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

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 is a clinically heterogeneous neurodegenerative disease characterized by progressive cognitive decline, parkinsonism and visual hallucinations. 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 widespread cortical and limbic deposition of pathologically altered forms of α-synuclein proteins in the form of Lewy bodies and Lewy neurites, which are also a hallmark feature of Parkinson’s disease. The vast majority of LBD patients additionally exhibit Alzheimer’s disease copathology. These neuropathological observations have led to the hypothesis, as yet unproven, that LBD lies on a disease continuum between Parkinson’s disease and Alzheimer’s disease. Although relatively common in the community, with an estimated 1.4 million cases prevalent in the United States, 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. We also performed gene-aggregation tests and modeled the relative contributions of Alzheimer’s disease and Parkinson’s disease risk variants to this fatal neurodegenerative disease (Fig. 1 gives 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.** Following quality control, whole-genome sequence data from 2,591 individuals diagnosed with LBD and from 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, while the remaining GWAS signals, in BIN1 and TMEM175, represented new 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).

We examined the associations of BIN1 and TMEM175 risk alleles with the Consortium to Establish a Registry for Alzheimer’s Disease (CERAD) and Braak semiquantitative pathological measures of Alzheimer’s disease copathology. We found that the BIN1 risk allele (rs6733839-T) was significantly associated with increased neurofibrillary tangle pathology (Fisher’s exact test, \( P = 0.0002 \) based on Braak neurofibrillary tangle staging; Extended Data Fig. 1). In contrast, there was no significant association of the TMEM175 risk allele with Alzheimer’s disease copathology. Conditional analyses detected a second signal at the APOE locus (Extended Data Fig. 2 provides regional association plots and Extended Data Fig. 3 gives 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).

**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 of ≤1% and a minor allele count (MAC) of ≥3 across the genome. This rare variant analysis identified GBA as being associated with LBD (Fig. 2c). GBA, encoding the lysosomal enzyme glucocerebrosidase, is a known pleomorphic risk gene for LBD and Parkinson’s disease, 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. 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\textsuperscript{18,19}, the largest available human blood and brain eQTL datasets. We found evidence of colocalization between the \textit{TMEM175} locus and an eQTL regulating \textit{TMEM175} expression in blood (posterior probability for H4 (PPH4) \(= 0.99\); Fig. 3a and Supplementary Table 1). There was also colocalization between the association signal at the \textit{SNCA} locus and an eQTL regulating \textit{SNCA}-AS1 expression in the brain (PPH4 \(= 0.96\); Fig. 3b and Supplementary Table 1). Interestingly, the index variant at the \textit{SNCA} locus was located within the \textit{SNCA}-AS1 gene, which overlaps with the 5’ end of \textit{SNCA} and encodes a long noncoding antisense RNA species known to regulate \textit{SNCA} expression. Sensitivity analyses confirmed that these colocalizations were robust to changes in the previous probability of a variant associating with both traits (Extended Data Fig. 4).

We interrogated the effect of each single-nucleotide polymorphism (SNP) in the region surrounding \textit{SNCA}-AS1 on LBD risk using our GWAS data, and \textit{SNCA}-AS1 expression using the PsychENCODE data (Extended Data Fig. 5a). All genome-wide-significant-risk

**Table 1 | Genome-wide-significant association signals in LBD GWAS**

| Chr | Position (SNP ID) | Closest gene | Alleles (A1, A2) | Genotype | OR (95% CI) | P   |
|-----|------------------|--------------|----------------|---------|-------------|-----|
| 1   | 155,236,376 (rs2230288) | GBA | C, T | 0.028 | 2.89 (2.16–3.87) | \(1.28 \times 10^{-12}\) |
| 2   | 127,135,234 (rs6733839) | BIN1 | C, T | 0.416 | 1.25 (1.16–1.35) | \(4.16 \times 10^{-9}\) |
| 4   | 945,299 (rs6599388) | TMEM175 | C, T | 0.338 | 1.25 (1.15–1.35) | \(3.54 \times 10^{-8}\) |
| 4   | 89,842,209 (rs7680557) | SNCA-AS1 | A, C | 0.440 | 0.79 (0.73–0.85) | \(9.73 \times 10^{-11}\) |
| 19  | 44,906,745 (rs769449) | APOE | G, A | 0.213 | 2.46 (2.22–2.74) | \(4.65 \times 10^{-43}\) |

For each of the five loci, the variant with the lowest P value is listed. The gene 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 \(P < 5 \times 10^{-8}\). A1, other allele; A2, effect allele; Chr, chromosome; EAF, effect allele frequency.
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 (RNA-seq) data from the Genotype-Tissue Expression (GTEx) consortium and single-nucleus RNA-seq data of the medial temporal gyrus from the Allen Institute of Brain Science 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).
Articles

The Alzheimer's disease genetic risk score was also found to be significantly associated with an earlier age of death in LBD ($\beta = -1.77$ years per s.d. increase in genetic risk score from the population mean, s.e. $= 0.19$, $P < 2 \times 10^{-16}$) and shorter disease duration ($\beta = -0.90$ years, s.e. $= 0.27$, $P = 0.0007$). In contrast, the Parkinson's disease genetic risk score was associated with an earlier age at onset among patients diagnosed with LBD ($\beta = -0.98$, s.e. $= 0.28$, $P = 0.00045$), 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

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. We tested the associations of 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 LBD cases (Fig. 4).

