cumulative minor allele count or cMAC ≥ 10) appertaining to the LOF+REVEL≥25 variant threshold. For the remaining genes, the burden of variants per gene was considered too low (cMAC<10) to infer any dependable signal.
For the LOF+REVEL≥50, the LOF+REVEL≥75 and the LOF-only thresholds, respectively 9,255, 5,781 and 3,233 genes reached the minimum of having at least cMAC ≥10 to allow testing (**Figure 2**). In sum, 31,568 tests were performed across 13,299 genes. Of note, since we tested each gene for

having a differential variant burden in cases and controls for different deleteriousness thresholds, a single gene could theoretically be identified multiple times in the burden test.

Identification of genes for which rare variant-burden associates with AD risk

We performed 31,568 tests in our analysis, and the genetic inflation of our analysis model was λ=1.038 **(Figure 3)**. Of all tests, 19 tests passed the FDR<0.2 threshold for having a suggestive differential variant burden in AD cases and controls **(Table 2, Figure 3)**. These tests covered 11 genes (in order of significance): *SORL1*, *TREM2*, *ABCA7*, *ATP8B4*, *ADAM10*, *ABCA1*, *ORC6*, *CBX3*, *PRSS3*, *B3GNT4* and *SRC*. Of these, 6 tests (covering 4 genes) were significant when using a more conservative family-wise error rate correction for multiple testing (Holm-Bonferoni corrected p<0.05): *SORL1*, *TREM2*, *ABCA7*, and *ATP8B4*.

The predicted deleteriousness and the number of identified rare variants varied per gene. We aimed to accommodate for this variability by using different deleteriousness predictions thresholds. Tests using the LOF+REVEL≥25 threshold provided the most evidence for an association between variant-burden and AD risk (i.e. lowest p value) for the *TREM2*, *ABCA7*, *ATP8B4*, *ORC6*, *CBX3*, *PRSS3*, *B3GNT4* genes. Tests using the LOF+REVEL≥50 threshold provided the most evidence for *SORL1*, *ABCA1* and *SRC*, and testing using the LOF+REVEL≥75 threshold provided the most evidence for an association for the *ADAM10* gene **(Table 2, Figure 3).** The *SORL1*, *ABCA7*, *ATP8B4*, *ADAM10*, and *ABCA1* genes were identified using multiple thresholds (light grey gene names in **Figure 3**). Most genes were associated with an increased burden in cases, but at the FDR<0.2 significance level we identified *CBX3* and *PRSS3* which exhibited a lower burden of LOF+REVEL≥25 variants in cases than in controls, indicating potential protective association **(Table 2)**.

Dependence of effect sizes on variant deleteriousness category

Next, we investigated the effect on AD risk for variants from the four predicted variant deleteriousness categories. In our dataset all genes (except *CBX3)* included LOF variants. For 7 genes, we identified at least 3 carriers with LOF variants (*SORL1*, *TREM2*, *ABCA7*, *ATP8B4*, *ADAM10*, *ABCA1*, *ORC6*). For 6 of these 7 genes, we observed that the LOF variant category had a higher ordinal OR point-estimate than the (missense) variant categories (p=0.06, binomial test) (**Figure 4**). Finally, when tested whether variant impact was ordered according to predicted

deleteriousness: LOF ≥ REVEL 75-100 ≥ REVEL 50-75 ≥ REVEL 25-50 using a trend test (see methods), this test was significant (FDR<0.05) for *SORL1*, *ADAM10*, and *ABCA1.*

Relation between variant-burden and age at onset

Subsequently, we investigated the relationship between age and variant-burden by testing if variant-burden in AD patients decreased with the age at onset categories ≤65 (EOAD), 65-70, 70-80 and >80 (**Figure 5**). The median age at onset in the complete dataset was 73. For most of the identified genes, the burden of damaging variants was highest in younger cases, and decreased with increasing age at onset. The median age at onset of case carriers, was lowest in ORC6 (60y), followed by *ADAM10* (62y), *SRC* (64y), *B3GNT4* (66y), *SORL1* (67y), *ABCA1* (70y), *TREM2* (70y), *ABCA7* (70y) and was the highest in *ATP8B4* (72y). Notably, while the median age at onset of missense variants in SORL1 was 68, it was lower for LOF variant carriers (60). In the *ATP8B4*, *CBX3*, and *PRSS3* genes we observed no relationship between the variant burden and age at onset. Note that the variants in the latter two genes were associated with a protective effect, and therefore most carriers are controls.

Carrier or variant frequency

In line with the above, the fraction of variant carriers generally decreased with increasing age (**Figure 5**). However, a considerable fraction of older AD patients carries variants in the *SORL1*, *TREM2*, *ABCA7, ATP8B4 and ABCA1* genes, suggesting that variants in these genes also contribute to an increased risk of late-onset AD. Of note, there were only a few carriers of damaging variants in the *ADAM10*, *ORC6*, *B3GNT4* and *SRC* genes (respectively 13, 16, 29 and 27 carriers), such that impairment of these genes is likely to contribute to AD in only a few patients.

A relatively large fraction of variants from the most significant variant threshold per gene were singletons, i.e. variants that were carried by only a single individual in our dataset **(Figure 6a)**. There were 126 carriers of a singleton variant in SORL1 (43%), 9 in *ADAM10* (69%), 105 in *ABCA1* (48%), 14 in *ORC6* (88%), 17 in *B3GNT4* (59%) and 10 in *SRC* (37%). However, the AD-association of the *TREM2*, *ABCA7* and *ATP8B4* genes was carried by more common variants: singletons were identified in only 8 carriers (3%), 167 carriers (13%) and 45 carriers (6%). Finally, in the protective genes we also found relatively low numbers of singletons: 0 in *CBX3* (0%) as the association signal was driven by a single recurrent variant and 14 in *PRSS3* (13%), indicating that their protective signal was effectuated by more common (but still rare) variants. We further tested if the effect

size trended to be higher for the rarer variants: a significant trend (FDR<0.05) was observed for *SORL1* (p≤0.00004) and *ABCA1* (p≤0.00004), and a suggestive trend in *TREM2* (p=0.04) **(Figure 6).**

Age-matched analysis

To investigate whether the observed variant burden-effects were AD-specific, or whether they could also be explained by other age-related diseases, we performed a sensitivity analysis with strict age-matching. There was a strong agreement between the effect sizes when comparing age-matched case-control analysis and the case-control analysis unselected for age **(Figure S3)**. The age-matched analysis supported for each gene a role in AD, but based on the confidence intervals for the effect of the *SRC* gene, we cannot exclude the possibility that observed effects might also be attributable to a non-AD age-related disease. We observed a slight reduction in the effect size in the age-matched analysis, as observed for *SORL1* and *TREM2*. This was according to expectations, as mortality due to AD causes an additional age-related effect between young cases and old controls, which is removed by the age-matching.

APOE-ε4 sensitivity analysis

We did not correct our analysis for the common APOE genotype because this is not a confounder for the identification of a differential burden of rare variants between cases and controls. To investigate the validity of this assumption, we performed a sensitivity analysis in which we compared analysis corrected and uncorrected for carriership of the *APOE-ε4* allele, which did not change our results (**Figure S2**).

Gene specific analysis

For our genome wide burden analysis variant selection criteria and thresholds were uniformly applied to all variants in each gene. Therefore, it was necessary to refine burden effects by correcting for variants with divergent effects compared to the variants in the burden (see Methods and **Table S3**). Gene-specific analyses are described for each gene in the Supplementary Material. This led to a refinement of the associations of *SORL1*, *TREM2*, *ABCA7*, and *ABCA1* (**Table 1, Figure 7**).

Carriers of multiple variants

We finally measured the presence of multiple damaging variants in carriers. Of the cases, 1,963/12,652 cases (15.5%) carried at least one damaging variant in at least one gene. Of these, 101 cases carried damaging variants in two genes, and 1 case carried damaging variants in three genes. This was slightly lower than expected under a model in which damaging variants were randomly distributed across the cases (114.3 double and 3.4 triple carriers expected, ratio=0.86, p=0.082). In particular, we observed that there were significantly less carriers of damaging *ATP8B4* variants that also carried a damaging variant in another gene (41 observed, 62.2 expected, ratio=0.66, p=0.0028). Of the individuals who carried damaging variants in multiple genes, 48.0% were classified as EOAD, compared to 36.9% of the cases that carried only a single damaging variant (p=0.027, fisher-exact test).

