HHS Public Access

Author manuscript

Nat Genet. Author manuscript; available in PMC 2021 August 15.

Published in final edited form as:

Nat Genet. 2021 March ; 53(3): 294–303. doi:10.1038/s41588-021-00785-3.

Genome sequencing analysis identifies new loci associated with Lewy body dementia and provides insights into its genetic architecture

A full list of authors and affiliations appears at the end of the article.

Abstract

The genetic basis of Lewy body dementia (LBD) is not well understood. Here, we performed whole-genome sequencing in large cohorts of LBD cases and neurologically healthy controls to study the genetic architecture of this understudied form of dementia and to generate a resource for the scientific community. Genome-wide association analysis identified five independent risk loci, whereas genome-wide gene-aggregation tests implicated mutations in the gene *GBA*. Genetic risk scores demonstrate that LBD shares risk profiles and pathways with Alzheimer's disease and Parkinson's disease, providing a deeper molecular understanding of the complex genetic architecture of this age-related neurodegenerative condition.

Lewy body dementia (LBD) is a clinically heterogeneous neurodegenerative disease characterized by progressive cognitive decline, parkinsonism, and visual hallucinations[1]. There are no effective disease-modifying treatments available to slow disease progression, and current therapy is limited to symptomatic and supportive care. At postmortem, the disorder is distinguished by the widespread cortical and limbic deposition of pathologically altered forms of α-synuclein proteins in the form of Lewy bodies and Lewy neurites that are

* sonja.scholz@nih.gov.

AUTHOR CONTRIBUTIONS

C.L.D, B.J.T., and S.W.S. conceptualized and supervised the study. M.S.S., S.A., R.L.W., J.T.G., and Y.A. performed sample preparations. C.V. performed library preparations and genome sequencing. J.D., A.M., J.R.G., and C.L.D. performed genome sequence alignment, variant calling, and initial quality control checks. R.C., S.W.S., and B.J.T. curated the data. R.C. performed quality control checks and genome-wide association analysis, and Z.S. contributed to this analysis. R.C. also led the genome-wide gene-based rare variant analysis; M.B.M., M.D.-F., and C.B. contributed to this analysis. M.S.S. performed the heritability analysis. S.B.-C. performed the genetic risk score analysis. S.S.-A. computed polygenic risk scores and performed enrichment analyses. R.H.R., E.G., and M.R. performed eQTL analyses. M.A.N. consulted on the statistical analysis. M.K.P. performed validation experiments. R.H.R., M.R., M.G., A. Calvo, G.M., A. Canosa, G.F., R.C.B., F.B., Z.G.-O., P.M., R.K., D.S.G., G.L., N.T., E.S., L.N.-K., J.-A.P., H.K., V.G.S., M.P., K.L.N., E.M., R.C.K., C.A.C., E.S.M., M.B., T.M.D., L.S.R., M.S.A., O.P., J.C.T., M.E.F., Q.M., E.H.B., E.R.-R., J.I., C.L., I.G.-A., P.S.-J., B.G., J.K., S.E.B., M.M., E.R., C.D., A.B., S.L., G.X., M.J.B., B.S.T., S.G., G.L., G.E.S., T.G.B., I.G.M., A.J.T., J.A., C.M.M., L.P., S.L., C.T., S.A.-S., A.K.H., D. Aarsland, G.K., S.M.K., R.W., P.P., L.M. Bekris, J.B.L., L.M. Besser, A.K., A.E.R., A.G., D.A.B., C.R.S., H.R.M., R.F., D. Albani, S.P.-B., K.F., W.A.K., E.M.-R., A.L., J.F., D. Alcolea, J.C., L.F., S.M.R., T.T., T.M.F., N.R.G.-R., Z.K.W., T.F., B.F.B., J.A.H., D.W.D., A.B.S., A. Chiò, O.A.R., B.J.T., and S.W.S. provided biospecimens and clinical data. D.J.S., J.E., L.P., A.B.S., J.A.H., O.A., L.C., L.S.H., K. Marder, A.L., P.St.G.-H., E.L., K. Morgan, T.L., T.T.W., Z.J., D.G., I.S., P.J.T., L.M., M.O., N.J.C., J.C.M., G.M.H., V.M.V.D., J.Q.T., T.G., C.S., F.L., E.S., D.C., A. Chiò, B.J.T., and S.W.S. provided replication data. A.T., E.J.T., D.G.H., J.R.G., and A.B.S. provided convenience control genomes. S.W.S. wrote the initial manuscript. All authors critically reviewed and edited the article.

The American Genome Center

Anthony R. Soltis[129], Coralie Viollet[8,9], Gauthaman Sukumar[129], Camille Alba[129], Nathaniel Lott[129], Elisa McGrath Martinez[129], Meila Tuck[129], Jatinder Singh[129], Dagmar Bacikova[129], Xijun Zhang[129], Daniel N. Hupalo[129], Adelani Adeleye[129], Matthew D. Wilkerson[129], Harvey B. Pollard[129], and Clifton L. Dalgard[128,129]

also a hallmark feature of Parkinson's disease. The vast majority of LBD patients additionally exhibit Alzheimer's disease co-pathology[2]. These neuropathological observations have led to the, as yet unproven, hypothesis that LBD lies on a disease continuum between Parkinson's disease and Alzheimer's disease[3]. Though relatively common in the community, with an estimated 1.4 million prevalent cases in the United States[4], the genetic contributions to this underserved condition are poorly understood.

The rapid advances in genome sequencing technologies offer unprecedented opportunities to identify and characterize disease-associated genetic variation. Here, we performed whole-genome sequencing in a cohort of 2,981 patients diagnosed with LBD and 4,391 neurologically healthy individuals. We analyzed these data using a genome-wide association study (GWAS) approach. This investigation identified five risk loci that were replicated in an independent case-control cohort[5,6]. We also performed gene aggregation tests, and we modeled the relative contributions of Alzheimer's disease and Parkinson's disease risk variants to this fatal neurodegenerative disease (see Fig. 1 for an analysis overview). Additionally, we created a resource for the scientific community to mine for new insights into the genetic etiology of LBD and to expedite the development of targeted therapeutics.

Results

Genome-wide association analysis identifies new loci associated with LBD.

After quality control, whole-genome sequence data from 2,591 individuals diagnosed with LBD and 4,027 neurologically healthy individuals were available for study. Participants were recruited across 44 institutions/consortia and were diagnosed according to established consensus criteria. Using a GWAS approach, we identified five loci that surpassed the genome-wide significance threshold (Table 1 and Fig. 2a). Three of these signals were located at known LBD risk loci within the genes *GBA, APOE,* and *SNCA*[7–10]. The remaining GWAS signals in *BIN1* and *TMEM175* represented novel LBD risk loci. Notably, these loci have been implicated in other age-related neurodegenerative diseases, including Alzheimer's disease (*BIN1*) and Parkinson's disease (*TMEM175*)[11,12]. We examined the associations of *BIN1* and *TMEM175* risk alleles with CERAD and Braak semi-quantitative pathological measures of Alzheimer's disease co-pathology. We found that the *BIN1* risk allele (rs6733839-T) was significantly associated with increased neurofibrillary tangle pathology (Fisher's exact test *P*-value based on Braak neurofibrillary tangle staging = 0.0002; Extended Data Fig. 1). In contrast, there was no significant association of the *TMEM175* risk allele with Alzheimer's disease co-pathology. Conditional analyses detected a second signal at the *APOE* locus (see Extended Data Fig. 2 for regional association plots and Extended Data Fig. 3 for conditional association analyses). Subanalysis GWAS of pathologically defined LBD cases only versus control subjects identified the same risk loci (Fig. 2b). Finally, we replicated each of the observed risk loci in an independent sample of 970 European-ancestry LBD cases and 8,928 controls (Table 1)[5,6].

Gene-level aggregation testing identifies *GBA* as a pleomorphic risk gene.

The significant loci from our GWAS explained only a small fraction (1%) of the conservatively estimated narrow-sense heritability of LBD of 10.81% (95% confidence

interval [CI]: 8.28%–13.32%, $P = 9.17 \times 10^{-4}$). To explore whether rare variants contribute to the remaining risk of LBD, we performed gene-level sequence kernel association – optimized (SKAT-O) tests of missense mutations with a minor allele frequency (MAF) threshold ≤ 1% and a minor allele count (MAC) of ≥ 3 across the genome[13]. This rare variant analysis identified *GBA* as associated with LBD (Fig. 2c). *GBA*, encoding the lysosomal enzyme glucocerebrosidase, is a known pleomorphic risk gene for LBD and Parkinson's disease[7,14,15], and our rare and common variant analyses confirm a prominent role of this gene in the pathogenesis of Lewy body diseases.

Functional inferences from colocalization and gene expression analyses.

Most GWAS loci are thought to operate through the regulation of gene expression[16,17]. Thus, we performed a colocalization analysis to determine whether a shared causal variant drives association signals for LBD risk and gene expression. Expression quantitative trait loci (eQTL) were obtained from eQTLGen and PsychENCODE[18,19], the largest available human blood and brain eQTL datasets. We found evidence of colocalization between the *TMEM175* locus and an eQTL regulating *TMEM175* expression in blood (posterior probability for H_4 (PPH4) = 0.99; Fig. 3a and Supplementary Table 1). There was also colocalization between the association signal at the *SNCA* locus and an eQTL regulating *SNCA-AS1* expression in the brain (PPH4 = 0.96; Fig. 3b and Supplementary Table 1). Interestingly, the index variant at the *SNCA* locus was located within the *SNCA-AS1* gene, which overlaps with the 5'-end of *SNCA* and encodes a long noncoding antisense RNA species known to regulate *SNCA* expression. Sensitivity analyses confirmed that these colocalizations were robust to changes in the prior probability of a variant associating with both traits (Extended Data Fig. 4).

We interrogated the effect of each SNP in the region surrounding *SNCA-AS1* on LBD risk using our GWAS data and *SNCA-AS1* expression using the PsychENCODE data (Extended Data Fig. 5a). All genome-wide significant risk SNPs in the locus had a negative beta coefficient, while the shared *SNCA-AS1* eQTL had a positive beta coefficient. This negative correlation suggested that increased *SNCA-AS1* expression is associated with reduced LBD risk (Spearman's rho = −0.42; $P = 0.0012$; Extended Data Fig. 5b).

Analysis of human bulk-tissue RNA-sequencing data from the Genotype-Tissue Expression (GTEx) consortium and single-nucleus RNA-sequencing data of the medial temporal gyrus from the Allen Institute of Brain Science[20,21] demonstrated that *TMEM175* is ubiquitously expressed, whereas *SNCA-AS1* is predominantly expressed in brain tissue (Extended Data Fig. 6a and Supplementary Table 2). At the cellular level, *TMEM175* is highly expressed in oligodendrocyte progenitor cells, while *SNCA-AS1* demonstrates neuronal specificity (Extended Data Fig. 6b and Supplementary Table 2). *SNCA* and *SNCA-AS1* share a similar, though not identical, tissue expression profile (Extended Data Fig. 7).

LBD risk overlaps with risk profiles of Alzheimer's disease and Parkinson's disease.

We leveraged our whole-genome sequence data to explore the etiological relationship between Alzheimer's disease, Parkinson's disease, and LBD. To do this, we applied genetic risk scores derived from large-scale GWAS analyses of Alzheimer's disease and Parkinson's

disease to individual-level genetic data from our LBD case-control cohort[22,23]. We tested the associations of the Alzheimer's disease and Parkinson's disease genetic risk scores with LBD disease status, and with age at death, age at onset, and the duration of illness observed among the LBD cases.

Individuals diagnosed with LBD had a higher genetic risk for developing both Alzheimer's disease (odds ratio [OR] = 1.66 per standard deviation of Alzheimer's disease genetic risk, 95% CI = 1.58–1.74, $P < 2 \times 10^{-16}$, Fig. 5a) and Parkinson's disease (OR = 1.20, 95% CI =1.14–1.26, $P = 4.34 \times 10^{-12}$, Fig. 5b). These risk scores remained significant after adjusting for genes that substantially contribute to Alzheimer's disease (model after adjustment for *APOE*: OR = 1.53, 95% CI = 1.37–1.72, $P = 3.29 \times 10^{-14}$) and Parkinson's disease heritable risk (model after adjustment for *GBA, SNCA*, and *LRRK2*: OR = 1.26, 95% CI = 1.19–1.34, $P = 5.91 \times 10^{-14}$). The Alzheimer's disease genetic risk score was also found to be significantly associated with an earlier age of death in LBD (β = −1.77 years per standard deviation increase in the genetic risk score from the population mean, standard error [SE] = 0.19, $P < 2 \times 10^{-16}$) and shorter disease duration (β = −0.90 years, SE = 0.27, P = 0.0007). In contrast, the Parkinson's disease genetic risk score was associated with an earlier age at onset among patients diagnosed with LBD (β = −0.98, SE = 0.28, P = 0.00045), indicating that higher Parkinson's disease risk is associated with earlier age at onset in LBD. We found no evidence of interaction between the genetic risk scores of Alzheimer's disease and Parkinson's disease in the LBD cohort (OR = 0.99, 95 % CI = 0.95–1.03, $P = 0.59$), implying that Alzheimer's disease and Parkinson's disease risk variants are independently associated with LBD risk.

Enrichment analysis identifies pathways involved in LBD.

Pathway enrichment analysis of LBD, using a polygenic risk score based on the GWAS risk variants, found several significantly enriched gene ontology processes associated with LBD (Fig. 5). These related to the *regulation of amyloid-beta formation* (adjusted $P = 0.04$), *regulation of endocytosis* (adjusted $P = 0.02$), *tau protein binding* (adjusted $P = 1.85 \times 10^{-5}$), and others. Among these, the regulation of *amyloid precursor protein, amyloid-beta formation*, and *tau protein binding* have been previously implicated in the pathogenesis of Alzheimer's disease, while regulation of endocytosis is particularly important in the pathogenesis of Parkinson's disease[24,25]. These observations support the notion of overlapping disease-associated pathways in these common age-related neurodegenerative diseases.

Association of polygenic risk with clinical dementia severity.

We performed an association analysis of LBD polygenic risk with dementia severity, as measured by the Clinical Dementia Rating scale[26]. We found that LBD patients in the highest polygenic risk score quintile had more severe impairment at baseline evaluation compared to LBD patients in the lowest quintile ($\chi^2 = 5.60$, df = 1, $P = 0.009$; Extended Data Fig. 8).

Discussion

Our analyses highlight the contributions of common and rare variants to the complex genetic architecture of LBD, a common and fatal neurodegenerative disease. Specifically, our GWAS identified five independent genome-wide significant loci (*GBA, BIN1, TMEM175, SNCA-AS1, APOE*) that influence risk for developing LBD, whereas the genome-wide gene-based aggregation tests implicated mutations in *GBA* as being critical in the pathogenesis of the disease. We further detected strong cis-eQTL colocalization signals at the *TMEM175* and *SNCA-AS1* loci, indicating that the risk of disease at these genomic regions may be driven by expression changes of these particular genes. Finally, we provided definitive evidence that the risk of LBD is driven, at least in part, by genetic variants associated with the risk of developing both Alzheimer's disease and Parkinson's disease.

We replicated all five GWAS signals in an independent LBD case-control dataset derived from imputed genotyping array data. Among these, *GBA* (encoding the lysosomal enzyme glucocerebrosidase), *APOE* (encoding apolipoprotein E), and *SNCA* (encoding α-synuclein) are known LBD risk genes[7–9]. In addition to these previously described loci, we identified a novel locus on chromosome 2q14.3, located 28 kb downstream of the *BIN1* gene, which is a known risk locus for Alzheimer's disease[11]. *BIN1* encodes the bridging integrator 1 protein that is involved in endosomal trafficking. The depletion of *BIN1* reduces the lysosomal degradation of β-site APP-cleaving enzyme 1 (BACE1), resulting in increased amyloid-β production[27]. Furthermore, the loss of *BIN1* promotes the propagation of tau pathology by increasing aggregate internalization via endocytosis and endosomal trafficking[28]. The direction of effect observed in LBD is the same as in Alzheimer's disease (Supplementary Table 3). The observed pleiotropic effects between LBD and Alzheimer's disease prompt us to speculate that mitigating *BIN1*-mediated endosomal dysfunction could have therapeutic implications in both neurodegenerative diseases.