Individuals diagnosed with LBD had a higher genetic risk for developing both Alzheimer's disease (odds ratio (OR) = 1.66 per s.d. 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 adjustment for genes that substantially contribute to Alzheimer's disease (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 ($\beta = -1.77$ years per s.d. increase in genetic risk score from the population mean, s.e. $= 0.19$, $P < 2 \times 10^{-16}$) and shorter disease duration ($\beta = -0.90$ years, s.e. $= 0.27$, $P = 0.0007$). In contrast, the Parkinson's disease genetic risk score was associated with an earlier age at onset among patients diagnosed with LBD ($\beta = -0.98$, s.e. $= 0.28$, $P = 0.00045$), 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.

Fig. 3 | Regional association plots for eQTL and LBD GWAS colocalizations. a, b, Regional association plots for eQTL (upper) and LBD GWAS signals (lower) 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 $\text{–log}_{10}$ scale.
based on GWAS risk variants, found several significantly enriched Gene Ontology (GO) 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 $P=1.85 \times 10^{-4}$) and others. Among these, the regulation of amyloid precursor protein, amyloid-beta formation and tau protein binding has previously been implicated in the pathogenesis of Alzheimer’s disease while regulation of endocytosis is particularly important in the pathogenesis of Parkinson’s disease\footnote{25,26}. Those 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\textsuperscript{18}. 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 ($\chi^2 = 5.60$, d.f. = 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 ($\text{GBA, BIN1, TMEM175, SNCA-ASI and APOE}$) that influence risk for developing LBD, whereas the genome-wide, gene-based aggregation tests implicated mutations in $\text{GBA}$ as being critical in the pathogenesis of the disease. We further detected strong cis-eQTL colocalization signals at the $\text{TMEM175}$ and $\text{SNCA-ASI}$ 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, $\text{GBA}$ (encoding the lysosomal enzyme glucocerebrosidase), $\text{APOE}$ (encoding apolipoprotein E) and $\text{SNCA}$ (encoding $\alpha$-synuclein) are known LBD risk genes\textsuperscript{7–9}. In addition to these previously described loci, we identified a new locus on chromosome 2q14.3, located 28kb downstream of the $\text{BIN1}$ gene, which is a known risk locus for Alzheimer’s disease\textsuperscript{11}. $\text{BIN1}$ encodes the bridging integrator 1 protein that is involved in endosomal trafficking. The depletion of $\text{BIN1}$ reduces the lysosomal degradation of $\beta$-site APP-cleaving enzyme 1 (BACE1), resulting in increased amyloid-$\beta$ production\textsuperscript{27}. Furthermore, the loss of $\text{BIN1}$ promotes the propagation of tau pathology by increasing aggregate internalization via endocytosis and endosomal trafficking\textsuperscript{26}. 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 mitigation of $\text{BIN1}$-mediated endosomal dysfunction could have therapeutic implications in both neurodegenerative diseases.

A second new LBD signal was detected within the lysosomal $\text{TMEM175}$ gene on chromosome 4p16.3, a known Parkinson’s disease risk locus\textsuperscript{12}. Deficiency of $\text{TMEM175}$, encoding a trans-membrane potassium channel, impairs lysosomal function, lysosome-mediated autophagosome clearance and mitochondrial respiratory capacity. Loss of function further increases the deposition of phosphorylated $\alpha$-synuclein\textsuperscript{26}, which makes $\text{TMEM175}$ a plausible LBD risk gene. The direction of effect is the same in LBD as in Parkinson’s disease (Supplementary Table 3), and identification of $\text{TMEM175}$ 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 either previously been described in the Alzheimer’s disease literature ($\text{APOE}$ and $\text{BIN1}$) or have been associated with risk of developing Parkinson’s disease ($\text{GBA}$, $\text{SNCA-ASI}$, $\text{APOE}$ and $\text{SNCA}$) and Alzheimer’s disease ($\text{APOE}$ and $\text{SNCA-ASI}$), respectively. The identification of overlapping genetic risk profiles further highlights the importance of understanding the molecular mechanisms underlying the pathogenesis of these neurodegenerative diseases.
disease. Our GWAS and colocalization analyses implicated **SNCA-AS1**, a noncoding RNA that regulates **SNCA** expression, as the main signal at the **SNCA** locus. In contrast, the main signal in Parkinson’s disease was detected at the 3’ end of **SNCA**. 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, to our knowledge, 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

---

**Fig. 5 | Insights into LBD pathways based on polygenic risk score enrichment analysis.** Functional enrichment analyses of 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 a given pathway. Significant GO enrichments for biological processes (BP, orange), cellular functions (CC, blue), molecular functions (MF, green) and pathways from WikiPathways (WP, pink) are shown. The size of each respective dot indicates the P value on a $-\log_{10}$ scale.