Discussion

In our WES study we identified four genes in which carrying a rare deleterious variant associated with AD at exome-wide significance. Of these, we identified rare predicted damaging variants in the *ATP8B4* gene as a novel AD risk factor, the other three genes were previously established AD risk factors, i.e. *SORL1*, *TREM2* and *ABCA7(7, 9, 20, 21).* Additionally, we identified seven genes with suggestive evidence for an association with AD risk. Of these, the *ADAM10* and *ABCA1* genes were previously identified to be associated with AD-related mechanisms (22, 23), while for rare variants in the *ORC6*, *CBX3*, *PRSS3*, *B3GNT4*, and *SRC* genes we provide a first report for a suggestive association with AD risk. Almost all genes showed an increased variant burden in the younger cases, with the exception of the variants in *CBX3* and *PRSS3*, which were associated with a protective effect. For several genes we observed trends that the rarest variants associated with the highest effect sizes. Also, a large fraction of the signal in *SORL1*, *ADAM10*, *ABCA1*, *ORC6*, *B3GNT4* and *SRC* came from singleton variants, while in *TREM2*, *ABCA7*, *ATP8B4 CBX3*, and *PRSS3* the majority of the signal was carried by more common (but still rare) variants. Common missense variants (MAF > 1%), which occur in *TREM2*, *SORL1* and *ABCA7*, had relatively small (or protective) effects compared to the effect size observed in the burden test. Investigation of gene-functions indicated that most identified genes were associated with aspects of the Alzheimer Disease pathophysiology.

Impaired ***SORL1*** function (Sortilin Related Receptor 1) has been associated with increased Aβ production due to a disruption of APP processing (24, 25) and a decrease in the degradation of

intracellular nascent Aβ peptides by lysosomes (26). In the present dataset, we identified a total of 168 damaging variants in the *SORL1* gene, carried by 291 individuals. The association with AD is mainly driven by variants which are individually extremely rare and mostly singletons. The burden of predicted damaging *SORL1* variants was highest in EOAD cases and decreased with increasing AAO (9, 16, 27). We observed a relationship between the predicted variant deleteriousness level and the effect on AD risk: LOF variants associated with a 36-fold increased risk of EOAD and 7-fold increased risk of LOAD, while missense variants associated with a 2.7 and 1.9-fold increase risk of EOAD and LOAD, respectively.

TREM2 (Triggering Receptor Expressed On Myeloid Cells 2) is involved in microglia-dependent pathophysiological processes in AD through Aβ phagocytosis and clearance and/or compaction in amyloid plaques (28, 29). In our dataset, we identified 17 damaging *TREM2* variants carried by 291 individuals. Although damaging *TREM2* variants are rare, most variants were observed in several individuals, which is different from what is observed in, for example, *SORL1*. We found a clear relation with predicted variant deleteriousness and the association with AD: *TREM2* LOF variants after refinement associated with a 10.8-fold increased risk of AD, while missense variants associated with a 3.5-fold increased AD risk.

One of the functions of ***ABCA7*** (ATP Binding Cassette Subfamily A Member 7) is to clear the blood brain barrier from Aβ (30). Impaired ABCA7 protein function was also associated with a faster APP endocytosis, an increased *in vitro* Aβ production, and an accelerated amyloid pathology accumulation in young transgenic mice (31-33). In our dataset, we found an AD-association of damaging variants in the *ABCA7* gene based on 272 variants carried by 1,267 individuals. As many as ~7.5% of all AD cases with an AAO<70 years and 5% of all controls carried such an *ABCA7* variant. The association with AD is driven by damaging variants with different features: some are individually extremely rare or singletons, while others occur in several individuals. Both LOF and missense variants in the *ABCA7* gene were associated with a ~1.4-1.8-fold increased AD risk, but the burden of damaging variants concentrated in younger AD patients.

We identified a new signal in the ***ATP8B4*** gene (ATPase Phospholipid Transporting 8B4) which encodes a member of the cation transport ATPase which is involved in phospholipid transport at the cell membrane. *ATP8B4* is expressed in macrophages/microglia in the brain and rare variants in this gene have been associated with the risk of developing systemic sclerosis, an autoimmune disease (34). Approximately 4% of the AD cases and 2.5% of the controls carried a rare, predicted deleterious variant in *ATP8B4*. The burden reaches exome wide significance based on 74 variants

carried by 767 individuals. The association with AD was mainly driven by 3 missense variants (G395S, C874R, and H987R), while the burden of highly rare variants (allele count < 5) did not associate with AD. In contrast to *SORL1*, *TREM2* and *ABCA7*, the variant burden was not associated with AAO. A common variant in the *ATP8B4* locus (rs6493386) was previously associated with both AD risk and LDL (35, 36)**.** A signal in the proximity of the *ATP8B4* locus was reported in a large GWAS meta-analysis, which was tagged to the neighboring *SSP2L* gene (4). It cannot be excluded that the *SSP2L* association with AD might be driven by *ATP8B4* rather than by *SSP2L*. Our observations highlight potential implication of *ATP8B4* in inflammation and may provide additional support for the importance of microglia/inflammation in the AD pathophysiology.

α-secretase ***ADAM10*** (a disintegrin and metalloproteinase domain-containing protein 10) plays a major role in APP metabolism (37). In our analysis, we identified only 11 damaging *ADAM10* variants in 12 carriers. With the rare occurrence of such variants only a very strong association with AD will enable the detection of an exome-wide significant signal, even in the current large sample. Indeed, we found that damaging LOF variants and missense variants were suggestively associated with a 15-fold and 6-fold increased AD-risk, respectively. In addition, similar to the association signals identified in *SORL1* and *ABCA7* genes, these LOF and high-impact missense variants showed suggestive association with an increased risk of EOAD. Notably, LOF variants in *ADAM10* were previously reported to be associated with an autosomal dominant inheritance of abnormal pigmentation of the skin (38), such that skin pigmentation might represent a clinical proxy for carrying a rare LOF variant in the *ADAM10* gene. We could not retrospectively investigate skin pigmentation in our cohort. Common variants in *ADAM10* were recently associated with AD risk in a GWAS meta-analysis (REF), which aligns with the independent AD-associations with common variants and rare variant-burden also observed for *SORL1*, *ABCA7,* and, most likely, *ATP8B4* . Previous reports identified the Q170H and the R181G variants in *ADAM10* in LOAD families (39). While we did detect these variants in our sample, the single variant analysis indicated that these were not significantly associated with AD.

The role of the ***ABCA1*** transporter (ATP Binding Cassette Subfamily A Member 1) gene, has been assessed extensively (40). ABCA1 protein lipidates APOE in the CNS (41), and poor ABCA1-dependent lipidation of APOE-containing lipoprotein particles may increase Aβ deposition and fibrillogenesis (42). Indeed, mice overexpressing *ABCA1* in an AD-like mouse model had significantly less Aβ deposition (41). A rare deleterious missense variant (A937V) was previously proposed to be implicated in a LOAD family (43) **a**nd another rare deleterious missense variant (N1800H) was previously associated with AD risk (44)**.** Based on 142 variants carried by 216

individuals, we found that the burden of rare variants in the *ABCA1* gene was suggestively associated with increased risk of AD. This variant burden did not include the A937V and N1800H variants, which were previously associated with AD (43, 44), respectively due to differential missingness and a low REVEL score. We were able to manually include the N1800H variant in a post hoc analysis, which improved the association of ABCA1 from p=2.4e-5 to p=4.5e-7, crossing the conservative Bonferroni threshold. Damaging variants in *ABCA1* associated with AD with a pattern similar as *SORL1:* early onset cases carried the highest fraction of predicted deleterious variants and a higher level of variant deleteriousness associated with a higher AD risk. While LOF variants in *ABCA1* were suggestively associated with a relatively modest >4-fold increased early onset AD risk (i.e. compared to damaging variants in *SORL1* or *TREM2*), the large number of damaging *ABCA1*-variants in our sample enabled the detection of the suggestive association.

The protein encoded by ***ORC6*** (Origin Recognition Complex Subunit 6) is part of a highly conserved six subunit protein complex essential for the initiation of the DNA replication in eukaryotic cells (45). It is expressed at a low level in neurons (46). We identified 15 rare damaging mutations in 16 individuals (14 of whom were cases), which were suggestively associated with a strong >9-fold increased risk for having early onset of AD, in a pattern resembling the AD-association of damaging *SORL1* variants. When this association replicates, further functional investigation is necessary to explain the involvement of the ORC6 protein in AD pathophysiology.

The protein encoded by the ***B3GNT4*** gene is a member of the beta-1,3-N-acetylglucosaminyltransferase protein family. *B3GNT4* was associated with serum urate and triglyceride concentration in GWAS (47, 48) which were both associated with increased risk for dementia and AD. While the protein is highly expressed in the brain (49), its function in the brain is not well explored. We identified 22 rare damaging mutations in 29 individuals, and the burden of damaging variants was highest in the early onset cases as evidenced by a suggestive >12-fold increased risk for early onset AD. The few variants identified included only one LOF variant, such that the number of variants was too low to infer a relation with variant-damagingness.