A second novel LBD signal was detected within the lysosomal *TMEM175* gene on chromosome 4p16.3, a known Parkinson's disease risk locus[12]. Deficiency of *TMEM175*, encoding a transmembrane potassium channel, impairs lysosomal function, lysosome-mediated autophagosome clearance, and mitochondrial respiratory capacity. Loss-of-function further increases the deposition of phosphorylated α-synuclein[29], which makes *TMEM175* a plausible LBD risk gene. The direction of effect is the same in LBD as it is in Parkinson's disease (Supplementary Table 3), and identification of *TMEM175* underscores the role of lysosomal dysfunction in the pathogenesis of Lewy body diseases.

Our data confirm the hypothesis that the LBD genetic architecture is complex and overlaps with the risk profiles of Alzheimer's disease and Parkinson's disease. First, several genome-wide significant risk loci in our GWAS analysis have been previously described either in the Alzheimer's disease literature (*APOE, BIN1*) or have been associated with risk of developing Parkinson's disease (*GBA, TMEM175, SNCA*)[11,12,30–32]. Second, genome-wide gene-based aggregation tests of rare mutations similarly identified *GBA*, which has been previously implicated in Parkinson's disease[7]. Third, genetic risk scores derived from Alzheimer's disease and Parkinson's disease GWAS meta-analyses predicted risk for LBD independently, even after removal of the strongest signals (*APOE, GBA, SNCA,* and

LRRK2). Interestingly, our data did not show a synergistic effect between the risk of Parkinson's disease and Alzheimer's disease in the pathogenesis of LBD, though analysis of larger cohorts will be required to confirm this observation.

Comparing the patterns of the risk loci in LBD with the patterns of risk in published Parkinson's disease and Alzheimer's disease GWAS meta-analyses provided additional insights into this complex relationship. The directions of effect at the index variants of the *GBA* and *TMEM175* loci were the same in LBD as the directions observed in Parkinson's disease[23]. Likewise, the directions of effect for the *BIN1* and *APOE* signals were the same as the directions detected in Alzheimer's disease (Supplementary Table 3)[33]. However, we observed a notably different profile at the *SNCA* locus in LBD compared to Parkinson's disease. Our GWAS and colocalization analyses implicated *SNCA-AS1*, a non-coding RNA that regulates *SNCA* expression, as the main signal at the *SNCA* locus. In contrast, the main signal in Parkinson's disease is detected at the 3'-end of *SNCA*[34]. This finding suggests that the regulation of *SNCA* expression may be different in LBD compared to Parkinson's disease and that only specific *SNCA* transcripts that are regulated by *SNCA-AS1* drive risk for developing dementia. Further, *SNCA-AS1* may prove to be a more amenable therapeutic target than *SNCA* itself due to its neuronal specificity.

As part of this study, we created a foundational resource that will facilitate the study of molecular mechanisms across a broad spectrum of neurodegenerative diseases. We anticipate that these data will be widely accessed for several reasons. First, the resource is the largest whole-genome sequence repository in LBD to date. Second, the nearly 2,000 neurologically healthy, aged individuals included within this resource can be used as control subjects for the study of other neurological and age-related diseases. Third, we prioritized the inclusion of pathologically confirmed LBD patients, representing more than two-thirds of the case cohort, to ensure high diagnostic accuracy among our case cohort participants. Finally, all genomes are of high quality and were generated using a uniform genome sequencing, alignment, and variant-calling pipeline. Whole genome sequencing data on this large case-control cohort has allowed us to undertake a comprehensive genomic evaluation of both common and rare variants, including immediate fine-mapping of association signals to pinpoint the functional variants at the *TMEM175* and *SNCA-AS1* loci. The availability of genome-sequence data will facilitate similar comprehensive evaluations of less commonly studied variant types, such as repeat expansions and structural variants.

Our study has limitations. We focused on individuals of European ancestry, as this is the population in which large cohorts of LBD patients were readily available. Recruiting patients and healthy controls from diverse populations will be crucial for future research to understand the genetic architecture of LBD. Another constraint is the use of short-read sequencing, rather than long-read sequencing applications, that limits the resolution of complex, repetitive, and GC-rich genomic regions[35]. Most study participants did not have in-depth phenotype information using standardized rating scales available. Further, despite our large sample size, we had limited power to detect common genetic variants of small effect size, and additional large-scale genomic studies will be required to unravel the missing heritability of LBD.

In conclusion, our study identified novel loci as relevant in the pathogenesis of LBD. Our findings confirmed that LBD genetically intersects with Alzheimer's disease and Parkinson's disease and highlighted the polygenic contributions of these other neurodegenerative diseases to its pathogenesis. Determining shared molecular genetic relationships among complex neurodegenerative diseases paves the way for precision medicine and has implications for prioritizing targets for therapeutic development. We have made the whole-genome sequence data available to the research community. These genomes constitute the largest sequencing effort in LBD to date and are designed to accelerate the pace of discovery in dementia.

Methods

Cohort description and study design.

A total of 5,154 participants of European ancestry (2,981 LBD cases, 2,173 neurologically healthy controls) were recruited across 17 European and 27 North American sites/consortia to create a genomic resource for LBD research (Supplementary Table 4). In addition to these resource genomes, we obtained convenience control genomes from (i) the Wellderly cohort (n = 1,202), a cohort of healthy, aged European-ancestry individuals recruited in the United States[36], and (ii) European-ancestry control genomes generated by the National Institute on Aging and the Accelerating Medicine Partnership - Parkinson's Disease Initiative (www.amp-pd.org; n = 1,016). This brought the total number of control individuals available for this study to 4,391.

All control cohorts were selected based on a lack of evidence of cognitive decline in their clinical history and absence of neurological deficits on neurological examination. Pathologically confirmed control individuals (n = 605) had no evidence of significant neurodegenerative disease on histopathological examination. LBD patients were diagnosed with pathologically definite or clinically probable disease according to consensus criteria[2,37]. The case cohort included 1,789 (69.0%) autopsy-confirmed LBD cases and 802 (31.0%) clinically probable LBD patients. 63.4% of LBD cases were male, as is typical for the LBD patient population[38]. The demographic characteristics of the cohorts are summarized in Supplementary Table 5. The appropriate institutional review boards of participating institutions approved the study (03-AG-N329, NCT02014246), and informed consent was obtained from all subjects or their surrogate decision-makers, according to the Declaration of Helsinki.

Whole-genome sequencing.

Fluorometric quantitation of the genomic DNA samples was performed using the PicoGreen dsDNA assay (Thermo Fisher). PCR-free, paired-end libraries were constructed by automated liquid handlers using the Illumina TruSeq chemistry according to the manufacturer's protocol. DNA samples underwent sequencing on an Illumina HiSeq X Ten sequencer (v.2.5 chemistry, Illumina) using 150 bp, paired-end cycles.

Sequence alignment, variant calling.

Genome sequence data were processed using the pipeline standard developed by the Centers for Common Disease Genomics (CCDG; https://www.genome.gov/27563570/). This standard allows for whole-genome sequence data processed by different groups to generate 'functionally equivalent' results[39]. The GRCh38DH reference genome was used for alignment, as specified in the CCDG standard. For whole-genome sequence alignments and processing, the Broad Institute's implementation of the functional equivalence standardized pipeline was used. This pipeline, which incorporates the GATK (2016) Best Practices[40], was implemented in the workflow description language for deployment and execution on the Google Cloud Platform. Single-nucleotide variants and indels were called from the processed whole-genome sequence data following the GATK Best Practices using another Broad Institute workflow for joint discovery and Variant Quality Score Recalibration. Both Broad workflows for WGS sample processing and joint discovery are publicly available (https://github.com/gatk-workflows/broad-prod-wgs-germline-snps-indels). All whole-genome sequence data were processed using the same pipeline.

Quality control.

For sample-level quality control checks, genomes were excluded from the analysis for the following reasons: (1) a high contamination rate (>5% based on VerifyBamID freemix metric)[41], (2) an excessive heterozygosity rate (exceeding +/− 0.15 F-statistic), (3) a low call rate (≤ 95%), (4) discordance between reported sex and genotypic sex, (5) duplicate samples (determined by pi-hat statistics > 0.8), (6) non-European ancestry based on principal components analysis when compared to the HapMap 3 Genome Reference Panel (Extended Data Fig. 9a)[42], and (7) samples that were related (defined as having a pi-hat > 0.125).

For variant-level quality control, we excluded: (1) variants that showed non-random missingness between cases and controls ($P \leq 1 \times 10^{-4}$), (2) variants with haplotype-based non-random missingness ($P \leq 1 \times 10^{-4}$), (3) variants with an overall missingness rate of ≥ 5%, (4) non-autosomal variants (X, Y, and mitochondrial chromosomes), (5) variants that significantly departed from Hardy-Weinberg equilibrium in the control cohort ($P \leq 1 \times 10^{-6}$), (6) variants mapping to variable, diversity, and joining (VDJ) recombination sites, as well as variants in centromeric regions +/− 10 kb (due to poor sequence alignment and incomplete resolution of the reference genome assembly at these sites)[43], (7) variants for which the allele frequency in the aged control subjects (Wellderly cohort) significantly deviated from the other control cohorts (non-Wellderly) based on FDR-corrected chi-square tests ($P < 0.05$), (8) variants for which the MAFs in our control cohorts significantly differed from reported frequencies in the NHLBI Trans-Omics TOPMed database (freeze 5b; www.nhlbiwgs.org) or gnomAD (version 3.0) (FDR-corrected chi-square test $P < 0.05$)[44], (9) variants that failed TOPMed variant calling filters, and (10) spanning deletions.

After these quality control filters were applied, there were 6,651 samples available for analysis. Extended Data Figure 10 shows quality control metrics.

Statistical analysis for single-variant association.

We performed a GWAS in LBD (n = 2,591 cases and 4,027 controls) using logistic regression in PLINK (v.2.0) with a minor allele frequency threshold of >1% based on the allele frequency estimates in the LBD case cohort[45]. We used the step function in the R MASS package to determine the minimum number of principal components (generated from common single nucleotide variants) required to correct for population substructure[46]. The first two principal components in our study cohorts compared to the HapMap3 Genomic Resource Panel are shown in Extended Data Figure 9a. Based on this analysis, we incorporated sex, age, and five principal components (PC1, PC3, PC4, PC5, PC7) as covariates in our model. Quantile-quantile plots revealed minimal residual population substructure, as estimated by the sample size-adjusted genome-wide inflation factor λ_{1000} of 1.004 (Extended Data Fig. 9b). The Bonferroni threshold for genome-wide significance was 5.0×10^{-8}. A conditional analysis was performed for each GWAS locus by adding each respective index variant to the covariates (Extended Data Fig. 3).

For the LBD GWAS replication analysis, we obtained genotyping array data from two independent, non-overlapping, European-ancestry LBD case-control cohorts, totaling 970 LBD cases and 8,928 controls, as described elsewhere[5,6]. The data were cleaned by applying the same sample- and variant-level quality control steps that were used in the discovery genomes. We imputed the data against the NHLBI TOPMed imputation reference panel under default settings with Eagle v.2.4 phasing[47–49]. Variants with an R^2 value < 0.3 were excluded. A meta-analysis of the two cohorts was performed with METAL under a fixed-effects model and variants that were significant in the discovery stage were extracted[50].

Genotype-pathology association analysis.

We evaluated the association of the newly identified LBD risk alleles in *BIN1* (rs6733839-T) and *TMEM175* (rs6599388-T) with the pathological changes of Alzheimer's disease. Neuritic plaque staging information, assessed by the CERAD method[51], was available for 700 pathologically confirmed LBD cases, while neurofibrillary tangle pathology staging, as assessed by Braak method[52], was available for 1,459 definite LBD cases. Association testing between the risk alleles and the semi-quantitative neuritic plaque and neurofibrillary tangle burden was performed using Fisher's exact tests.

Colocalization analyses.

Coloc (v.4.0.1) was used to evaluate the probability of LBD loci and expression quantitative trait loci (eQTL) sharing a single causal variant[53]. This tool incorporates a Bayesian statistical framework that computes posterior probabilities for five hypotheses: namely, there is no association with either trait (hypothesis 0, H_0); an associated LBD variant exists but no associated eQTL variant (H_1); there is an associated eQTL variant but no associated LBD variant (H_2); there is an association with an eQTL and LBD risk variant, but they are two independent variants (H_3); and there is a shared associated LBD variant and eQTL variant within the analyzed region (H_4). Cis-eQTL were derived from eQTLGen (n = 31,684 individuals; accessed 19 February 2020) and PsychENCODE (n = 1,387 individuals; accessed 20 February 2020)[18,19]. For each locus, we examined all genes within 1 Mb of a significant region of interest, as defined by our LBD GWAS ($P < 5.0 \times 10^{-8}$). Coloc was run

using the default $p_1 = 10^{-4}$ and $p_2 = 10^{-4}$ priors, while the p_{12} prior was set to $p_{12} = 5 \times 10^{-6}$ [54]. Loci with a posterior probability for H_4 (PPH4) ≥ 0.90 were considered colocalized. All colocalizations were subjected to sensitivity analyses to explore the robustness of our conclusions to changes in the p_{12} prior (i.e., the probability that a given variant affects both traits).

Cell-type and tissue specificity measures.

To determine specificity of a gene's expression to a tissue or cell-type, specificity values were generated from two independent gene expression datasets: (1) bulk-tissue RNA-sequencing of 53 human tissues from the Genotype-Tissue Expression consortium (GTEx; v.8)[21]; and (2) human single-nucleus RNA-sequencing of the middle temporal gyrus from the Allen Institute for Brain Science (n = 7 cell types)[20]. Specificity values for GTEx were generated using modified code from a previous publication[55]. Expression of tissues was averaged by organ (except in the case of brain; n = 35 tissues in total). Specificity values for the Allen Institute for Brain Science-derived dataset were generated using gene-level exonic reads and the 'generate.celltype.data' function of the EWCE package[56]. The specificity values for both datasets and the code used to generate these values are available at https://github.com/RHReynolds/MarkerGenes.

Heritability analysis.

The narrow-sense heritability (h^2), a measure of the additive genetic variance, was calculated using GREML-LDMS to determine how much of the genetic liability for LBD is explained by common genetic variants[57]. This analysis included unrelated individuals (pi-hat < 0.125, n = 2,591 LBD cases, and n = 4,027 controls) and autosomal variants with a MAF >1%. The analysis was adjusted for sex, age, and five principal components (PC1, PC3, PC4, PC5, PC7), and a disease prevalence of 0.1% to account for ascertainment bias.

Gene-based rare variant association analysis.

We conducted a genome-wide, gene-based sequence kernel association test - optimized (SKAT-O) analysis of missense mutations to determine the difference in the aggregate burden of rare coding variants between LBD cases and controls[64]. This analysis was performed in RVTESTS (v.2.1.0) using default parameters after annotating variants in ANNOVAR (v.2018–04/16)[58,59]. The study cohort for this analysis consisted of 2,591 LBD cases and 4,027 control subjects. We used a MAF threshold of ≤ 1% and a minor allele count (MAC) of ≥ 3 as filters. The covariates used in this analysis included sex, age, and five principal components (PC1, PC3, PC4, PC5, PC7). The Bonferroni threshold for genome-wide significance was 2.86×10^{-6} (0.05 / 17,483 autosomal genes tested).

Predictions of LBD risk using Alzheimer's disease and Parkinson's disease risk scores.

Genetic risk scores were generated using PLINK (v.1.9) based on summary statistics from recent Alzheimer's disease and Parkinson's disease GWAS meta-analyses. Considering the LBD cohort as our target dataset, risk allele dosages were counted across Alzheimer's disease or Parkinson's disease loci per sample (i.e., giving a dose of two if homozygous for the risk allele, one if heterozygous, and zero if homozygous for the alternate allele). The

SNPs were weighted by their log odds ratios, giving greater weight to alleles with higher risk estimates, and a composite genetic risk score was generated across all risk loci. Genetic risk scores were *z*-transformed prior to analysis, centered on controls, with a mean of zero and a standard deviation of one in the control subjects. Regression models were then applied to test for association with the risk of developing LBD (based on logistic regression) or the age at death, age at onset, and disease duration (linear regression), adjusting for sex, age (risk and disease duration only), and five principal components (PC1, PC3, PC4, PC5, PC7) to account for population stratification.

Polygenic risk score generation for pathway enrichment and phenotype associations.