*TMEM175* and **SNCA**. Second, genome-wide, gene-based aggregation tests of rare mutations similarly identified **GBA**, which has previously been implicated in Parkinson’s disease. 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 those 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 those observed in Parkinson’s disease. Likewise, the directions of effect for the **BIN1** and **APOE** signals were the same as those detected in Alzheimer’s disease (Supplementary Table 3). 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 noncoding RNA that regulates **SNCA** expression, as the main signal at the **SNCA** locus. In contrast, the main signal in Parkinson’s disease was detected at the 3’ end of **SNCA**. 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, to our knowledge, 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

---

**Pathways**

| Pathway                                                                 | -log$_{10}$(P) |
|------------------------------------------------------------------------|----------------|
| GO:BP Regulation of cellular component biogenesis                       |                |
| GO:BP Negative regulation of catalytic activity                         |                |
| GO:BP Positive regulation of proteolysis                               |                |
| GO:BP Regulation of proteolysis                                         |                |
| GO:BP Negative regulation of phosphorylation                           |                |
| GO:BP Negative regulation of MAP kinase activity                       |                |
| GO:BP Regulation of proteolysis                                         |                |
| GO:BP Positive regulation of proteolysis                               |                |
| GO:BP Negative regulation of catalytic activity                        |                |
| GO:BP Regulation of cellular component biogenesis                       |                |
| GO:OC Protein–lipid complex                                             |                |
| GO:MF Phospholipid binding                                              |                |
| GO:MF Tau protein binding                                               |                |
| GO:MF Phospholipid binding                                              |                |
| WP Alzheimer’s disease                                                 |                |

---

**Enrichment (%)**

0 5 10 15

0 5 10 15

0 5 10 15

0 5 10 15

---

**Fig. 5 | Insights into LBD pathways based on polygenic risk score enrichment analysis.** Functional enrichment analyses of 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 a given pathway. Significant GO enrichments for biological processes (BP, orange), cellular functions (CC, blue), molecular functions (MF, green) and pathways from WikiPathways (WP, pink) are shown. The size of each respective dot indicates the P value on a $-\log_{10}$ scale.
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 have 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-ASI 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. First, we focused on individuals of European ancestry because 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- rather than long-read sequencing applications, which limits the resolution of complex, repetitive and guanine- and cytosine-rich regions. Most study participants did not have in-depth phenotype information using the 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 new loci as relevant in the pathogenesis of LBD. Our findings confirm that LBD genetically intersects with Alzheimer's disease and Parkinson's disease and highlights 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 prioritization of targets for therapeutic development. We have made the whole-genome sequence data available to the research community. These genomes constitute, to our knowledge, the largest sequencing effort in LBD to date and are designed to accelerate the pace of discovery in dementia.

Online content
Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41588-021-00785-3.

Received: 6 July 2020; Accepted: 12 January 2021; Published online: 15 February 2021