The protein encoded by ***SRC*** (Proto-Oncogene, Non-Receptor Tyrosine Kinase) is a non-receptor protein tyrosine kinase that belongs to the same family as Pyk2, an AD genetic risk factor, and Fyn. Moreover, SRC is known to bind Pyk2, which is critical for Pyk2 activity (50) SRC is activated by many different classes of cellular receptors including immune response receptors, integrins and other adhesion receptors (51) The suggestive AD-risk increasing signal in *SRC*-variants was based on 15 damaging variants carried by 27 individuals, and the strongest association was found

in early onset cases (OR=6.6). SRC has been described to potentially modulate APP trafficking/metabolism (52), but also Tau phosphorylation (53).

We identified a single variant in the ***CBX3*** gene (Chromobox 3) that suggestively associated with a *decreased* AD risk, with an odds ratio of 0.2. The variant was carried by 30 individuals, mostly controls and several EOAD cases. The protein encoded by *CBX3* binds DNA and is a component of heterochromatin (54). it is ubiquitously present and, in the brain, mainly expressed in neurons (46). Little is known about CBX3 functions in the brain and this protein has been described to maintain lineage specificity during neural differentiation (55), as well as promoting glioma cell proliferation (56). The *CBX3* variant was previously identified to have a suggestive signal in an AD WES sequencing analysis (which included overlapping samples with this study)(12).

Last, we identified a suggestive association between variants in the ***PRSS3*** (Serine Protease 3) gene and two-fold *decreased* risk for AD (OR=0.5). We identified 21 variants in this gene carried by 111 individuals, of which 14 were singletons. This indicates that the majority of this protective signal was effectuated by more common (but still rare) variants. *PRSS3* encodes a serine protease of the trypsin family which is mainly expressed in pancreas and in the neurons of the brain (46). The Kunitz inhibitor domain in APP has been reported to be a highly specific substrate of the *PRSS3* protease (57), but the protective effect of these variants needs to be replicated and further explored in future studies.

This comparison of between exomes from AD cases and controls represents one of the largest performed thus far, which allows the detection of differential burden of damaging variants in genes that were not yet associated with AD. Across all genes, a large part of the signal depended on singletons, indicating that high level of accuracy is warranted. We applied several approaches to maximize the statistical power and the accuracy of the discovery study. (i). We collected and merged raw WES data on one server which allowed us to uniformly apply a quality control pipeline. (ii) We designed custom algorithms that detected and removed the prevalent batch effects across all data simultaneously, which were highly prevalent due to the use of different WES kits and sequencing laboratories. (iii). We confirmed that the variants were not somatic by checking allele balance, indicating that the protective signal in *PRSS3* and *CBX3* was not a consequence of age-related clonal hematopoiesis (ARCH) in our controls (58), who were on average older than our cases. (iv). We were able to accommodate differential variant effects by performing burden analyses across four different levels of predicted variant deleteriousness. (v).

We took into consideration that cases with a higher age at onset may have a lower burden of damaging variants.
Further, we performed several complementary analyses to explore additional potential biases. (vi). In an age-matched analysis we investigated whether burden associations with AD could also be due to a confounding factor such as age. This analysis supported a role in AD for all the eleven genes. (vii) A sensitivity analysis in which we compared our results when corrected and uncorrected for *APOE-ε4* indicated that the observed associations between variant burden and AD risk are independent of *APOE* genotype. We could not explore possible synergistic or additive effects between carrying a damaging genetic variant in one of the identified genes and *APOE* genotype, because part of our sample was selected according to *APOE* genotype, which complicates such an analysis. Moreover, stratification by *APOE* genotype would reduce statistical power.

In conclusion, our study provides further evidence for a pivotal role of APP processing, lipid metabolism, and microglia and neuroinflammatory processes in AD pathophysiology (59-61). Of the genes identified here, five belong to the Aβ network, either through Aβ production (APP processing) or through increased aggregation / decreased clearance. More specifically, the suggestive association of rare variants in *ADAM10* with increased AD risk is in line with the important role of APP processing on top of the contribution of *APP*, *PSEN1*, *PSEN2*, *SORL1* and *ABCA7*. Furthermore, next to the known AD-associations of variants in *APOE*, *PLCG2*, *ABI3*, *ABCA7* and *TREM2*, we find a suggestive association of rare variants in *ABCA1* with AD risk, providing a novel genetic determinant with a role in Aβ aggregation and clearance. Moreover, with the identification of *ATP8B4* as a novel AD genetic risk factor, further strengthening the evidence for the involvement of microglia and neuroinflammation in AD. We acknowledge that the novel genetic associations we identified will require further investigation and replication in independent samples before they can be accepted as genuine AD genetic determinants. Notably, with this sample we were able to assess 13,299 genes of the total 19,822 autosomal protein-coding genes and not all types of genetic variation. A larger sample size and the use of whole genome sequencing will allow the investigation of even more genes, which will require continued efforts in combining and jointly analyzing samples.

ACKNOWLEDGMENTS:
The authors are grateful to the study participants, their family members, and the participating general practitioners, pharmacists and all laboratory personnel involved in blood collection, DNA

isolation, and DNA biobanking. The work in this manuscript was carried out on the Cartesius supercomputer, which is embedded in the Dutch national e-infrastructure with the support of SURF Cooperative. Computing hours were granted in 2016, 2017, 2018 and 2019 to H. Holstege by the Dutch Research Council (project name: '100plus'; project numbers 15318 and 17232). See the Supplemental Materials for acknowledgements for all contributing studies.

REFERENCES

1. Bintener C, Miller O. Estimating the prevalence of dementia in Europe. Alzheimer Europe. 2020.
2. Gatz M, Reynolds CA, Fratiglioni L, Johansson B, Mortimer JA, Berg S, et al. Role of genes and environments for explaining Alzheimer disease. Archives of general psychiatry. 2006;63(2):168-74.
3. Lambert J, Ibrahim-Verbaas C, Harold D, Naj A, Sims R, Bellenguez C, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nature genetics. 2013;45(12):1452-8.
4. Kunkle BW, Grenier-Boley B, Sims R, Bis JC, Damotte V, Naj AC, et al. Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing. Nature genetics. 2019;51(3):414-30.
5. Pottier C, Hannequin D, Coutant S, Rovelet-Lecrux A, Wallon D, Rousseau S, et al. High frequency of potentially pathogenic SORL1 mutations in autosomal dominant early-onset Alzheimer disease. Molecular psychiatry. 2012;17(9):875-9.
6. Cuyvers E, De Roeck A, Van den Bossche T, Van Cauwenberghe C, Bettens K, Vermeulen S, et al. Mutations in ABCA7 in a Belgian cohort of Alzheimer's disease patients: a targeted resequencing study. Lancet neurology. 2015;14(8):814-22.
7. Jonsson T, Stefansson H, Steinberg S, Jonsdottir I, Jonsson PV, Snaedal J, et al. Variant of TREM2 associated with the risk of Alzheimer's disease. The New England journal of medicine. 2013;368(2):107-16.
8. Guerreiro R, Wojtas A, Bras J, Carrasquillo M, Rogaeva E, Majounie E, et al. TREM2 variants in Alzheimer's disease. The New England journal of medicine. 2013;368(2):117-27.
9. Holstege H, van der Lee SJ, Hulsman M, Wong TH, van Rooij JG, Weiss M, et al. Characterization of pathogenic SORL1 genetic variants for association with Alzheimer's disease: a clinical interpretation strategy. European journal of human genetics : EJHG. 2017;25(8):973-81.
10. Corder EH, Saunders AM, Strittmatter WJ, Schmechel DE, Gaskell PC, Small GW, et al. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. Science. 1993;261(5123):921-3.
11. Lo M-T, Kauppi K, Fan C-C, Sanyal N, Reas ET, Sundar VS, et al. Identification of genetic heterogeneity of Alzheimer's disease across age. Neurobiology of Aging. 2019;84:243.e1-.e9.
12. Bis JC, Jian X, Kunkle BW, Chen Y, Hamilton-Nelson KL, Bush WS, et al. Whole exome sequencing study identifies novel rare and common Alzheimer's-Associated variants involved in immune response and transcriptional regulation. Molecular psychiatry. 2018.
13. Ioannidis NM, Rothstein JH, Pejaver V, Middha S, McDonnell SK, Baheti S, et al. REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. The American Journal of Human Genetics. 2016;99(4):877-85.