A genome-wide LBD polygenic risk score was generated using PRSice-2. The polygenic risk score was computed by summing the risk alleles associated with LBD that had been weighted by the effect size estimates generated by performing a GWAS in the pathologically confirmed LBD cases and controls. This workflow identified the optimum *P*-value threshold (1×10^{-4} in our dataset) for variant selection, allowing for the inclusion of variants that failed to reach genome-wide significance but that contributed to disease risk, nonetheless. After excluding variants without an rs-identifier, the remaining 122 variants were ranked based on their GWAS *P*-values, with the *APOE*, *GBA*, *SNCA*, *BIN1* and *TMEM175* genes added to the top five positions. The list was then analyzed for pathway enrichment using the g:Profiler toolkit (v.0.1.8). We defined the genes involved in the pathways and gene sets using the following databases: (i) Gene Ontology, (ii) Kyoto Encyclopedia of Genes and Genomes, (iii) Reactome, and (iv) WikiPathways[60,61]. Significant pathways and gene lists with a single gene or containing more than 1,000 genes were discarded. Significance was defined as $P < 0.05$. The g:Profiler algorithm applies a Bonferroni correction to the *P*-value for each pathway to correct for multiple testing.

Next, we tested whether the same LBD polygenic risk scores were associated with cognitive impairment, as measured by the Clinical Dementia Rating scale. This analysis was performed in the 214 LBD cases provided by the National Alzheimer's Coordinating Center, as this was the only cohort for which the Clinical Dementia Rating scale had been collected at baseline evaluation. Genetic risk scores were *z*-transformed before separating all cases into quintiles based on their individual polygenic risk scores. A two-proportions *z*-test was performed to compare the proportion of severe LBD cases within the highest genetic risk score quintile group versus the lowest quintile.

Data availability.

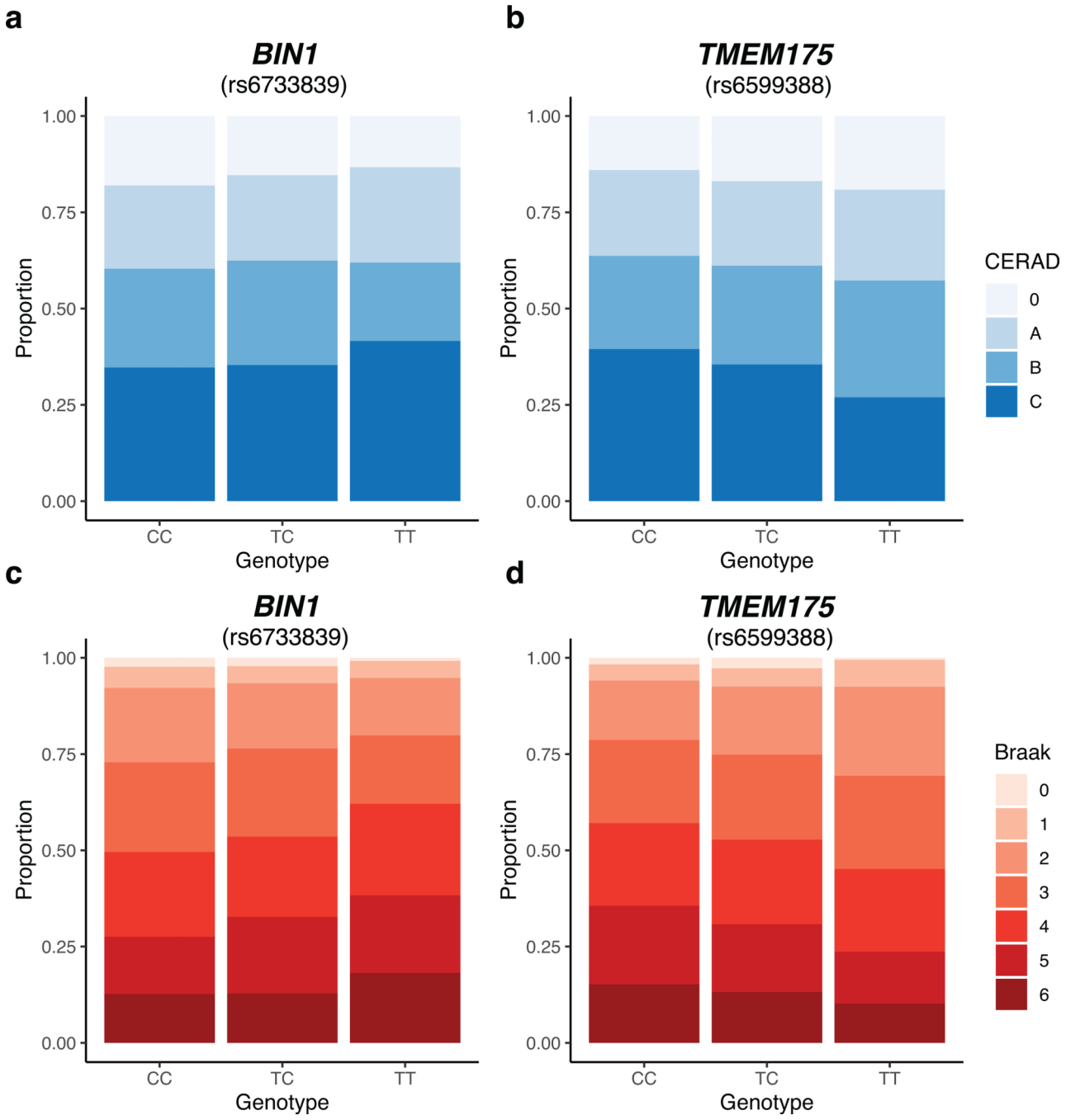
The individual-level sequence data for the resource genomes have been deposited at dbGaP (accession number: phs001963.v1.p1 NIA DementiaSeq). The GWAS summary statistics have been deposited in the GWAS catalog: https://www.ebi.ac.uk/gwas/home. eQTLGen data are available at https://www.eqtlgen.org/cis-eqtls.html. PsychENCODE QTL data are available at http://resource.psychencode.org/. Bulk-tissue RNA sequence data (GTEx version 8) are available at the Genotype-Tissue Expression consortium portal (https://www.gtexportal.org/home/). Human single-nucleus RNA sequence data are available at the Allen Institute for Brain Science portal (portal.brain-map.org/atlases-and-data/rnaseq/

human-mtg/smart-seq). Specificity values for the Allen Institute for Brain Science and GTEx data are available at: https://github.com.RHReynolds/MarkerGenes.

Code availability.

Analyses were performed using open-source tools and code for analysis is available at the associated website of each software package. Genome sequence alignment and variant calling followed the implementation of the GATK Best Practices pipeline (v.2016-June) (https://github.com/gatk-workflows/broad-prod-wgs-germline-snps-indels). Contamination rates were assessed using VerifyBamID (v.1.1.3) (https://genome.sph.umich.edu/wiki/VerifyBamID). Quality control checks, association analyses, and conditional analyses were performed in PLINK2 (v.2.0-dev-20191128) (https://www.cog-genomics.org/plink/2.0/). Data formatting and visualizations were performed in R (version 3.5.2; https://www.r-project.org) using the following packages: MASS (v.7.3–51.4), tidyverse (v.1.2.1), stringr (v.1.4.0), ggrepel (v.0.8.1), data.table (v.1.12), viridis (v.0.5.1), ggplot2 (v.3.3.2), gridExtra (v.2.3), grid (v.3.5.2). Imputation was performed using Minimac4 on data phased by Eagle (v.2.4) (https://github.com/poruloh/Eagle). Meta-analysis was performed using METAL (v.2018–08-28) (https://genome.sph.umich.edu/wiki/METAL). Heritability analysis was performed using GRML-LDMS in GCTA (v.1.26.0) (https://cnsgenomics.com/software/gcta). Rare variant analysis was performed using RVTESTS (v.2.1.0) (http://zhanxw.github.io/rvtests/) after annotating variant files in ANNOVAR (v.2018–04/16) (https://doc-openbio.readthedocs.io/projects/annovar/en/latest/). Genetic risk score analyses were performed in PLINK 1.9 (v.1.9.0-beta4.4) (https://www.cog-genomics.org/plink). LBD summary statistics were converted from hg38 to hg19 using the R implementation of the LiftOver tool, which is available from the rtracklayer package (v.1.42.2) (genome.sph.umich.edu/wiki/LiftOver). Colocalization analyses were performed in R-3.2 using the packages coloc (v.4.0.1) (https://github.com/chr1swallace/coloc). Specificity values for the AIBS-derived dataset were generated using gene-level exonic reads and the ‘generate.celltype.data’ function of the EWCE package (v.0.99.2) (https://github.com/NathanSkene/EWCE). Polygenic risk scores were constructed using PRSice-2 (v.2.1.1) (https://www.prsice.info). Pathway enrichment analysis was performed using the R package gprofiler2 (v.0.2.0) (https://cran.r-project.org/web/packages/gprofiler2/vignettes/gprofiler2.html).

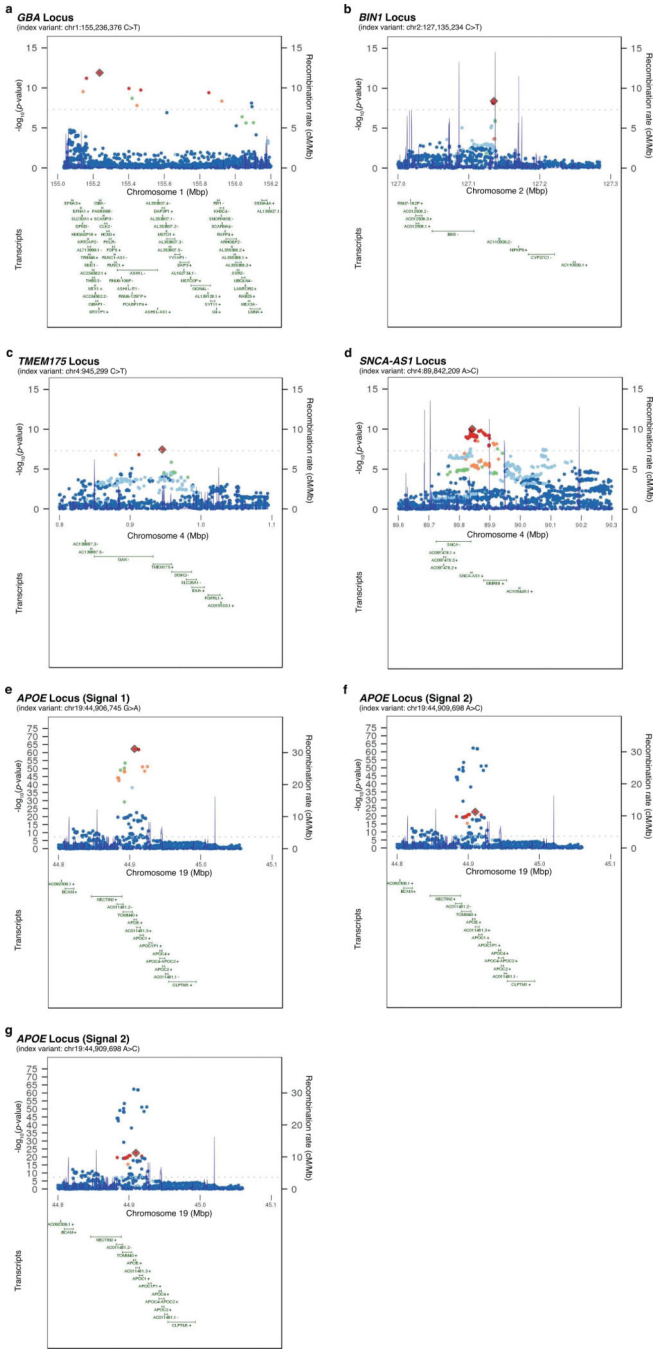
Extended Data



Extended Data Fig. 1. *BIN1* and *TMEM175* genotype-phenotype analysis

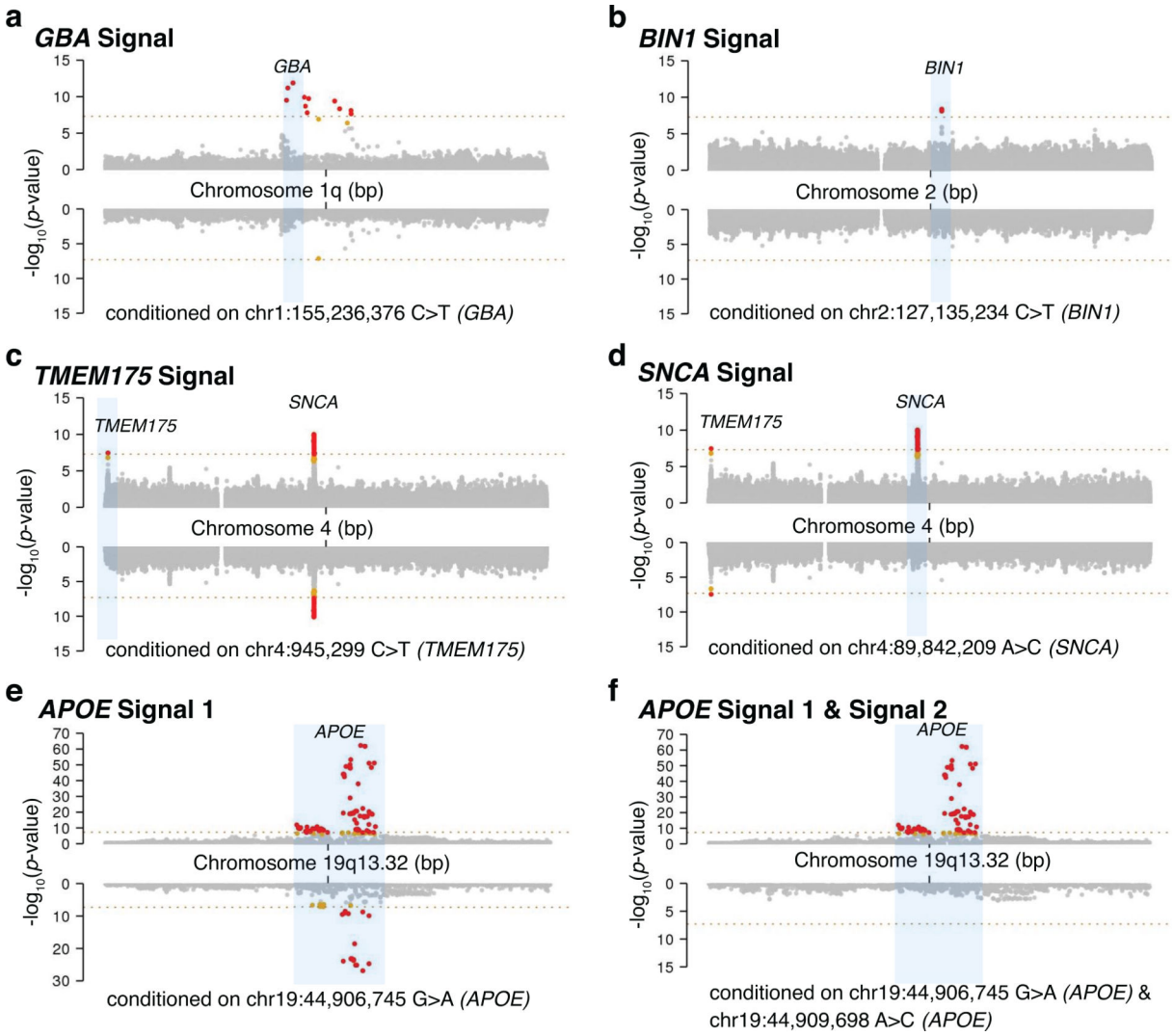
Relationship between *BIN1* and *TMEM175* genotypes and the presence of Alzheimer's disease co-pathology in definite LBD cases. The color gradation refers to semi-quantitative pathological measures of neuritic plaques (assessed by CERAD method) and neurofibrillary tangles (assessed by Braak stage). Darker colors refer to higher burden of pathology. Homozygous *BIN1* risk allele carriers (TT) were found to have significantly increased neurofibrillary tangle pathology compared to homozygous major allele carriers (CC;

Fisher's exact test *P*-value on Braak staging = 0.0002). Although the proportion of LBD cases that had high neuritic plaque burden was higher in homozygous risk allele carriers compared to homozygous major allele carriers, the difference between these groups was not statistically significant ($P = 0.23$). There was no association of *TMEM175* risk allele dosage and Alzheimer's disease co-pathology, though a trend toward lower Alzheimer's disease co-pathology was observed among homozygous *TMEM175* risk allele carriers.