14. Liu X, Wu C, Li C, Boerwinkle E. dbNSFP v3.0: A One-Stop Database of Functional Predictions and Annotations for Human Nonsynonymous and Splice-Site SNVs. Human mutation. 2016;37(3):235-41.
15. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature. 2020;581(7809):434-43.
16. Bellenguez C, Charbonnier C, Grenier-Boley B, Quenez O, Le Guennec K, Nicolas G, et al. Contribution to Alzheimer's disease risk of rare variants in TREM2, SORL1, and ABCA7 in 1779 cases and 1273 controls. Neurobiol Aging. 2017;59:220 e1- e9.
17. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B (Methodological). 1995;57(1):289-300.
18. Holm S. A Simple Sequentially Rejective Multiple Test Procedure. Scandinavian Journal of Statistics. 1979;6(2):65-70.
19. Posner MA, Ash AS. Comparing weighting methods in propensity score analysis. Unpublished working paper, Columbia University. 2012;http://www.stat.columbia.edu/~gelman/stuff_for_blog/posner.pdf.
20. Guerreiro R, Escott-Price V, Darwent L, Parkkinen L, Ansorge O, Hernandez DG, et al. Genome-wide analysis of genetic correlation in dementia with Lewy bodies, Parkinson's and Alzheimer's diseases. Neurobiology of aging. 2016;38:214. e7-. e10.
21. Steinberg S, Stefansson H, Jonsson T, Johannsdottir H, Ingason A, Helgason H, et al. Loss-of-function variants in ABCA7 confer risk of Alzheimer's disease. Nature genetics. 2015;47(5):445-7.
22. Yang P, Baker KA, Hagg T. The ADAMs family: Coordinators of nervous system development, plasticity and repair. Progress in Neurobiology. 2006;79(2):73-94.
23. Koldamova R, Fitz NF, Lefterov I. The role of ATP-binding cassette transporter A1 in Alzheimer's disease and neurodegeneration. Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids. 2010;1801(8):824-30.
24. Rogaeva E, Meng Y, Lee JH, Gu Y, Kawarai T, Zou F, et al. The neuronal sortilin-related receptor SORL1 is genetically associated with Alzheimer disease. Nature genetics. 2007;39(2):168-77.
25. Knupp A, Mishra S, Martinez R, Braggin JE, Szabo M, Kinoshita C, et al. Depletion of the AD Risk Gene SORL1 Selectively Impairs Neuronal Endosomal Traffic Independent of Amyloidogenic APP Processing. Cell Reports. 2020;31(9).
26. Caglayan S, Takagi-Niidome S, Liao F, Carlo AS, Schmidt V, Burgert T, et al. Lysosomal sorting of amyloid-beta by the SORLA receptor is impaired by a familial Alzheimer's disease mutation. Science translational medicine. 2014;6(223):223ra20.
27. Campion D, Charbonnier C, Nicolas G. SORL1 genetic variants and Alzheimer disease risk: a literature review and meta-analysis of sequencing data. Acta neuropathologica. 2019.
28. Jay TR, Hirsch AM, Broihier ML, Miller CM, Neilson LE, Ransohoff RM, et al. Disease Progression-Dependent Effects of TREM2 Deficiency in a Mouse Model of Alzheimer's Disease. The Journal of Neuroscience. 2017;37(3):637-47.
29. Colonna M, Holtzman DM, Cirrito JR, DeMattos RB, Grutzendler J, Cella M, et al. TREM2-mediated early microglial response limits diffusion and toxicity of amyloid plaques. Journal of Experimental Medicine. 2016;213(5):667-75.
30. Lamartinière Y, Boucau M-C, Dehouck L, Krohn M, Pahnke J, Candela P, et al. ABCA7 Downregulation Modifies Cellular Cholesterol Homeostasis and Decreases Amyloid-β Peptide Efflux in an in vitro Model of the Blood-Brain Barrier. Journal of Alzheimer's Disease. 2018;64(4):1195-211.

31. Sakae N, Liu C-C, Shinohara M, Frisch-Daiello J, Ma L, Yamazaki Y, et al. ABCA7 Deficiency Accelerates Amyloid-β Generation and Alzheimer's Neuronal Pathology. The Journal of Neuroscience. 2016;36(13):3848-59.
32. Satoh K, Abe-Dohmae S, Yokoyama S, St George-Hyslop P, Fraser PE. ATP-binding Cassette Transporter A7 (ABCA7) Loss of Function Alters Alzheimer Amyloid Processing. Journal of Biological Chemistry. 2015;290(40):24152-65.
33. Kim WS, Li H, Ruberu K, Chan S, Elliott DA, Low JK, et al. Deletion of Abca7 Increases Cerebral Amyloid- Accumulation in the J20 Mouse Model of Alzheimer's Disease. Journal of Neuroscience. 2013;33(10):4387-94.
34. Gao L, Emond MJ, Louie T, Cheadle C, Berger AE, Rafaels N, et al. Identification of Rare Variants inATP8B4as a Risk Factor for Systemic Sclerosis by Whole-Exome Sequencing. Arthritis & Rheumatology. 2016;68(1):191-200.
35. Broce IJ, Tan CH, Fan CC, Witoelar A, Wen N, Jansen I, et al. 2018.
36. Li H, Wetten S, Li L, St. Jean PL, Upmanyu R, Surh L, et al. Candidate Single-Nucleotide Polymorphisms From a Genomewide Association Study of Alzheimer Disease. Archives of neurology. 2008;65(1).
37. Saftig P, Lichtenthaler SF. The alpha secretase ADAM10: A metalloprotease with multiple functions in the brain. Progress in Neurobiology. 2015;135:1-20.
38. Kono M, Sugiura K, Suganuma M, Hayashi M, Takama H, Suzuki T, et al. Whole-exome sequencing identifies ADAM10 mutations as a cause of reticulate acropigmentation of Kitamura, a clinical entity distinct from Dowling-Degos disease. Human molecular genetics. 2013;22(17):3524-33.
39. Kim M, Suh J, Romano D, Truong MH, Mullin K, Hooli B, et al. Potential late-onset Alzheimer's disease-associated mutations in the ADAM10 gene attenuate α-secretase activity. Human molecular genetics. 2009;18(20):3987-96.
40. Koldamova R, Fitz NF, Lefterov I. ATP-binding cassette transporter A1: From metabolism to neurodegeneration. Neurobiology of disease. 2014;72:13-21.
41. Wahrle SE, Jiang H, Parsadanian M, Kim J, Li A, Knoten A, et al. Overexpression of ABCA1 reduces amyloid deposition in the PDAPP mouse model of Alzheimer disease. Journal of Clinical Investigation. 2008.
42. Koldamova R, Staufenbiel M, Lefterov I. Lack of ABCA1 Considerably Decreases Brain ApoE Level and Increases Amyloid Deposition in APP23 Mice. Journal of Biological Chemistry. 2005;280(52):43224-35.
43. Beecham GW, Vardarajan B, Blue E, Bush W, Jaworski J, Barral S, et al. Rare genetic variation implicated in non-Hispanic white families with Alzheimer disease. Neurology Genetics. 2018;4(6).
44. Nordestgaard LT, Tybjaerg-Hansen A, Nordestgaard BG, Frikke-Schmidt R. Loss-of-function mutation in ABCA1
and risk of Alzheimer's disease and cerebrovascular disease. Alzheimer's & Dementia. 2015;11(12):1430-8.
45. Duncker BP, Chesnokov IN, McConkey BJ. The origin recognition complex protein family. Genome biology. 2009;10(3).
46. Aguet F, Barbeira AN, Bonazzola R, Brown A, Castel SE, Jo B, et al. 2019.
47. Bentley AR, Sung YJ, Brown MR, Winkler TW, Kraja AT, Ntalla I, et al. Multi-ancestry genome-wide gene–smoking interaction study of 387,272 individuals identifies new loci associated with serum lipids. Nature genetics. 2019;51(4):636-48.
48. Köttgen A, Albrecht E, Teumer A, Vitart V, Krumsiek J, Hundertmark C, et al. Genome-wide association analyses identify 18 new loci associated with serum urate concentrations. Nature genetics. 2012;45(2):145-54.