Extended Data Fig. 2. Regional association plots

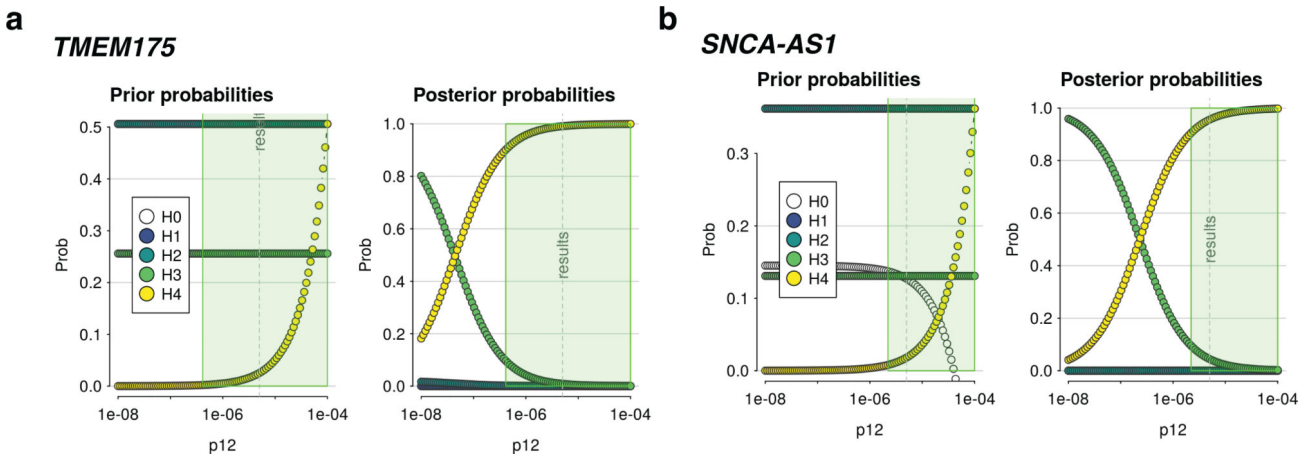
a-g, Regional association plots, local linkage disequilibrium, and recombination rates at the significantly associated LBD GWAS risk signals. Regional associations are plotted as a function of their genomic position, denoting the index variant by a red diamond. Single nucleotide variants or indels surrounding the index variant are color-coded to reflect the strength of linkage disequilibrium with the index variant based on pairwise r^2-values in the study cohort (red, $1.0 \geq r^2 \geq 0.8$; orange, $0.8 > r^2 \geq 0.6$; green $0.6 > r^2 \geq 0.4$; light blue, $0.4 > r^2 \geq 0.2$; dark blue, $0.2 > r^2 \geq 0$; gray, no r^2 value available). Transcript annotations according to the University of California Santa Cruz genome browser are depicted under each association plot.



Extended Data Fig. 3. Conditional analysis

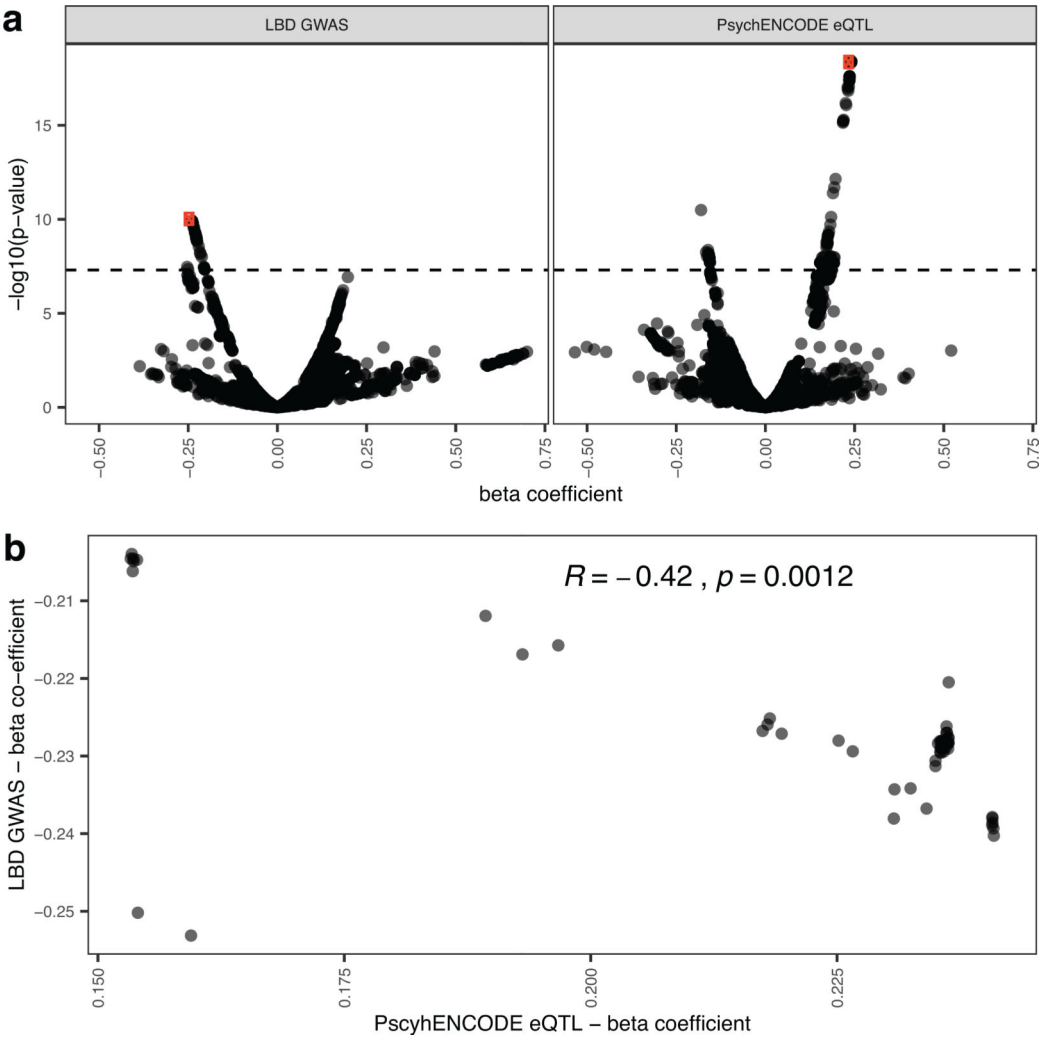
a-f, Conditional analyses for all genome-wide significant GWAS signals are depicted. For each panel, the *x*-axis denotes the chromosomal position in build 38, and the *y*-axis indicates the association *P*-values on a $-\log_{10}$ scale. The unconditioned GWAS signal is shown in the upper pane of each panel, while the lower pane illustrates the association results after correction for the index variant(s) at each respective signal. This analysis demonstrated two

signals at the *APOE* locus (**e, f**). The locus name is based on the closest gene to the index variant.



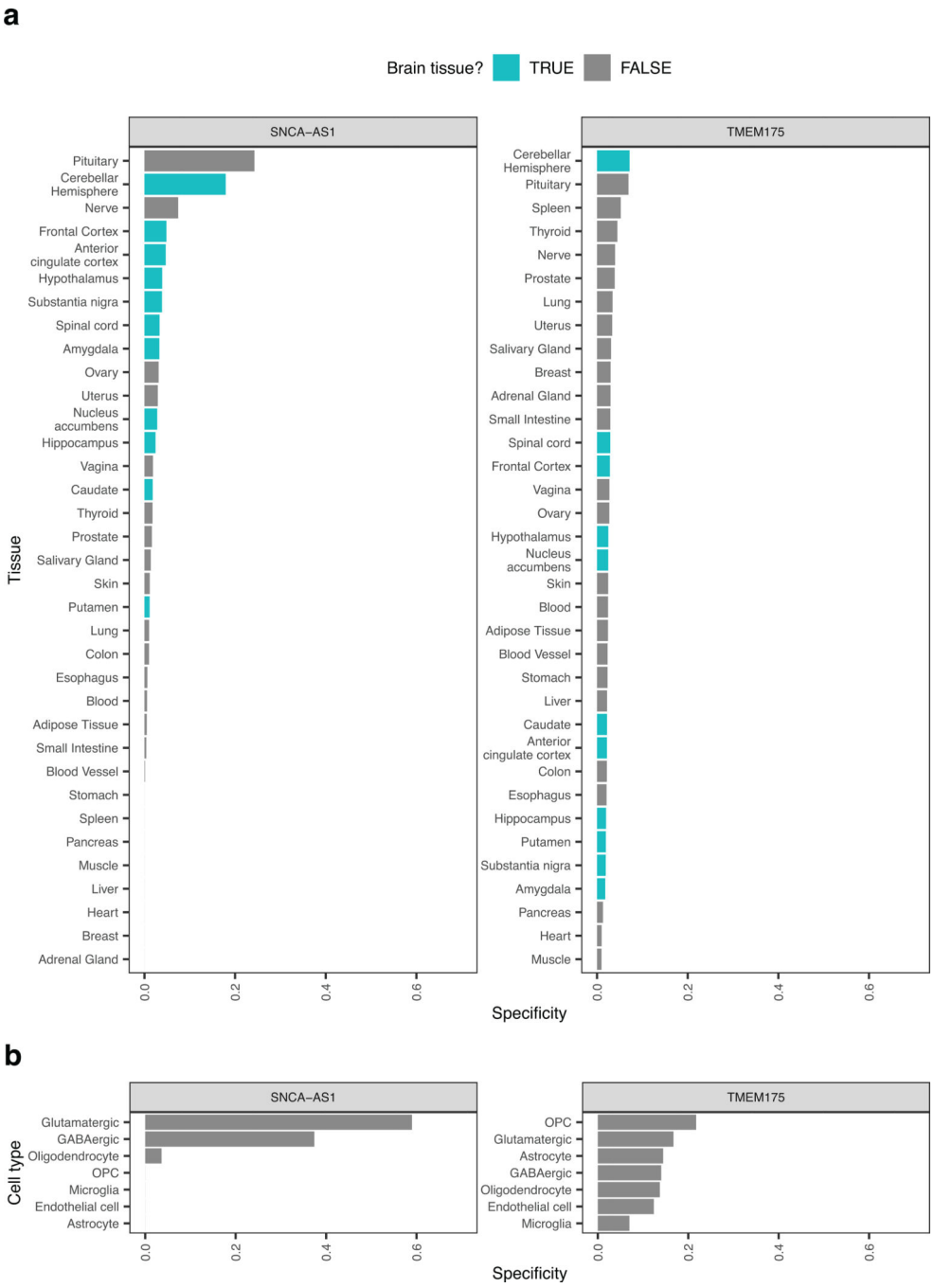
Extended Data Fig. 4. Sensitivity analyses

a,b, Sensitivity analyses of colocalization between eQTLs regulating *TMEM175* expression and LBD GWAS signals (**a**) and *SNCA-AS1* expression and LBD GWAS signals (**b**). eQTLs for *TMEM175* were derived from eQTL-Gen, while eQTLs for *SNCA-AS1* were derived from PsychENCODE. Plots of prior (left) and posterior (right) probabilities for H_0-H_4 hypotheses across varying p_{12} priors are shown. A dashed vertical line indicates the value of p_{12} used in the initial analysis ($p_{12} = 5 \times 10^{-6}$). The green shaded areas in these plots show the regions for which the posterior probability of H4 ≥ 0.90 would still be supported. Abbreviations: H0, hypothesis 0 (no association with either trait); H1, hypothesis 1 (association with trait 1, not with trait 2); H2, hypothesis 2 (association with trait 2, not with trait 1); H3, hypothesis 3 (association with trait 1 and trait 2, two independent SNPs); H4, hypothesis 4 (association with trait 1 and trait 2, one shared SNP).



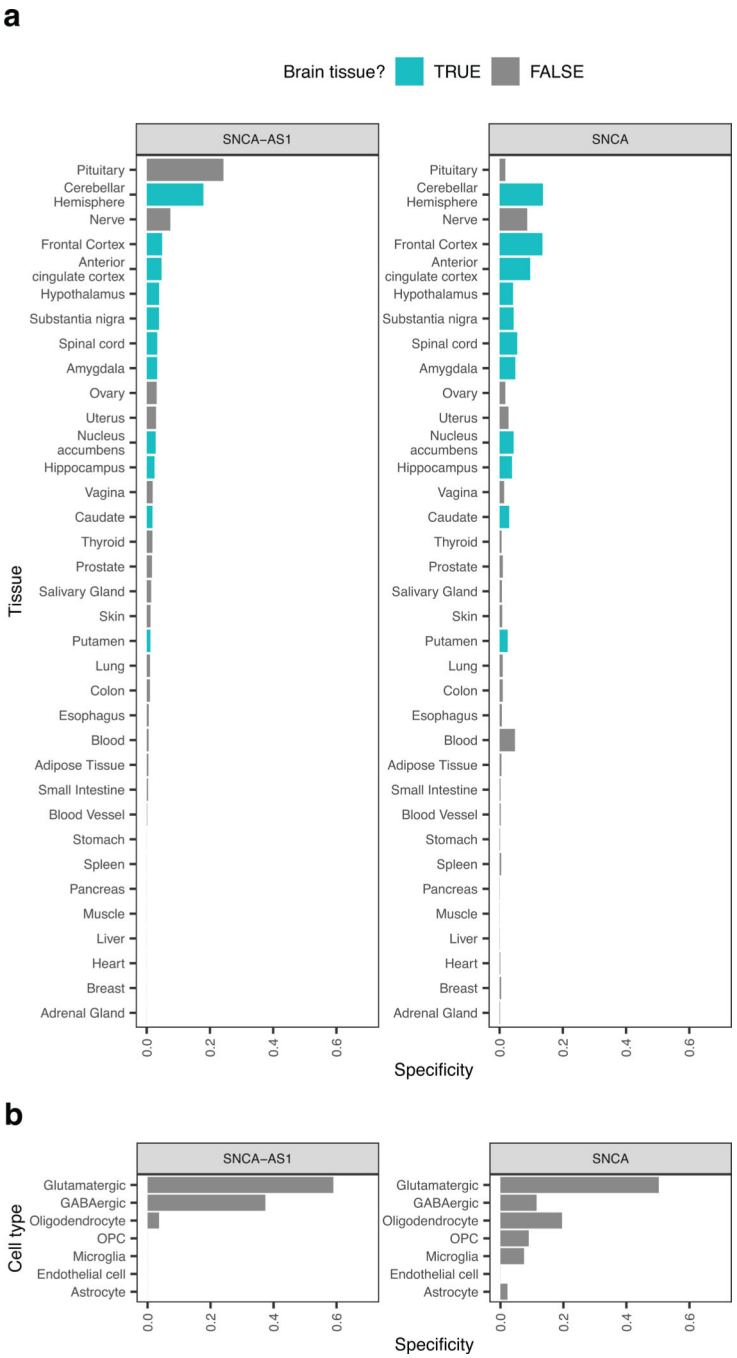
Extended Data Fig. 5. GWAS variants correlate with increased *SNCA-AS1* expression

Shown here are genome-wide significant SNPs that decrease risk for LBD and their correlation with increased *SNCA-AS1* expression. **a,** Scatterplot of beta coefficients and association *P*-values (on a -log10 scale) for SNPs shared between the LBD GWAS (left) and PsychENCODE (right). The SNPs represented in this plot are those that are eQTLs regulating *SNCA-AS1* expression. The top SNP in the LBD GWAS (as determined by the lowest association test *P*-value) is indicated in both scatterplots by a red point. The dashed line represents the cut-off for genome-wide significance (5×10^{-8}). **b,** Scatterplot of SNPs shared between the LBD GWAS and PsychENCODE, which pass genome-wide significance in the LBD GWAS. Spearman's rho (R) and associated *P*-value are displayed.