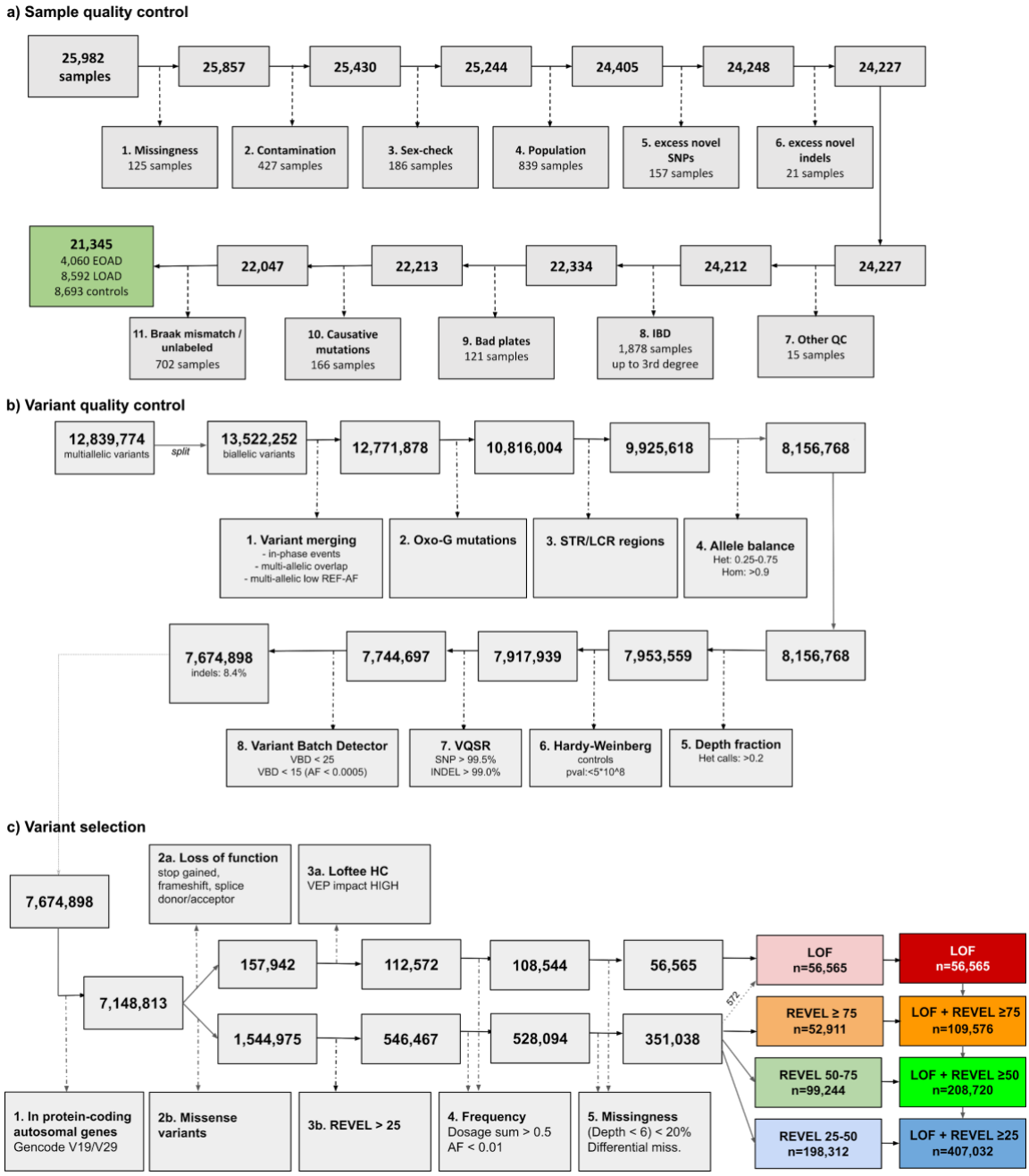
49. Nägga K, Gustavsson A-M, Stomrud E, Lindqvist D, van Westen D, Blennow K, et al. Increased midlife triglycerides predict brain β-amyloid and tau pathology 20 years later. Neurology. 2018;90(1):e73-e81.
50. Bruzzaniti A, Neff L, Sandoval A, Du L, Horne WC, Baron R. Dynamin Reduces Pyk2 Y402 Phosphorylation and Src Binding in Osteoclasts. Molecular and Cellular Biology. 2009;29(13):3644-56.
51. Roskoski R. Src protein–tyrosine kinase structure and regulation. Biochemical and biophysical research communications. 2004;324(4):1155-64.
52. Chaufty J, Sullivan SE, Ho A. Intracellular Amyloid Precursor Protein Sorting and Amyloid- Secretion Are Regulated by Src-Mediated Phosphorylation of Mint2. Journal of Neuroscience. 2012;32(28):9613-25.
53. Lee G. Tau and src family tyrosine kinases. Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease. 2005;1739(2-3):323-30.
54. Aydin E, Kloos D-P, Gay E, Jonker W, Hu L, Bullwinkel J, et al. A hypomorphic Cbx3 allele causes prenatal growth restriction and perinatal energy homeostasis defects. Journal of Biosciences. 2015;40(2):325-38.
55. Huang C, Su T, Xue Y, Cheng C, Lay FD, McKee RA, et al. Cbx3 maintains lineage specificity during neural differentiation. Genes & Development. 2017;31(3):241-6.
56. Zhao S-P, Wang F, Yang M, Wang X-Y, Jin C-L, Ji Q-K, et al. CBX3 promotes glioma U87 cell proliferation and predicts an unfavorable prognosis. Journal of neuro-oncology. 2019;145(1):35-48.
57. Salameh MdA, Robinson JL, Navaneetham D, Sinha D, Madden BJ, Walsh PN, et al. The Amyloid Precursor Protein/Protease Nexin 2 Kunitz Inhibitor Domain Is a Highly Specific Substrate of Mesotrypsin. Journal of Biological Chemistry. 2010;285(3):1939-49.
58. Shlush LI. Age-related clonal hematopoiesis. Blood. 2018;131(5):496-504.
59. Campion D, Pottier C, Nicolas G, Le Guennec K, Rovelet-Lecrux A. Alzheimer disease: modeling an Aβ-centered biological network. Molecular psychiatry. 2016;21(7):861-71.
60. Hardy J, Bogdanovic N, Winblad B, Portelius E, Andreasen N, Cedazo-Minguez A, et al. Pathways to Alzheimer's disease. Journal of internal medicine. 2014;275(3):296-303.
61. Webers A, Heneka MT, Gleeson PA. The role of innate immune responses and neuroinflammation in amyloid accumulation and progression of Alzheimer's disease. Immunology & Cell Biology. 2019;98(1):28-41.
62. Costello M, Pugh TJ, Fennell TJ, Stewart C, Lichtenstein L, Meldrim JC, et al. Discovery and characterization of artifactual mutations in deep coverage targeted capture sequencing data due to oxidative DNA damage during sample preparation. Nucleic acids research. 2013;41(6):e67-e.
63. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, et al. The Ensembl Variant Effect Predictor. Genome biology. 2016;17(1):122.

Table 1

gene	deleteriousness threshold	p-value	FDR	#variants / #carriers	carrier frequency	odds ratio (95% CI)			age at onset
					EOAD / LOAD / control	case / control	EOAD / control	LOAD / control	median (IQR)
SORL1	LOF+REVEL≥50	1.80E-18	<<0.01%	168 / 291	2.66% / 1.45% / 0.67%	2.6 (2.1-3.3)	3.6 (2.7-4.9)	2.1 (1.5-2.8)	67 (60-74)
	LOF	*9.00E-16*		*38 / 49*	*0.81% / 0.16% / 0.02%*	*16.4 (9.0-29.8)*	*36.1 (10.8-inf)*	*7.2 (2.0-50.9)*	*60 (56-68)*
	REVEL 50-100	*4.80E-10*		*130 / 245*	*1.92% / 1.29% / 0.64%*	*2.2 (1.7-2.8)*	*2.7 (2.0-3.8)*	*1.9 (1.4-2.6)*	*68 (60-75)*
	REVEL 50-100 [refined]	*6.20E-12*		*129 / 261*	*2.02% / 1.44% / 0.63%*	*2.5 (1.9-3.2)*	*3.0 (2.1-4.1)*	*2.2 (1.6-3.0)*	*68 (60-75)*
TREM2	LOF+REVEL≥25	2.80E-16	<<0.01%	17 / 291	2.12% / 1.83% / 0.55%	3.6 (2.8-4.6)	4.2 (2.9-6.0)	3.4 (2.4-4.7)	70 (63-76)
	LOF	*7.60E-03*		*9 / 39*	*0.25% / 0.26% / 0.08%*	*3.3 (1.7-6.5)*	*3.4 (1.3-9.0)*	*3.3 (1.4-7.7)*	*72 (63-76)*
	LOF [refined]	*4.70E-03*		*8 / 21*	*0.20% / 0.14% / 0.01%*	*10.8 (4.4-26.9)*	*14.2 (3.3-460.5)*	*9.4 (2.6-320.4)*	*70 (63-75)*
	REVEL 25-100	*8.90E-15*		*8 / 253*	*1.87% / 1.58% / 0.47%*	*3.7 (2.8-4.8)*	*4.3 (2.9-6.4)*	*3.4 (2.4-4.9)*	*69 (63-76)*
	REVEL 25-100 [refined]	*9.00E-20*		*10 / 336*	*2.56% / 2.04% / 0.66%*	*3.5 (2.8-4.4)*	*4.4 (3.1-6.1)*	*3.2 (2.3-4.3)*	*69 (63-76)*
ABCA7	LOF+REVEL≥25	8.80E-08	0.06%	272 / 1267	7.41% / 6.15% / 5.04%	1.3 (1.2-1.5)	1.5 (1.3-1.7)	1.3 (1.1-1.4)	70 (62-78)
	LOF	*1.50E-03*		*47 / 107*	*0.81% / 0.54% / 0.32%*	*1.8 (1.2-2.6)*	*2.2 (1.4-3.7)*	*1.5 (1.0-2.4)*	*69 (60-74)*
	REVEL 25-100	*4.20E-06*		*225 / 1162*	*6.60% / 5.62% / 4.73%*	*1.3 (1.2-1.5)*	*1.4 (1.2-1.7)*	*1.2 (1.1-1.4)*	*70 (62-79)*
	REVEL 25-100 [refined]	*4.10E-08*		*223 / 983*	*5.91% / 4.91% / 3.69%*	*1.4 (1.3-1.6)*	*1.6 (1.4-1.9)*	*1.3 (1.2-1.6)*	*70 (62-78)*
ATP8B4	LOF+REVEL≥25	4.60E-07	0.24%	74 / 767	4.43% / 4.12% / 2.68%	1.5 (1.3-1.8)	1.6 (1.3-1.9)	1.5 (1.3-1.8)	72 (62-79)
	LOF	*2.10E-01*		*13 / 34*	*0.25% / 0.16% / 0.12%*	*1.5 (0.7-3.1)*	*1.8 (0.7-4.4)*	*1.4 (0.6-3.1)*	*73 (59-78)*
	REVEL 25-100	*1.10E-06*		*61 / 733*	*4.19% / 3.96% / 2.57%*	*1.5 (1.3-1.8)*	*1.6 (1.3-1.9)*	*1.5 (1.3-1.8)*	*72 (63-79)*
ADAM10	LOF+REVEL≥75	2.70E-06	1%	11 / 12	0.25% / 0.01% / 0.01%	7.3 (1.3-46.0)	19.8 (4.3-inf)	1.1 (0.0-32.2)	62 (59-64)
	LOF	*2.40E-04*		*9 / 9*	*0.17% / 0.01% / 0.01%*	*5.4 (1.6-17.9)*	*13.4 (2.9-inf)*	*1.1 (0.0-28.7)*	*63 (59-64)*
	REVEL 75-100	*0.0016*		*2 / 3*	*0.07% / 0.00% / 0.00%*	--	--	--	--
ABCA1	LOF+REVEL≥50	2.50E-05	6.5%	142 / 216	1.55% / 1.05% / 0.72%	1.7 (1.3-2.3)	2.3 (1.6-3.2)	1.5 (1.1-2.1)	70 (59-76)
	LOF	*5.70E-03*		*21 / 31*	*0.22% / 0.15% / 0.10%*	*3.2 (1.5-6.8)*	*4.2 (1.5-12.0)*	*2.7 (1.0-7.3)*	*70 (59-77)*
	LOF [refined]	*2.50E-03*		*20 / 24*	*0.22% / 0.14% / 0.03%*	*4.9 (2.1-11.4)*	*6.9 (1.8-25.9)*	*4.0 (1.1-14.4)*	*68 (59-77)*
	REVEL 50-100	*6.20E-04*		*121 / 185*	*1.33% / 0.90% / 0.62%*	*1.6 (1.2-2.2)*	*2.0 (1.4-3.0)*	*1.4 (1.0-2.0)*	*69 (59-76)*
	REVEL 50--100 [refined]	*1.20E-06*		*122 / 230*	*1.70% / 1.23% / 0.63%*	*2.1 (1.6-2.7)*	*2.5 (1.7-3.5)*	*1.9 (1.3-2.6)*	*68 (58-76)*
ORC6	LOF+REVEL≥25	5.60E-05	12%	15 / 16	0.27% / 0.03% / 0.02%	4.1 (1.3-24.7)	9.4 (3.1-84.2)	1.3 (0.2-12.9)	60 (59-65)
	LOF	*5.10E-02*		*4 / 4*	*0.07% / 0.00% / 0.01%*	--	--	--	--
	REVEL 25-100	*0.00042*		*11 / 12*	*0.20% / 0.03% / 0.01%*	*6.4 (1.9-21.3)*	*13.3 (3.1-inf)*	*2.7 (0.4-82.7)*	*61 (59-67)*
CBX3	LOF+REVEL≥25	6.00E-05	12%	1 / 30	0.12% / 0.02% / 0.26%	0.2 (0.1-0.3)	0.3 (0.1-0.9)	0.1 (0.0-0.3)	--
PRSS3	LOF+REVEL≥25	7.60E-05	14%	21 / 111	0.27% / 0.43% / 0.72%	0.5 (0.3-0.7)	0.3 (0.2-0.7)	0.6 (0.4-0.9)	--
B3GNT4	LOF+REVEL≥25	9.50E-05	16%	22 / 29	0.32% / 0.16% / 0.02%	8.1 (2.4-32.1)	12.6 (4.0-97.8)	6.0 (2.1-53.3)	66 (60-74)
SRC	LOF+REVEL≥50	1.10E-04	18%	15 / 27	0.32% / 0.10% / 0.06%	3.3 (1.5-7.4)	6.6 (2.3-18.8)	1.9 (0.6-5.8)	64 (58-73)

Results from the discovery analysis. Per gene, results are shown for the most significant deleteriousness threshold, and separately for LOF variants and missense variants (except for *CBX3, PRSS3, B3GNT4, SRC* which have ≤1 LOF variant carrier). A carrier is an individual with at least one or more minor alleles. Carrier frequency is the percentage of people that carry one or more variants. Tests were performed at the gene level, putatively gathering several transcripts of a same gene.

Figure 1



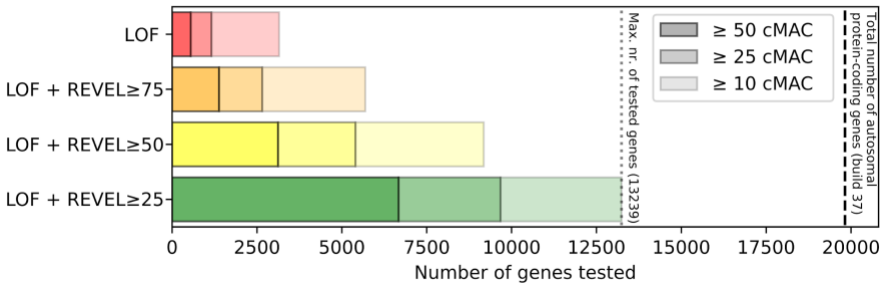
A) Sample QC We removed (1) samples with very low read coverage, (2) samples with excessive contamination, (3) samples for which the gender-annotation did not fit with the sex-chromosomal profile, (4) samples that were non-Caucasian, (5,6) samples with an excess of novel SNPs or indels, (7) samples that deviated in heterozygous/homozygous or transition/transversion ratios, (8) closely related samples (IBD), and (9) samples that were on PCR-plates that were

enriched for gender-annotation mismatches, (10) removal of samples that carried variants classified as pathogenic or likely pathogenic in Mendelian dementia genes (see supplemental data). (11) samples with a mismatch between Braak stage and AD label (AD case with Braak stage <= 1 or a control with Braak stage >= 5) or were not annotated as an AD case or control.

B) Variant QC, Multi-allelic SNPs were split into bi-allelic variants. (1) Variants that were in close vicinity, in *cis* and always occurred together, were merged into single events. (2) We designed a custom tool (see supplement to remove G>T and C>A variants, caused by the oxygenation of G bases (62). (3) Exclusion of variants in simple tandem repeat (STR) regions and low complexity regions (LCR). (4) Exclusion of variants that deviated in allele read balance (<0.25 or >0.75 for heterozygous calls and <0.9 for homozygous calls. (5) Exclusion of variants for which heterozygous calls had <20% of the coverage of reference calls. (6) Exclusion of variants that deviated from Hardy-Weinberg equilibrium in controls ($p < 5 * 10e-8$). (7) Exclusion of variants that failed VQSR (>99.5% tranche for SNPs, >99% tranche for indels). (8) Exclusion of variants that still presented batch effects that were not explainable by population structure or phenotype effects using a custom tool (see supplement). **C) Variant selection.** (1) variants in autosomal protein-coding genes that were annotated by VEP (version 94.5)(63), (2) selection of variants that directly affected the protein (missense or LOF annotation). (3) Missense variants with a REVEL score (Rare Exome Variant Ensemble Learner) (13) and LOF variants were annotated using LOFTEE (15). Selection of missense variants with a score ≥ 25 (score range 0 - 100). and LOF variants with a LOFTEE 'high-confidence' flag, and a VEP 'high impact' flag. (4) Selection of variants that were estimated to have at least one carrier, and had a minor allele frequency (MAF) of <1%. (5) Selection of variants with <20% genotyping missingness (genotypes with a read depth < 6 are considered missing) that passed a filter for differential missingness between the EOAD, LOAD and control groups. Variants were divided in 4 deleteriousness categories.