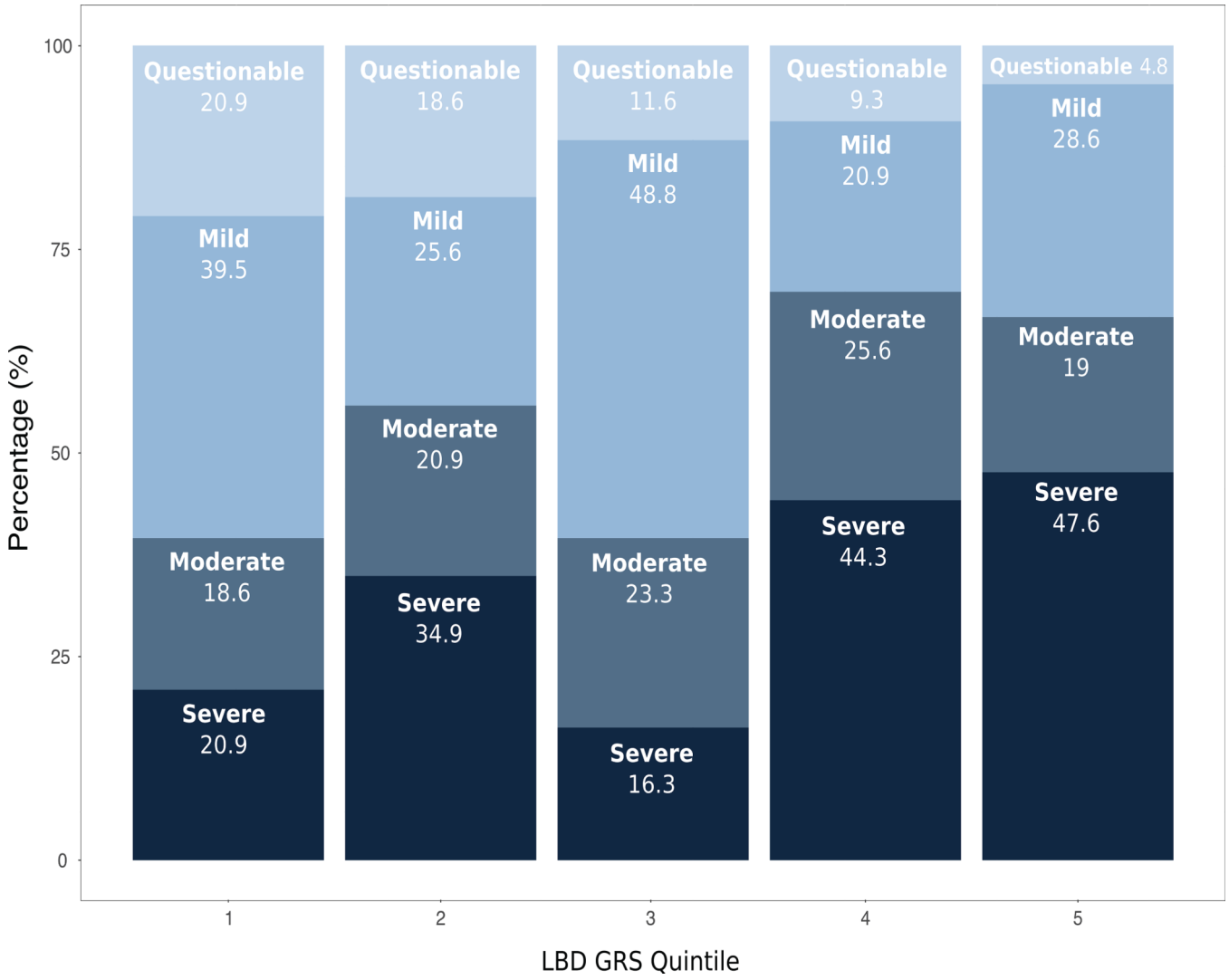
Extended Data Fig. 6. Tissue and cell-type specificity of *SNCA-AS1* and *TMEM175*

a,b, Plot of *SNCA-AS1* and *TMEM175* specificity in 35 human tissues (GTEx dataset) (**a**) and seven broad categories of cell types derived from human middle temporal gyrus (Allen Institute for Brain Science dataset) (**b**). Tissues are colored by whether they belong to the brain. In all plots, tissues and cell types have been ordered by specificity.



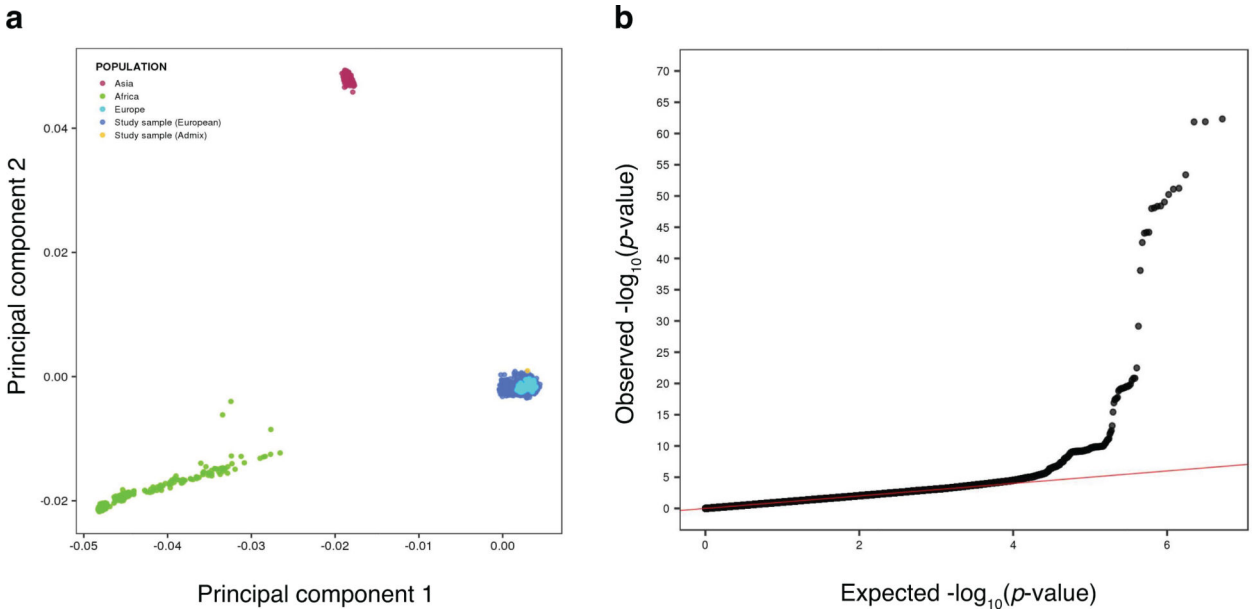
Extended Data Fig. 7. Tissue and cell-specificity of *SNCA-AS1* and *SNCA*

a,b, Plots of *SNCA-AS1* and *SNCA* specificity in 35 human tissues (GTEx dataset) (**a**) and seven broad categories of cell types derived from human middle temporal gyrus (Allen Institute for Brain Science dataset) (**b**). Tissues are colored by whether they belong to the brain. In all plots, tissues and cell types have been ordered by specificity.



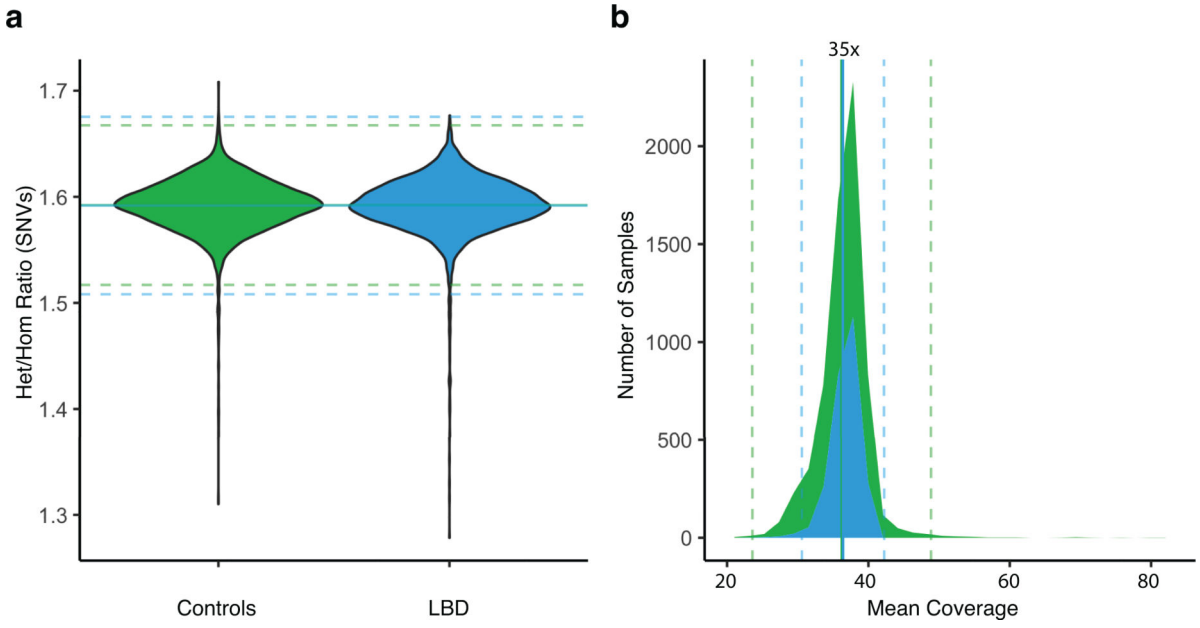
Extended Data Fig. 8. LBD polygenic risk score is associated with dementia severity

Dementia severity score proportions (measured by the Clinical Dementia Rating scale) at baseline evaluation relative to LBD polygenic risk score quintiles. LBD patients in the highest quintile had significantly more severe cognitive impairment at baseline compared to cases in the lowest quintile ($\chi^2 = 5.60$, df = 1, test *P*-value = 0.009).



Extended Data Fig. 9. Principal components analysis and QQ plot

Quality control metrics of GWAS data. **a,** Population structure is shown by plotting the first two principal components of the study cohorts (n = 2,591 LBD cases and n = 4,027 controls) compared to the HapMap3 Genome Reference panel. **b,** Quantile-quantile (QQ) plot of single-variant associations depicting observed (y-axis) versus expected P-values (x-axis). The sample size adjusted genomic inflation factor λ1000 was 1.004.



Extended Data Fig. 10. Quality control metrics

This figure depicts quality control metrics of the genome data across study cohorts. **a**, Heterozygous-to-homozygous single nucleotide variant (SNV) ratios. **b**, Mean coverage across the study cohorts.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Authors

Ruth Chia[1,130], Marya S. Sabir[2,130], Sara Bandres-Ciga[3], Sara Saez-Atienzar[1], Regina H. Reynolds[4,5,6], Emil Gustavsson[5,6], Ronald L. Walton[7], Sarah Ahmed[2], Coralie Viollet[8,9], Jinhui Ding[10], Mary B. Makarious[2], Monica Diez-Fairen[11], Makayla K. Portley[2], Zalak Shah[2], Yevgeniya Abramzon[1,12], Dena G. Hernandez[3], Cornelis Blauwendraat[3], David J. Stone[13], John Eicher[14], Laura Parkkinen[15], Olaf Ansorge[15], Lorraine Clark[16], Lawrence S. Honig[17], Karen Marder[17], Afina Lemstra[18], Peter St George-Hyslop[19,20], Elisabet Londos[21], Kevin Morgan[22], Tammaryn Lashley[4,23], Thomas T. Warner[12,23], Zane Jaunmuktane[23], Douglas Galasko[24,25], Isabel Santana[26,27,28,29], Pentti J. Tienari[30,31], Liisa Myllykangas[32,33], Minna Oinas[34], Nigel J. Cairns[35], John C. Morris[35], Glenda M. Halliday[36,37,38], Vivianna M. Van Deerlin[39], John Q. Trojanowski[39], Maurizio Grassano[1,40], Andrea Calvo[40,41], Gabriele Mora[42], Antonio Canosa[40,41], Gianluca Floris[43], Ryan C. Bohannan[44], Francesca Brett[45], Ziv Gan-Or[46], Joshua T. Geiger[2], Anni Moore[10], Patrick May[47], Rejko Krüger[47,48,49], David S. Goldstein[50], Grisel Lopez[51], Nahid Tayebi[51], Ellen Sidransky[51], The American Genome Center, Lucy Norcliffe-Kaufmann[52], Jose-Alberto Palma[52], Horacio Kaufmann[52], Vikram G. Shakkottai[53], Matthew Perkins[54], Kathy L. Newell[55], Thomas Gasser[56], Claudia Schulte[56], Francesco Landi[57], Erika Salvi[58], Daniele Cusi[59], Eliezer Masliah[60], Ronald C. Kim[61], Chad A. Caraway[62], Edwin S. Monuki[63], Maura Brunetti[40], Ted M. Dawson[64,65,66,67], Liana S. Rosenthal[64], Marilyn S. Albert[64], Olga Pletnikova[68,69], Juan C. Troncoso[68], Margaret E. Flanagan[70,71], Qinwen Mao[70,71], Eileen H. Bigio[70,71], Eloy Rodríguez-Rodríguez[72], Jon Infante[72], Carmen Lage[72], Isabel González-Aramburu[72], Pascual Sanchez-Juan[72], Bernardino Ghetti[55], Julia Keith[73], Sandra E. Black[74,75,76,77,78], Mario Masellis[77,78,79,80], Ekaterina Rogaeva[81], Charles Duyckaerts[82,83], Alexis Brice[83], Suzanne Lesage[83], Georgia Xiromerisiou[84], Matthew J. Barrett[85], Bension S. Tilley[86], Steve Gentleman[86], Giancarlo Logroscino[87,88], Geidy E. Serrano[89], Thomas G. Beach[89], Ian G. McKeith[90], Alan J. Thomas[90], Johannes Attems[90], Christopher M. Morris[90], Laura Palmer[91], Seth Love[92], Claire Troakes[93], Safa Al-Sarraj[94], Angela K. Hodges[93], Dag Aarsland[93,95], Gregory Klein[96], Scott M. Kaiser[97], Randy Woltjer[98], Pau Pastor[11], Lynn M. Bekris[99], James B. Leverenz[100], Lilah M. Besser[101], Amanda Kuzma[102], Alan E. Renton[103], Alison Goate[104], David A. Bennett[96], Clemens R. Scherzer[105], Huw R. Morris[106], Raffaele Ferrari[4], Diego Albani[107], Stuart Pickering-Brown[108], Kelley Faber[109], Walter A. Kukull[110], Estrella Morenas-Rodriguez[111,112,113], Alberto Lleó[112,113], Juan Fortea[112,113], Daniel Alcolea[112,113], Jordi Clarimon[112,113], Mike A. Nalls[114,115,116], Luigi Ferrucci[117], Susan M. Resnick[118], Toshiko Tanaka[117], Tatiana M. Foroud[109], Neill R. Graff-Radford[119], Zbigniew K. Wszolek[119], Tanis Ferman[120], Bradley F. Boeve[121], John A. Hardy[4,12,122,123,124], Eric J. Topol[125], Ali Torkamani[125], Andrew B. Singleton[3,116], Mina Ryten[5,6], Dennis W. Dickson[7], Adriano Chiò[40,41,126,131], Owen A.