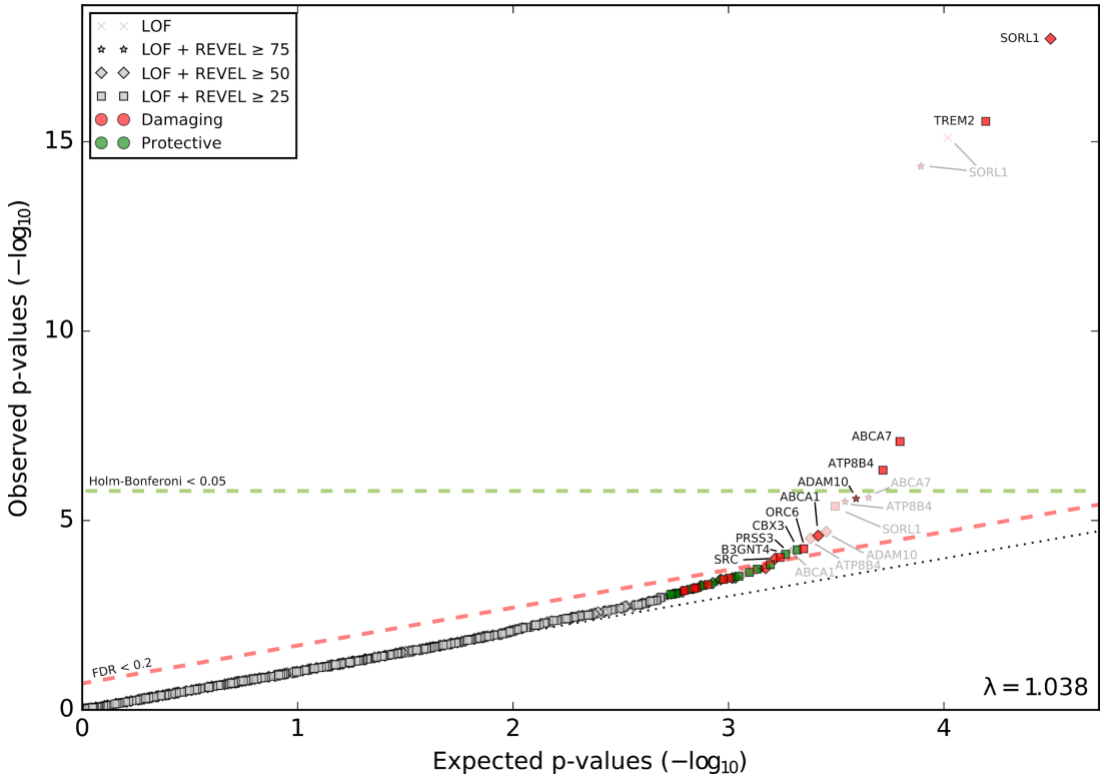
In colors the deleteriousness categories (translucent) used to construct the deleteriousness thresholds (opaque). Four different deleteriousness thresholds were used to perform burden tests. Of the missense variants, 572 were also classified as LOF variants and assigned to the LOF category.

Figure 2



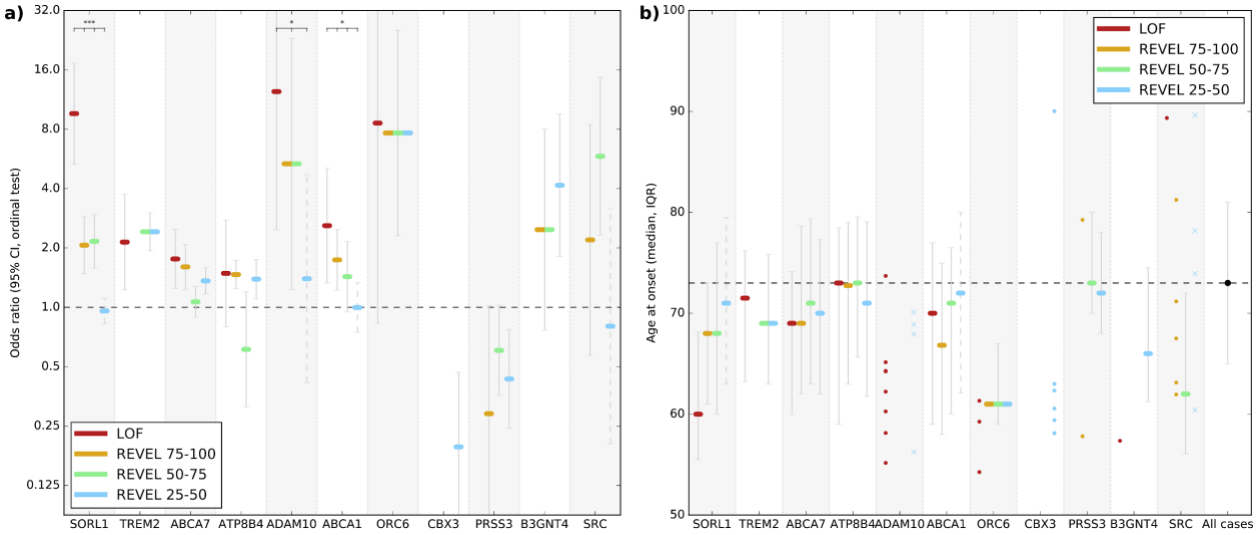
The number of genes tested per variant threshold. Only autosomal genes with a cumulative Minor Allele Count (cMAC) ≥10 were tested.

Figure 3



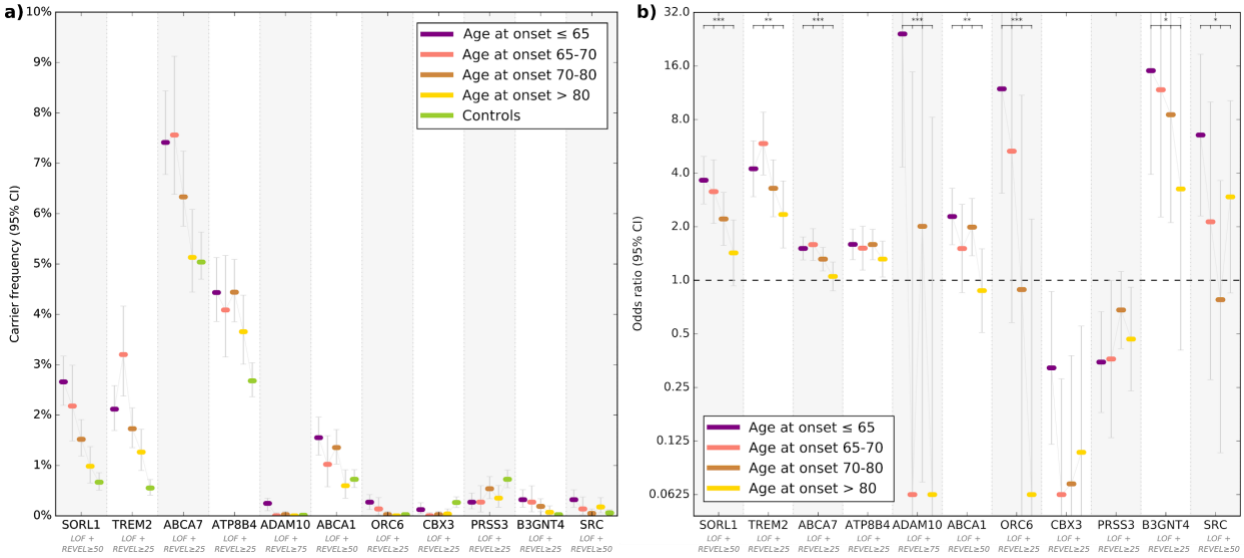
Quantile-quantile plot of observed p-values versus expected p-values in the absence of signal ($\log_{10}$ scale). In total, results of 31,568 different tests are shown, which were performed for 13,299 genes. For each gene, the most significant test is shown opaque, tests for which the signal was less significant were shown translucent. Multiple testing correction thresholds are shown for suggestive and conservative thresholds. Color indicates if burden is enriched in cases (‘Damaging’) or controls (‘Protective’).