Ross[7,127,131], J. Raphael Gibbs[10,131], Clifton L. Dalgard[128,129,131], Bryan J. Traynor[1,12,64,131], Sonja W. Scholz[2,64,131,*]

Affiliations

[1]Neuromuscular Diseases Research Section, Laboratory of Neurogenetics, National Institute on Aging, Bethesda, MD, USA. [2]Neurodegenerative Diseases Research Unit, Laboratory of Neurogenetics, National Institute of Neurological Disorders and Stroke, Bethesda, MD, USA. [3]Molecular Genetics Section, Laboratory of Neurogenetics, National Institute on Aging, Bethesda, MD, USA. [4]Department of Neurodegenerative Disease, UCL Queen Square Institute of Neurology, University College London, London, UK. [5]NIHR Great Ormond Street Hospital Biomedical Research Centre, University College London, London, UK. [6]Great Ormond Street Institute of Child Health, Genetics and Genomic Medicine, University College London, London, UK. [7]Department of Neuroscience, Mayo Clinic, Jacksonville, FL, USA. [8]Collaborative Health Initiative Research Program, Uniformed Services University of the Health Sciences, Bethesda, MD, USA. [9]Henry M. Jackson Foundation for the Advancement of Military Medicine, Inc., Bethesda, MD, USA. [10]Computational Biology Group, Laboratory of Neurogenetics, National Institute on Aging, Bethesda, MD, USA. [11]Memory and Movement Disorders Units, Department of Neurology, University Hospital Mutua de Terrassa, Barcelona, Spain. [12]Reta Lila Weston Institute, UCL Queen Square Institute of Neurology, University College London, London, UK. [13]Cerevel Therapeutics, Boston, MA, USA. [14]Genetics and Pharmacogenomics, Merck & Co., Inc., West Point, PA, USA. [15]Nuffield Department of Clinical Neurosciences, Oxford Parkinson's Disease Centre, University of Oxford, Oxford, UK. [16]Taub Institute for Alzheimer Disease and the Aging Brain, and Department of Pathology and Cell Biology, Columbia University, New York, NY, USA. [17]Taub Institute for Alzheimer Disease and the Aging Brain, G. H. Sergievsky Center, and Department of Neurology, Columbia University, New York, NY, USA. [18]Department of Neurology and Alzheimer Center, Neuroscience Campus Amsterdam, Amsterdam, The Netherlands. [19]Department of Clinical Neurosciences, Cambridge Institute of Medical Research, University of Cambridge, Cambridge, UK. [20]Department of Medicine, University of Toronto, Toronto, Ontario, Canada. [21]Clinical Memory Research Unit, Institution of Clinical Sciences Malmo, Lund University, Lund, Sweden. [22]Human Genetics, School of Life Sciences, Queens Medical Centre, University of Nottingham, Nottingham, UK. [23]Queen Square Brain Bank for Neurological Disorders, Department of Clinical and Movement Neurosciences, UCL Queen Square Institute of Neurology, University College London, London, UK. [24]Department of Neurosciences, University of California, San Diego, La Jolla, CA, USA. [25]Veterans Affairs San Diego Healthcare System, La Jolla, CA, USA. [26]Neurology Service, University of Coimbra Hospital, Coimbra, Portugal. [27]Centro Hospitalar e Universitário de Coimbra (CHUC), Coimbra, Portugal. [28]Faculty of Medicine, University of Coimbra, Coimbra, Portugal. [29]Center for Innovative Biomedicine and Biotechnology (CIBB), University of Coimbra, Coimbra, Portugal. [30]Translational Immunology, Research Programs Unit, University of Helsinki, Helsinki, Finland. [31]Department of Neurology, Helsinki University

Hospital, Helsinki, Finland. [32]Department of Pathology, Medicum, University of Helsinki, Helsinki, Finland. [33]HUS Diagnostic Center, Helsinki University Hospital, Helsinki, Finland. [34]Department of Clinical Medicine, Faculty of Health, UiT The Arctic University of Norway, Tromsø, Norway. [35]Knight Alzheimer's Disease Research Center, Department of Neurology, Washington University School of Medicine, Saint Louis, MO, USA. [36]Neuroscience Research Australia, Sydney, New South Wales, Australia. [37]School of Medical Sciences, Faculty of Medicine, University of New South Wales, Sydney, New South Wales, Australia. [38]Brain and Mind Centre, Sydney Medical School, University of Sydney, Sydney, New South Wales, Australia. [39]Department of Pathology and Laboratory Medicine, Center for Neurodegenerative Disease Research, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. [40]"Rita Levi Montalcini" Department of Neuroscience, University of Turin, Turin, Italy. [41]Azienda Ospedaliero Universitaria Città della Salute e della Scienza, Turin, Italy. [42]Istituti Clinici Scientifici Maugeri, IRCCS, Milan, Italy. [43]Department of Neurology, University Hospital of Cagliari, Cagliari, Italy. [44]Department of Neurobiology and Behavior, University of California Irvine, Irvine, CA, USA. [45]Dublin Brain Bank, Neuropathology Department, Beaumont Hospital, Dublin, Ireland. [46]Montreal Neurological Institute and Hospital, Department of Neurology & Neurosurgery, McGill University, Montreal, Quebec, Canada. [47]Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg. [48]Transversal Translational Medicine, Luxembourg Institute of Health (LIH), Strassen, Luxembourg. [49]Parkinson Research Clinic, Centre Hospitalier de Luxembourg (CHL), Luxembourg City, Luxembourg. [50]Clinical Neurocardiology Section, National Institute of Neurological Disorders and Stroke, Bethesda, MD, USA. [51]Medical Genetics Branch, National Human Genome Research Institute, Bethesda, MD, USA. [52]Department of Neurology, New York University School of Medicine, New York, NY, USA. [53]Department of Neurology, University of Michigan Medical School, Ann Arbor, MI, USA. [54]Michigan Brain Bank, University of Michigan Medical School, Ann Arbor, MI, USA. [55]Department of Pathology and Laboratory Medicine, Indiana University School of Medicine, Indianapolis, IN, USA. [56]Department of Neurodegenerative Diseases, Center of Neurology and Hertie-Institute for Clinical Brain Research, University of Tübingen and German Center for Neurodegenerative Diseases, Tübingen, Germany. [57]Fondazione Policlinico Universitario Agostino Gemelli IRCCS Università Cattolica del Sacro Cuore, Rome, Italy. [58]Neuroalgology Unit, Fondazione IRCCS Istituto Neurologico Carlo Besta, Milan, Italy. [59]Bio4Dreams-Business Nursery for Life, Milan, Italy. [60]Molecular Neuropathology Section, Laboratory of Neurogenetics, National Institute on Aging, Bethesda, MD, USA. [61]Department of Neuropathology, School of Medicine, University of California Irvine, Irvine, CA, USA. [62]Institute for Memory Impairments and Neurological Disorders, University of California Irvine, Irvine, CA, USA. [63]Department of Pathology & Laboratory Medicine, School of Medicine, University of California Irvine, Irvine, CA, USA. [64]Department of Neurology, Johns Hopkins University Medical Center, Baltimore, MD, USA. [65]Neuroregeneration and Stem Cell Programs, Institute of Cell Engineering, Johns

Hopkins University Medical Center, Baltimore, MD, USA. [66]Department of Pharmacology and Molecular Science, Johns Hopkins University Medical Center, Baltimore, MD, USA. [67]Solomon H. Snyder Department of Neuroscience, Johns Hopkins University Medical Center, Baltimore, MD, USA. [68]Department of Pathology (Neuropathology), Johns Hopkins University Medical Center, Baltimore, MD, USA. [69]Department of Pathology and Anatomical Sciences, Jacobs School of Medicine and Biomedical Sciences, University at Buffalo, Buffalo, NY, USA. [70]Mesulam Center for Cognitive Neurology and Alzheimer's Disease, Northwestern University Feinberg School of Medicine, Chicago, IL, USA. [71]Department of Pathology, Northwestern University Feinberg School of Medicine, IL, USA. [72]Neurology Service, University Hospital Marqués de Valdecilla-IDIVAL-UC-CIBERNED, Santander, Spain. [73]Department of Anatomical Pathology, Sunnybrook Health Sciences Centre, University of Toronto, Toronto, Ontario, Canada. [74]Institute of Medical Science, Faculty of Medicine, University of Toronto, Toronto, Ontario, Canada. [75]Division of Neurology, Department of Medicine, University of Toronto, Toronto, Ontario, Canada. [76]Heart and Stroke Foundation Canadian Partnership for Stroke Recovery, Sunnybrook Health Sciences Centre, University of Toronto, Toronto, Ontario, Canada. [77]Hurvitz Brain Sciences Research Program, Sunnybrook Research Institute, University of Toronto, Toronto, Ontario, Canada. [78]LC Campbell Cognitive Neurology Research Unit, Sunnybrook Research Institute, University of Toronto, Toronto, Ontario, Canada. [79]Cognitive & Movement Disorders Clinic, Sunnybrook Health Sciences Centre, University of Toronto, Toronto, Ontario, Canada. [80]Department of Medicine, Division of Neurology, University of Toronto, Toronto, Ontario, Canada. [81]Tanz Centre for Research in Neurodegenerative Diseases, University of Toronto, Toronto, Ontario, Canada. [82]Department of Neuropathology Escourolle, Paris Brain Institute, Sorbonne Universites, Paris, France. [83]Sorbonne Université, Institut du Cerveau - Paris Brain Institute - ICM, Inserm, CNRS, AP-HP, Hôpital Pitié-Salpêtrière, DMU Neuroscience 6, Paris, France. [84]Department of Neurology, University Hospital of Larissa, University of Thessalia, Larissa, Greece. [85]Department of Neurology, Virginia Commonwealth University, Richmond, VA, USA. [86]Neuropathology Unit, Department of Brain Sciences, Imperial College London, London, UK. [87]Department of Basic Medicine Neurosciences and Sense Organs, University Aldo Moro, Bari, Italy. [88]Center for Neurodegenerative Diseases and the Aging Brain - Department of Clinical Research in Neurology of the University of Bari at "Pia Fondazione Card G. Panico" Hospital Tricase (Le), Bari, Italy. [89]Civin Laboratory for Neuropathology, Banner Sun Health Research Institute, Sun City, AZ, USA. [90]Newcastle Brain Tissue Resource, Translational and Clinical Research Institute, Biomedical Research Building, Newcastle University, Newcastle upon Tyne, UK. [91]South West Dementia Brain Bank, Bristol Medical School, University of Bristol, Bristol, UK. [92]Dementia Research Group, Bristol Medical School, University of Bristol, Bristol, UK. [93]Institute of Psychiatry, Psychology and Neuroscience (IoPPN), King's College London, London, UK. [94]Department of Clinical Neuropathology and London Neurodegenerative Diseases Brain Bank, Institute of Psychiatry, Psychology and Neuroscience (IoPPN), King's College

Hospital and King's College London, London, UK. [95]Centre for Age-Related Medicine, Stavanger University Hospital, Stavanger, Norway. [96]Rush Alzheimer's Disease Center, Rush University, Chicago, IL, USA. [97]Department of Neuropathology, Indiana University School of Medicine, Indianapolis, IN, USA. [98]Department of Neurology, Oregon Health & Sciences University, Portland, OR, USA. [99]Genomic Medicine Institute, Cleveland Clinic, Cleveland, OH, USA. [100]Cleveland Lou Ruvo Center for Brain Health, Neurological Institute, Cleveland Clinic, Cleveland, OH, USA. [101]Institute for Human Health and Disease Intervention, Florida Atlantic University, Boca Raton, FL, USA. [102]Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. [103]Ronald M. Loeb Center for Alzheimer's Disease, Nash Family Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [104]Ronald M. Loeb Center for Alzheimer's Disease, Nash Family Department of Neuroscience, Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [105]Precision Neurology Program, Brigham & Women's Hospital, Harvard Medical School, Boston, MA, USA. [106]Department of Clinical and Movement Neuroscience, UCL Queen Square Institute of Neurology, University College London, London, UK. [107]Department of Neuroscience, Istituto di Ricerche Farmacologiche Mario Negri IRCCS, Milan, Italy. [108]Division of Neuroscience and Experimental Psychology, Faculty of Biology, Medicine and Health, The University of Manchester, Manchester, UK. [109]Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN, USA. [110]National Alzheimer's Coordinating Center (NACC), Department of Epidemiology, University of Washington, Seattle, WA, USA. [111]Biomedizinisches Centrum (BMC), Biochemie, Ludwig-Maximilians-Universität München & Deutsches Zentrum für Neurodegenerative Erkrankungen (DZNE), Munich, Germany. [112]Sant Pau Biomedical Research Institute, Hospital de la Santa Creu i Sant Pau, Universitat Autònoma de Barcelona, Barcelona, Spain. [113]The Network Center for Biomedical Research in Neurodegenerative Diseases (CIBERNED), Madrid, Spain. [114]Laboratory of Neurogenetics, National Institute on Aging, Bethesda, MD, USA. [115]Data Tecnica International, Glen Echo, MD, USA. [116]Center for Alzheimer's and Related Dementias, National Institute on Aging, Bethesda, MD, USA. [117]Longitudinal Studies Section, National Institute on Aging, Baltimore, MD, USA. [118]Laboratory of Behavioral Neuroscience, National Institute on Aging, Baltimore, MD, USA. [119]Department of Neurology, Mayo Clinic, Jacksonville, FL, USA. [120]Department of Psychiatry and Psychology, Mayo Clinic, Jacksonville, FL, USA. [121]Department of Neurology, Mayo Clinic, Rochester, MN, USA. [122]UK Dementia Research Institute of UCL, UCL Institute of Neurology, University College London, London, UK. [123]UCL Movement Disorders Centre, University College London, London, UK. [124]Institute for Advanced Study, The Hong Kong University of Science and Technology, Hong Kong SAR, China. [125]Scripps Research Translational Institute, Scripps Research, La Jolla, CA, USA. [126]Institute of Cognitive Sciences and Technologies, C.N.R., Rome, Italy. [127]Department of Clinical Genomics, Mayo Clinic, Jacksonville, FL, USA. [128]Department of Anatomy,

Physiology & Genetics, Uniformed Services University of the Health Sciences, Bethesda, MD, USA. [129]The American Genome Center, Uniformed Services University of the Health Sciences, Bethesda, MD, USA. [130]These authors contributed equally to this work. [131]These authors jointly supervised this work.

Acknowledgments

We thank contributors who collected samples used in this study, as well as patients and families, whose help and participation made this work possible. This research was supported in part by the Intramural Research Program of the National Institutes of Health (National Institute on Aging, National Institute of Neurological Disorders and Stroke; project numbers: 1ZIAAG000935 [PI Bryan J. Traynor], 1ZIANS003154 [PI Sonja W. Scholz], 1ZIANS0030033 and 1ZIANS003034 [David S. Goldstein]). This study used the high-performance computational capabilities of the Biowulf Linux cluster at the National Institutes of Health, Bethesda, MD, USA (http://biowulf.nih.gov). A complete list of acknowledgments is listed in the Supplementary Note.

COMPETING INTERESTS

T.G.B. is a consultant for Prothena Biosciences, Vivid Genomics and Avid Radiopharmaceutical, and is a scientific advisory board member for Vivid Genomics. J.A.H., H.R.M., S.P.-B., P.J.T. and B.J.T. hold US, EU and Canadian patents on the clinical testing and therapeutic intervention for the hexanucleotide repeat expansion of *C9orf72*. H.R.M. reports paid consultancy from Biogen, Biohaven, Lundbeck, UCB, Denali, and lecture fees/honoraria from the Wellcome Trust, and Movement Disorders Society. H.R.M. received research grants from Parkinson's UK, Cure Parkinson's Trust, PSP Association, CBD Solutions, Drake Foundation, and the Medical Research Council. H.R.M. is a co-applicant on a patent application related to C9orf72 – Method for diagnosing a neurodegenerative disease (PCT/GB2012/052140). J.E. was an employee of a for-profit company (Merck) at the time of the collaboration. M.A.N.'s participation is supported by a consulting contract between Data Tecnica International and the National Institute on Aging, NIH, Bethesda, MD, USA; as a possible conflict of interest, M.A.N. also consults for Neuron23 Inc., Lysosomal Therapeutics Inc., Illumina Inc., the Michael J. Fox Foundation and Vivid Genomics among others. A.B.S. is an associate editor for the journals *Brain*, *Movement Disorders*, and *npj Parkinson's Disease*. H.K. is Editor-in-Chief of *Clinical Autonomic Research*, serves as PI of a clinical trial sponsored by Biogen MA Inc. (TRACK MSA, S19–01846), received consultancy fees from Lilly USA LLC, Biohaven Pharmaceuticals Inc, Takeda Pharmaceutical Company Ltd, Ono Pharma UK Ltd, Lundbeck LLC, and Theravance Biopharma US Inc. J.A.P. is an editorial board member of *Movement Disorders*, *Parkinsonism & Related Disorders*, *BMC Neurology*, and *Clinical Autonomic Research*. B.F.B., J.B.L., and S.W.S. serve on the Scientific Advisory Council of the Lewy Body Dementia Association. S.W.S. is an editorial board member for the *Journal of Parkinson's Disease*, and *JAMA Neurology*. B.J.T. is an editorial board member for *JAMA Neurology*, *Journal of Neurology, Neurosurgery, and Psychiatry*, *Brain*, and *Neurobiology of Aging*. D.A. is associate editor of the *Journal of Alzheimer's Disease*. R.K. is Coordinator of the National Centre for Excellence in Research on Parkinson's disease (NCER-PD) and received speaker's honoraria and/or travel grants from Abbvie, Zambon and Medtronic, and he participated as PI or site-PI for industry sponsored clinical trials without receiving honoraria. Z.K.W. serves as a principal investigator or co-principal investigator on Biogen, Inc. (228PD201), Biohaven Pharmaceuticals, Inc. (BHV4157–206 and BHV3241–301), and Neuraly, Inc. (NLY01-PD-1) grants. Z.K.W. also serves as the co-principal investigator of the Mayo Clinic Florida American Parkinson Disease Association Center for Advanced Research. Z.G.-O. received consultancy fees from Lysosomal Therapeutics Inc. (LTI), Idorsia, Prevail Therapeutics, Inceptions Sciences (now Ventus), Ono Therapeutics, Denali, Deerfield, Neuron23, and Handle Therapeutics. Z.G.-O. is an Associate Editor of the *Journal of Parkinson's Disease* and editorial board member in *Parkinsonism and Related Disorders*. A.T. serves on the scientific advisory board for Vivid Genomics. All other authors report no competing interests.

REFERENCES

1. Walker Z, Possin KL, Boeve BF & Aarsland D Lewy body dementias. Lancet 386, 1683–1697 (2015). [PubMed: 26595642]
2. McKeith IG et al. Diagnosis and management of dementia with Lewy bodies: Fourth consensus report of the DLB Consortium. Neurology 89, 88–100 (2017). [PubMed: 28592453]
3. Meeus B, Theuns J & Van Broeckhoven C The genetics of dementia with Lewy bodies: what are we missing? Arch. Neurol. 69, 1113–1118 (2012). [PubMed: 22635379]
4. https://www.lbda.org/page/what-lbd.

5. Guerreiro R et al. Investigating the genetic architecture of dementia with Lewy bodies: a two-stage genome-wide association study. Lancet Neurol. 17, 64–74 (2018). [PubMed: 29263008]
6. Sabir MS et al. Assessment of APOE in atypical parkinsonism syndromes. Neurobiol. Dis 127, 142–146 (2019). [PubMed: 30798004]
7. Nalls MA et al. A multicenter study of glucocerebrosidase mutations in dementia with Lewy bodies. JAMA Neurol. 70, 727–735 (2013). [PubMed: 23588557]
8. Singleton AB et al. alpha-Synuclein locus triplication causes Parkinson's disease. Science 302, 841 (2003). [PubMed: 14593171]
9. Tsuang D et al. APOE epsilon4 increases risk for dementia in pure synucleinopathies. JAMA Neurol 70, 223–228 (2013). [PubMed: 23407718]
10. Pickering-Brown SM et al. Apolipoprotein E4 and Alzheimer's disease pathology in Lewy body disease and in other beta-amyloid-forming diseases. Lancet 343, 1155 (1994).
11. Seshadri S et al. Genome-wide analysis of genetic loci associated with Alzheimer disease. JAMA 303, 1832–1840 (2010). [PubMed: 20460622]
12. Pankratz N et al. Genomewide association study for susceptibility genes contributing to familial Parkinson disease. Hum. Genet 124, 593–605 (2009). [PubMed: 18985386]
13. Lee S et al. Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. Am. J. Hum. Genet 91, 224–237 (2012). [PubMed: 22863193]
14. Geiger JT et al. Next-generation sequencing reveals substantial genetic contribution to dementia with Lewy bodies. Neurobiol Dis. 94, 55–62 (2016). [PubMed: 27312774]
15. Singleton A & Hardy J A generalizable hypothesis for the genetic architecture of disease: pleomorphic risk loci. Hum. Mol. Genet 20, R158–R162 (2011). [PubMed: 21875901]
16. Nicolae DL et al. Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. PLoS Genet. 6, e1000888 (2010). [PubMed: 20369019]
17. Li YI et al. RNA splicing is a primary link between genetic variation and disease. Science 352, 600–604 (2016). [PubMed: 27126046]
18. Võsa U et al. Unraveling the polygenic architecture of complex traits using blood eQTL metaanalysis. Preprint at https://www.biorxiv.org/content/10.1101/447367v1 (2018).
19. Wang D et al. Comprehensive functional genomic resource and integrative model for the human brain. Science 362, eaat8464 (2018). [PubMed: 30545857]
20. Hawrylycz MJ et al. An anatomically comprehensive atlas of the adult human brain transcriptome. Nature 489, 391–399 (2012). [PubMed: 22996553]
21. GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science 348, 648–660 (2015). [PubMed: 25954001]
22. Kunkle BW et al. Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Abeta, tau, immunity and lipid processing. Nat. Genet 51, 414–430 (2019). [PubMed: 30820047]
23. Nalls MA et al. Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. Lancet Neurol. 18, 1091–1102 (2019). [PubMed: 31701892]
24. Kramarz B et al. Improving the Gene Ontology resource to facilitate more informative analysis and interpretation of Alzheimer's disease data. Genes (Basel) 9, 593 (2018).
25. Bandres-Ciga S et al. The endocytic membrane trafficking pathway plays a major role in the risk of Parkinson's disease. Mov. Disord 34, 460–468 (2019). [PubMed: 30675927]
26. Morris JC The Clinical Dementia Rating (CDR): current version and scoring rules. Neurology 43, 2412–2414 (1993).
27. Miyagawa T et al. BIN1 regulates BACE1 intracellular trafficking and amyloid-beta production. Hum. Mol. Genet 25, 2948–2958 (2016). [PubMed: 27179792]
28. Calafate S, Flavin W, Verstreken P & Moechars D Loss of Bin1 promotes the propagation of Tau pathology. Cell Rep. 17, 931–940 (2016). [PubMed: 27760323]

29. Jinn S et al. TMEM175 deficiency impairs lysosomal and mitochondrial function and increases alpha-synuclein aggregation. Proc. Natl. Acad. Sci. USA 114, 2389–2394 (2017). [PubMed: 28193887]
30. Corder EH et al. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. Science 261, 921–923 (1993). [PubMed: 8346443]
31. Sidransky E et al. Multicenter analysis of glucocerebrosidase mutations in Parkinson's disease. N. Engl. J. Med 361, 1651–1661 (2009). [PubMed: 19846850]
32. Simon-Sanchez J et al. Genome-wide association study reveals genetic risk underlying Parkinson's disease. Nat. Genet 41, 1308–1312 (2009). [PubMed: 19915575]
33. Jansen IE et al. Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. Nat. Genet 51, 404–413 (2019). [PubMed: 30617256]
34. Ross OA et al. Genomic investigation of alpha-synuclein multiplication and parkinsonism. Ann. Neurol 63, 743–750 (2008). [PubMed: 18571778]
35. Pollard MO, Gurdasani D, Mentzer AJ, Porter T & Sandhu MS Long reads: their purpose and place. Hum. Mol. Genet 27, R234–R241 (2018). [PubMed: 29767702]

Methods-only References

36. Erikson GA et al. Whole-genome sequencing of a healthy aging cohort. Cell 165, 1002–1011 (2016). [PubMed: 27114037]
37. Emre M et al. Clinical diagnostic criteria for dementia associated with Parkinson's disease. Mov. Disord 22, 1689–1707 (2007). [PubMed: 17542011]
38. Savica R et al. Incidence of dementia with Lewy bodies and Parkinson disease dementia. JAMA Neurol. 70, 1396–1402 (2013). [PubMed: 24042491]
39. Regier AA et al. Functional equivalence of genome sequencing analysis pipelines enables harmonized variant calling across human genetics projects. Nat. Commun 9, 4038 (2018). [PubMed: 30279509]
40. Van der Auwera GA et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr. Protoc. Bioinformatics 43, 11 10 1–11 10 33 (2013). [PubMed: 25431634]
41. Jun G et al. Detecting and estimating contamination of human DNA samples in sequencing and array-based genotype data. Am. J. Hum. Genet. 91, 839–848 (2012). [PubMed: 23103226]
42. International HapMap Consortium et al. Integrating common and rare genetic variation in diverse human populations. Nature 467, 52–58 (2010). [PubMed: 20811451]
43. Schneider VA et al. Evaluation of GRCh38 and de novo haploid genome assemblies demonstrates the enduring quality of the reference assembly. Genome Res. 27, 849–864 (2017). [PubMed: 28396521]
44. Karczewski KJ et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature 581, 434–443 (2020). [PubMed: 32461654]
45. Chang CC et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 4, 7 (2015). [PubMed: 25722852]
46. Venables WN & Ripley BD Modern Applied Statistics with S, (Springer, New York, 2002).
47. Das S et al. Next-generation genotype imputation service and methods. Nat. Genet. 48, 1284–1287 (2016). [PubMed: 27571263]
48. Fuchsberger C, Abecasis GR & Hinds DA minimac2: faster genotype imputation. Bioinformatics 31, 782–784 (2015). [PubMed: 25338720]
49. Taliun D et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. Preprint at https://www.biorxiv.org/content/10.1101/563866v1 (2019).
50. Willer CJ, Li Y & Abecasis GR METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 26, 2190–2191 (2010). [PubMed: 20616382]
51. Mirra SS et al. The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part II. Standardization of the neuropathologic assessment of Alzheimer's disease. Neurology 41, 479–486 (1991). [PubMed: 2011243]

52. Braak H, Alafuzoff I, Arzberger T, Kretzschmar H & Del Tredici K Staging of Alzheimer disease-associated neurofibrillary pathology using paraffin sections and immunocytochemistry. Acta Neuropathol 112, 389–404 (2006). [PubMed: 16906426]

53. Giambartolomei C et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet. 10, e1004383 (2014). [PubMed: 24830394]

54. Wallace C Eliciting priors and relaxing the single causal variant assumption in colocalisation analyses. PLoS Genet. 16, e1008720 (2020). [PubMed: 32310995]

55. Bryois J et al. Genetic identification of cell types underlying brain complex traits yields insights into the etiology of Parkinson's disease. Nat. Genet 52, 482–493 (2020). [PubMed: 32341526]

56. Skene NG & Grant SG Identification of vulnerable cell types in major brain disorders using single cell transcriptomes and expression weighted cell type enrichment. Front. Neurosci 10, 16 (2016). [PubMed: 26858593]

57. Yang J et al. Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. Nat. Genet 47, 1114–1120 (2015). [PubMed: 26323059]

58. Wang K, Li M & Hakonarson H ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 38, e164 (2010). [PubMed: 20601685]

59. Zhan X, Hu Y, Li B, Abecasis GR & Liu DJ RVTESTS: an efficient and comprehensive tool for rare variant association analysis using sequence data. Bioinformatics 32, 1423–1426 (2016). [PubMed: 27153000]

60. Bohler A et al. Reactome from a WikiPathways perspective. PLoS Comput. Biol 12, e1004941 (2016). [PubMed: 27203685]

61. Kanehisa M, Sato Y, Kawashima M, Furumichi M & Tanabe M KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res. 44, D457–D462 (2016). [PubMed: 26476454]

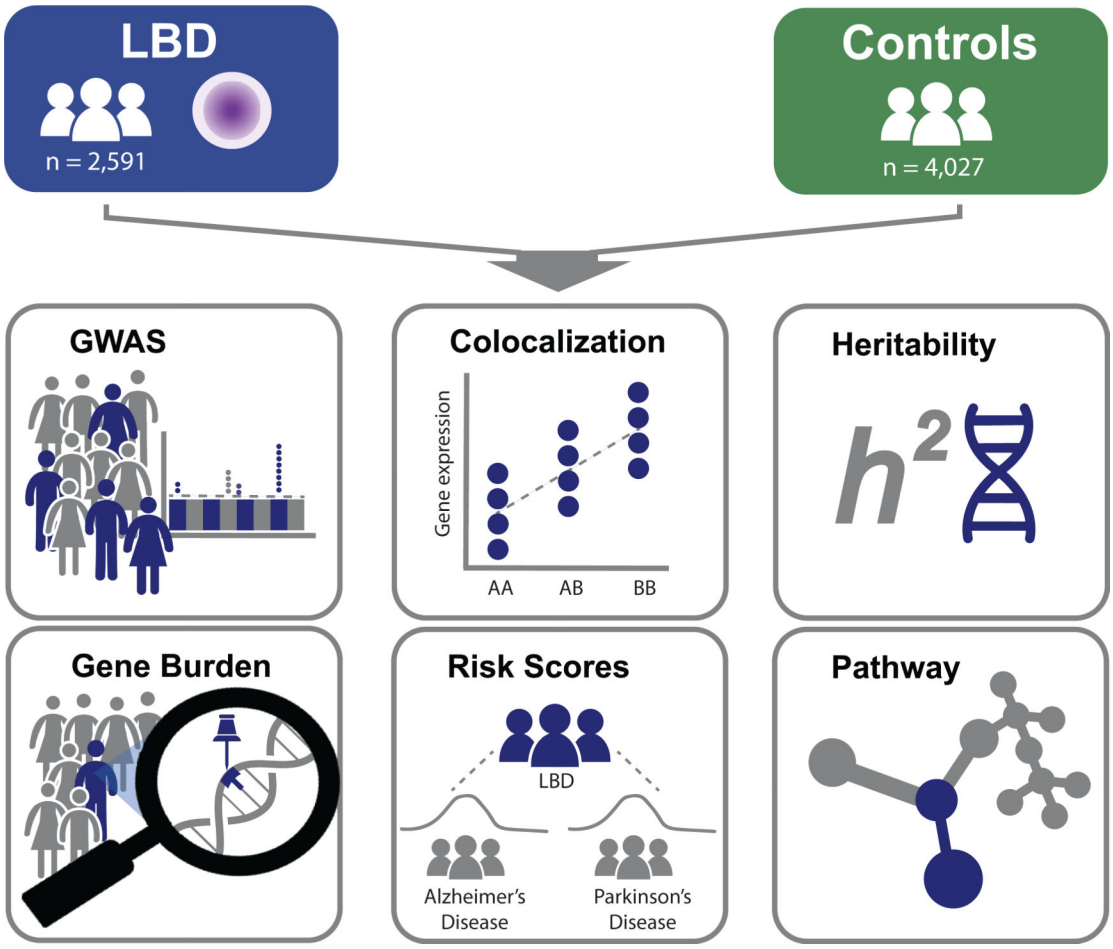


Fig. 1 |. Analysis workflow.
Schematic illustration of the analytical workflow.

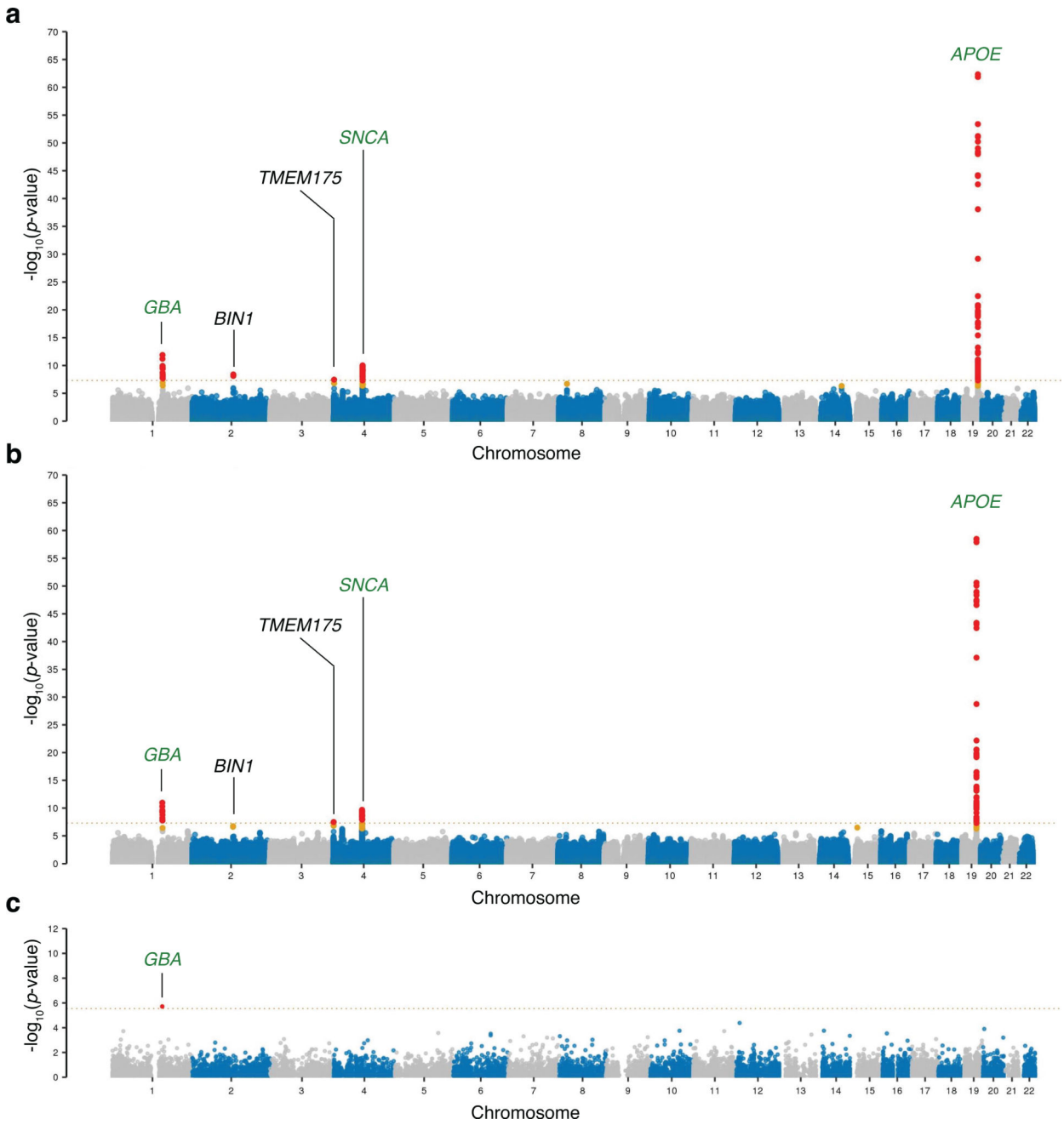


Fig. 2 |. Genome-wide representation of common and rare variant associations in LBD.

a-c, Manhattan plots depicting the GWAS results (n = 2,591 cases and 4,027 controls; MAF > 1%) (**a**), the GWAS subanalysis of pathologically confirmed LBD cases only (n = 1,789) versus controls (n = 4,027) (**b**), and gene-based genome-wide SKAT-O test associations of rare missense variants (MAF ≤ 1%, MAC ≥ 3) (**c**). The x-axis denotes the chromosomal position for all 22 autosomes in hg38, and the y-axis indicates the association P-values on a $-\log_{10}$ scale. Each dot in **a** and **b** indicates a single-nucleotide variant or indel, while each dot in **c** corresponds to a gene. Red dots highlight genome-wide significant signals, while suggestive variants are indicated with orange dots. A dashed line shows the conservative

Bonferroni threshold for genome-wide significance. For **a** and **b**, the gene with the closest proximity to the top variant at each significant locus is listed. Green font was used to highlight known LBD risk loci, while black font indicates novel association signals.