Figure 4



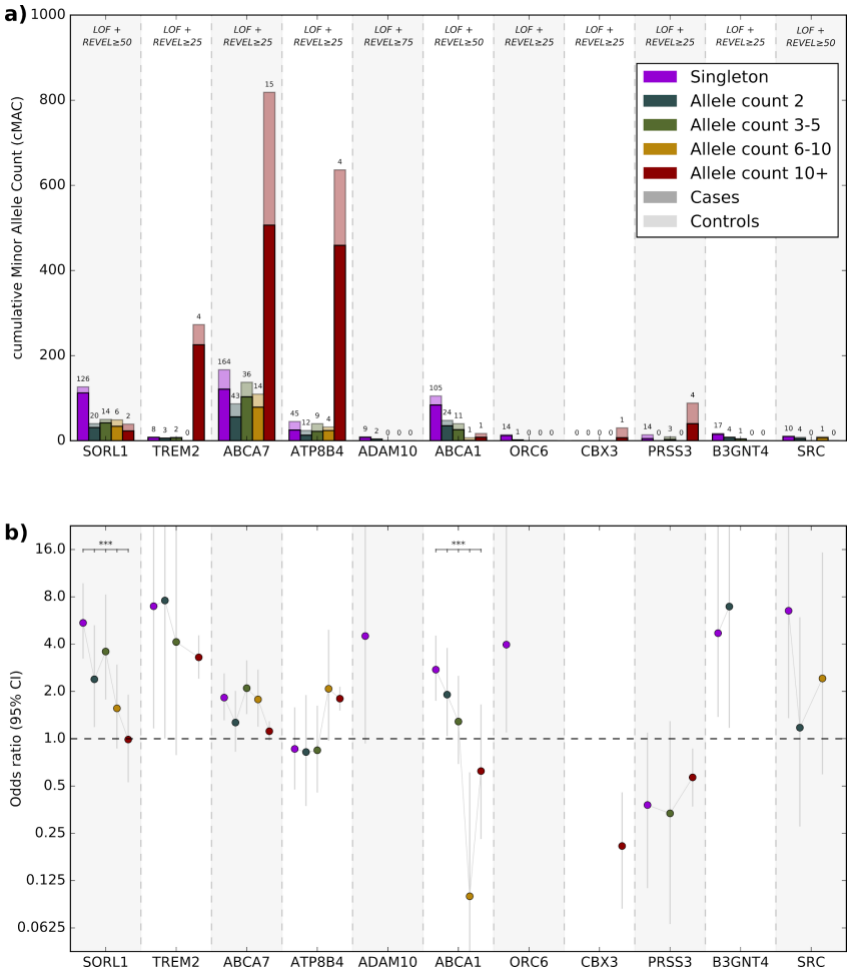
a) Odds ratios (ordinal test) per variant category. Significance is indicated if a trend in odds ratios was observed (i.e. a larger effect in the high deleteriousness categories and lower effect in lower deleteriousness categories). For missense variants, deleteriousness categories were merged when one category for REVEL (not LOF) categories if they had < 5 carriers; this was done, both for the visualization and the tests. When there were multiple neighboring deleteriousness categories to merge with, we merged with the smallest (in terms of carriers). Odds ratios for deleteriousness categories with 0 carriers and odds ratios with 0-inf confidence intervals are not shown. Categories with dashed confidence interval lines were not included in the most significant variant category. *: FDR < 0.05, **: FDR < 0.01, ***: FDR < 0.001. **b)** Age at onset per deleteriousness category and 95% CI. When the number of carrier cases per deleteriousness category was <10 carriers, the age at onset of these carriers was shown as individual dots.

Figure 5



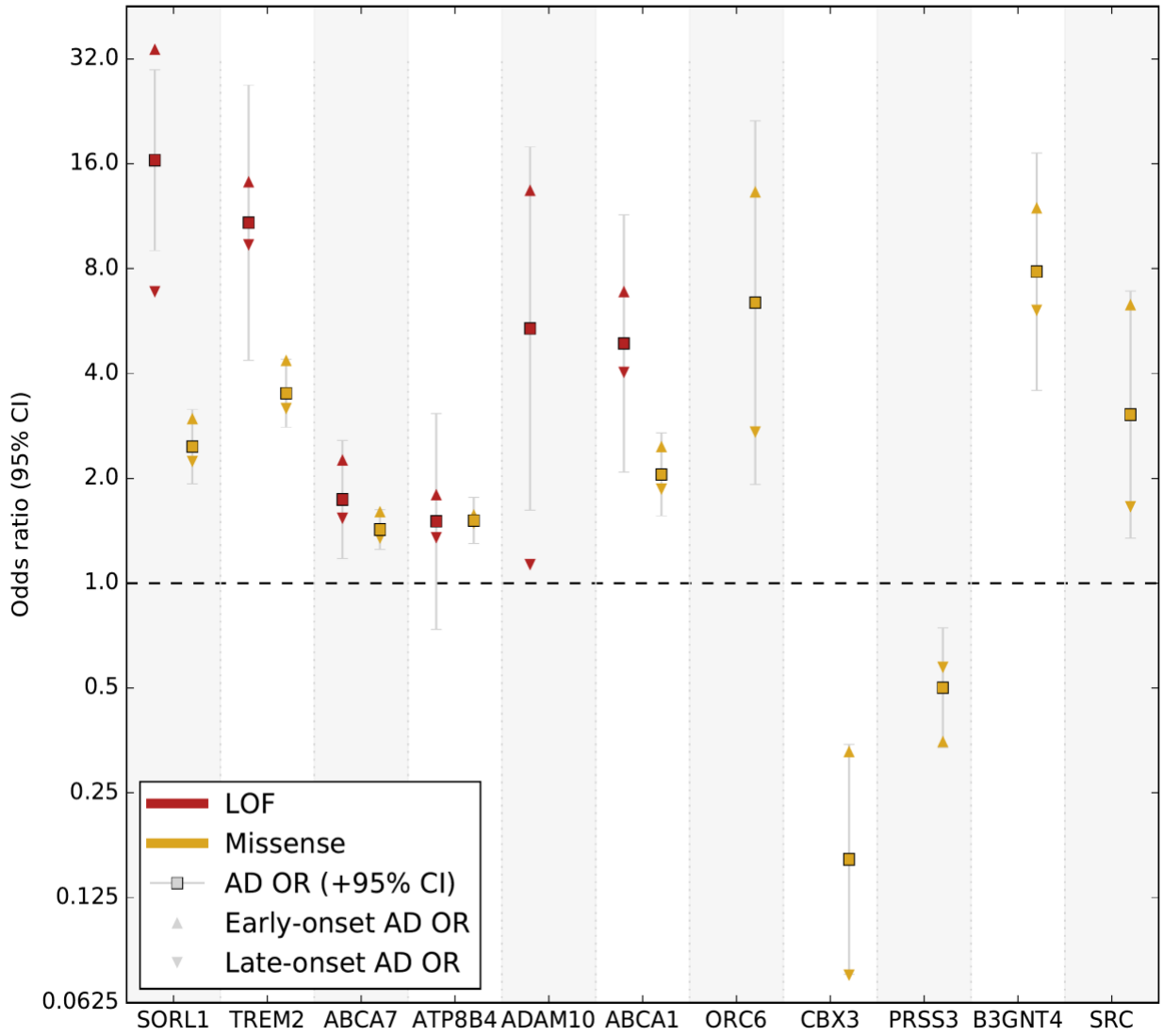
a) Carrier frequency by age at onset. Carriers have a cumulative dosage >0.5 **b)** Odds ratio by age. Odds ratios are calculated by multinomial logistic regression. Results are shown for variants in the most significant deleteriousness threshold (indicated below the gene names). The significance symbols indicate if there is a trend towards higher enrichment in younger patients (see methods). *: FDR < 0.05, **: FDR < 0.01, ***: FDR < 0.001.

Figure 6



a) Cumulative minor allele count by variant frequency For each gene, the number of variants (minor alleles) detected in cases and controls in the predicted damagingness levels threshold associated with the most significant association with AD (indicated at the top). Variants were binned according to "allele count", the occurrence of each unique variant in the sample (from extremely rare singletons to more common variants with more than 10 carriers). The number above each bar is the number of unique variants in the bin. **b) Odds ratio by variant frequency.** For the same variants and bins as in A), the odds ratio of the AD association and its confidence interval is shown. Odds ratios are not shown for bins with less than 5 carriers.

Figure 7



Odds ratios (logistic test) for LOF and missense variants after refinement analysis. Case/control (+95%CI), as well as EOAD- and LOAD-specific odds ratios are shown for variant categories with ≥5 carriers.