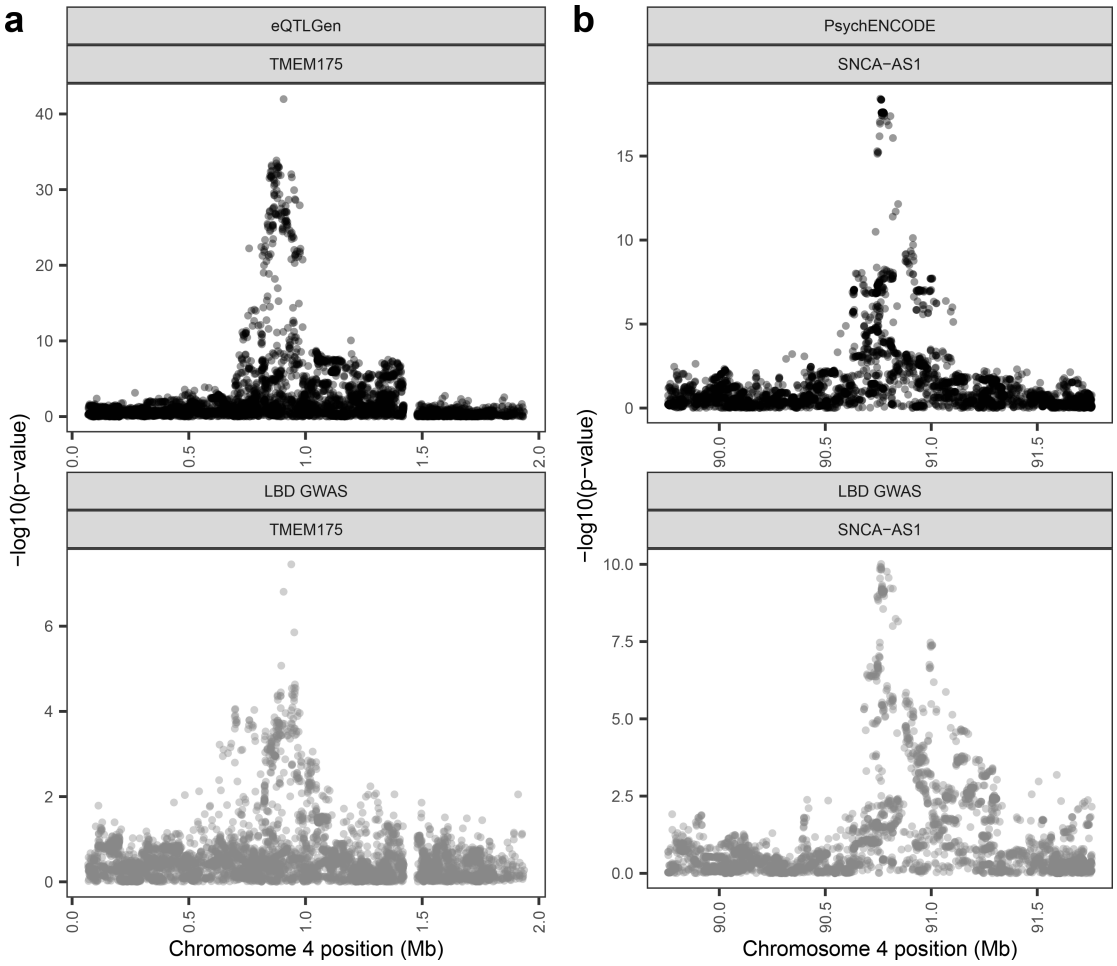


Fig. 3 |. Regional association plots for eQTL and LBD GWAS colocalizations.

a,**b**, Regional association plots for eQTL (upper pane) and LBD GWAS signals (lower pane) in the regions surrounding *TMEM175* (PPH4 = 0.99) (**a**) and *SNCA-AS1* (PPH4 = 0.96) (**b**). The *x*-axis denotes the chromosomal position in hg19, and the *y*-axis indicates the association *P*-values on a $-\log_{10}$ scale.

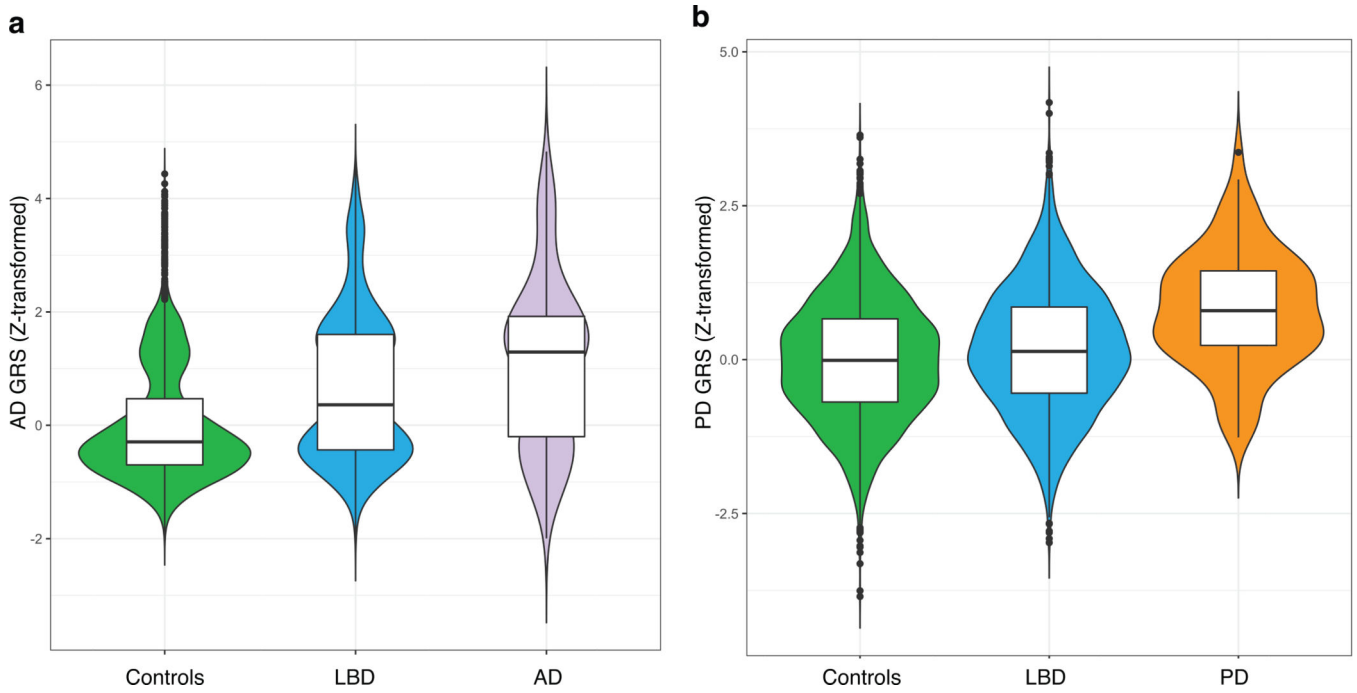


Fig. 4 |. Genetic risk scores from Alzheimer's disease and Parkinson's disease GWAS studies illustrate intersecting molecular genetic risk profiles with LBD.

Alzheimer's disease and Parkinson's disease genetic risk scores predict risk for LBD and highlight overlapping molecular risk profiles. **a**, Violin plots comparing *z*-transformed Alzheimer's disease genetic risk score distributions in LBD cases, controls, and 100 random Alzheimer's disease cases. **b**, Violin plots comparing *z*-transformed Parkinson's disease genetic risk score distributions for LBD cases, controls, and 100 random Parkinson's disease cases. The center line of each violin plot is the median, the box limits depict the interquartile range, and whiskers correspond to the 1.5x interquartile range. Abbreviations: GRS, genetic risk score; AD, Alzheimer's disease; PD, Parkinson's disease.

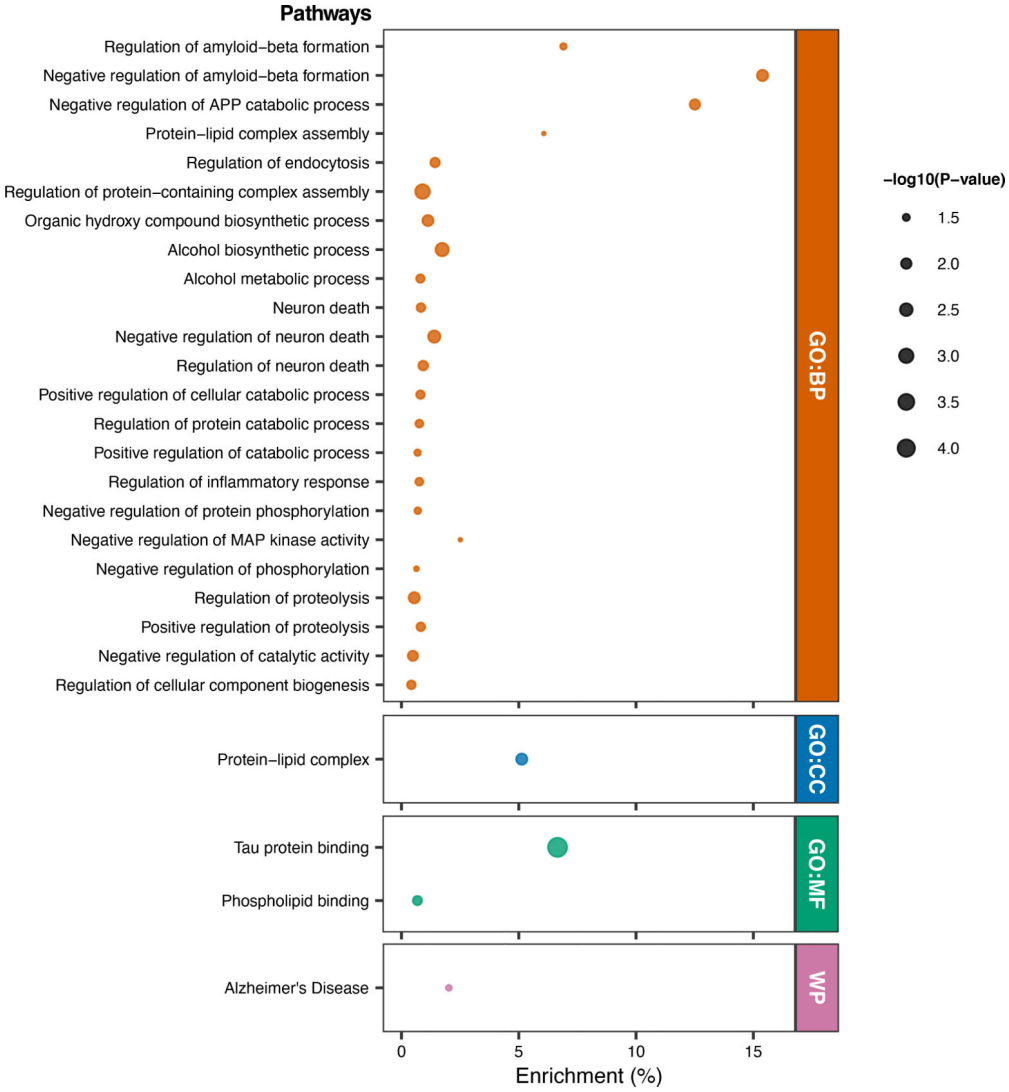


Fig. 5 |. Insights into LBD pathways based on polygenic risk score enrichment analysis.
Functional enrichment analyses of the LBD polygenic risk scores. The *x*-axis corresponds to the enrichment category in LBD cases compared to controls, and the *y*-axis shows the enrichment percentages of significant associations after multiple testing correction. The enrichment percentage refers to the percentage of input genes/variants that are within in a given pathway. Significant gene ontology (GO) enrichments for biological processes (BP, orange), cellular functions (CC, blue), molecular functions (MP, green), and pathways from WikiPathways (WP, pink) are shown. The size of each respective dot indicates the *P*-values on a $-\log_{10}$ scale.

Table 1 |

Genome-wide significant association signals in LBD GWAS

Chr	Position (SNP-ID)	Closest gene	Alleles (A1, A2)	WHOLE GENOME DISCOVERY COHORT				REPLICATION COHORT				Meta-analysis
				EAF		OR (95% CI)	*P*	EAF		OR (95% CI)	*P*	*P*
				Cases	Controls			Cases	Controls			
1	155,236,376 (rs2230288)	*GBA*	C, T	0.028	0.009	2.89 (2.16 – 3.87)	1.28×10^{-12}	0.031	0.013	2.01 (1.46 – 2.78)	1.95×10^{-5}	4.63×10^{-16}
2	127,135,234 (rs6733839)	*BIN1*	C, T	0.416	0.362	1.25 (1.16 – 1.35)	4.16×10^{-9}	0.413	0.382	1.16 (1.05 – 1.28)	3.28×10^{-3}	1.04×10^{-10}
4	945,299 (rs6599388)	*TMEM175*	C, T	0.338	0.310	1.25 (1.15 – 1.35)	3.54×10^{-8}	0.337	0.286	1.15 (1.03 – 1.28)	1.03×10^{-2}	2.61×10^{-9}
4	89,842,209 (rs7680557)	*SNCA-AS1*	A, C	0.440	0.504	0.79 (0.73 – 0.85)	9.73×10^{-11}	0.435	0.509	0.76 (0.64 – 0.90)	5.27×10^{-8}	3.28×10^{-17}
19	44,906,745 (rs769449)	*APOE*	G, A	0.213	0.100	2.46 (2.22 – 2.74)	4.65×10^{-63}	0.222	0.110	2.32 (2.05 – 2.63)	3.27×10^{-40}	2.57×10^{-101}

For each of the five loci, the variant with the lowest *P*-value is listed. The gene that is in closest proximity to the top variant at each locus is represented. The chromosomal position is shown according to hg38. Genome-wide significance was defined as $P < 5 \times 10^{-8}$. Abbreviations: A1, other allele; A2, effect allele; EAF, effect allele frequency; OR, odds ratio; Chr, chromosome; CI, confidence interval.