References
1. Walker, Z., Possin, K. L., Boeve, B. F. & Aarsland, D. Lewy body dementias. *Lancet* **386**, 1683–1697 (2015).
2. McKeith, I. G. et al. Diagnosis and management of dementia with Lewy bodies: fourth consensus report of the DLB Consortium. *Neurology* **89**, 88–100 (2017).
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).
4. 10 things you should know about LBD. *Lewy Body Dementia Association* https://www.lbda.org/10-things-you-should-know-about-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).
6. Sabir, M. S. et al. Assessment of APOE in atypical parkinsonism syndromes. *Neurobiol. Dis.* **127**, 142–146 (2019).
7. Nalls, M. A. et al. A multicenter study of glucocerebrosidase mutations in dementia with Lewy bodies. *JAMA Neurol.* **70**, 727–735 (2013).
8. Singleton, A. B. et al. Alpha-Synuclein locus triplication causes Parkinson’s disease. *Science* **302**, 841 (2003).
9. Tsuang, D. et al. APOE epsilon4 increases risk for dementia in pure synucleinopathies. *JAMA Neurol.* **70**, 223–228 (2013).
10. Pickering-Brown, S. M. 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).
12. Pankratz, N. et al. Genomewide association study for susceptibility genes contributing to familial Parkinson disease. *Hum. Genet.* **124**, 593–605 (2009).
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).
14. Geiger, J. T. et al. Next-generation sequencing reveals substantial genetic contribution to dementia with Lewy bodies. *Neurobiol. Dis.* **49**, 55–62 (2016).
15. Singleton, A. & Hardy, J. A generalizable hypothesis for the genetic architecture of disease: pheomorph risk loci. *Hum. Mol. Genet.* **20**, R158–R162 (2011).
16. Nicolae, D. L. et al. Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS Genet.* **6**, e1000888 (2010).
17. Li, Y. I. et al. RNA splicing is a primary link between genetic variation and disease. *Science* **352**, 600–605 (2016).
18. Vösa, U. et al. Unraveling the polygenic architecture of complex traits using whole eQTL metaanalysis. Preprint at bioRxiv https://doi.org/10.1101/447367 (2018).
19. Wang, D. et al. Comprehensive functional genomic resource and integrative model for the human brain. *Science* **362**, eaat4846 (2018).
20. Hawrylycz, M. J. et al. An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* **489**, 391–399 (2012).
21. GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multisite gene regulation in humans. *Science* **348**, 648–660 (2015).
22. Kunkle, B. W. 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).
23. Nalls, M. A. 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).
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. Disorder.* **34**, 460–468 (2019).
26. Morris, J. C. The Clinical Dementia Rating (CDR); current version and scoring rules. *Neurology* **43**, 2412–2414 (1993).
27. Miyagawa, T. et al. BACE1 intracellular trafficking and amyloid-beta production. *Hum. Mol. Genet.* **25**, 2948–2958 (2016).
28. Calafate, S., Flavin, W., Verstreken, P. & Moehrs, D. Loss of Bic1 promotes the propagation of Tau pathology. *Cell Rep.* **17**, 931–940 (2016).
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).
30. Corder, E. H. 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).
31. Sidransky, E. et al. Multicenter analysis of glucocerebrosidase mutations in Parkinson’s disease. *N. Engl. J. Med.* **361**, 1651–1661 (2009).
32. Simon-Sanchez, J. et al. Genome-wide association study reveals genetic risk underlying Parkinson’s disease. *Nat. Genet.* **41**, 1308–1312 (2009).
33. Jansen, I. E. et al. Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer’s disease risk. *Nat. Genet.* **41**, 404–413 (2019).
34. Ross, O. A. et al. Genomic investigation of alpha-synuclein multiplication and parkinsonism. *Annu. Neurol.* **63**, 743–750 (2008).
35. Pollard, M. O., Gurudasani, D., Mentzer, A. J., Porter, T. & Sandhu, M. S. Long reads: their purpose and place. *Hum. Mol. Genet.* **28**, R234–R241 (2018).

Publisher’s note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. This is a U.S. government work and not under copyright protection in the U.S.; foreign copyright protection may apply 2021.
Ruth Chia\textsuperscript{1,30}, Marya S. Sabir\textsuperscript{2,30}, Sara Bandres-Ciga\textsuperscript{3}, Sara Saez-Atienza\textsuperscript{1}, Regina H. Reynolds\textsuperscript{4,5,6}, Emil Gustavsson\textsuperscript{5,6}, Ronald L. Walton\textsuperscript{7}, Sarah Ahmed\textsuperscript{8}, Coralie Viollet\textsuperscript{8,9}, Jinhui Ding\textsuperscript{10}, Mary B. Makarious\textsuperscript{2}, Monica Diez-Fairen\textsuperscript{11}, Makayla K. Portley\textsuperscript{4}, Zalak Shah\textsuperscript{2}, Yevgeniya Abramzon\textsuperscript{12}, Dena G. Hernandez\textsuperscript{3}, Cornelis Blauwendraat\textsuperscript{3}, David J. Stone\textsuperscript{13}, John Eicher\textsuperscript{14}, Laura Parkkinnen\textsuperscript{15}, Olaf Ansorge\textsuperscript{16}, Lorraine Clark\textsuperscript{16}, Lawrence S. Honig\textsuperscript{17}, Karen Marder\textsuperscript{17}, Afina Lemstra\textsuperscript{18}, Peter St George-Hyslop\textsuperscript{19,20}, Elisabet Londos\textsuperscript{21}, Kevin Morgan\textsuperscript{22}, Tammaryn Lashley\textsuperscript{4,23}, Thomas T. Warner\textsuperscript{12,23}, Zane Jaunmuktane\textsuperscript{23}, Douglas Galasko\textsuperscript{24,25}, Isabel Santana\textsuperscript{26,27,28,29}, Pentti J. Tienari\textsuperscript{30,31}, Liisa Myllykangas\textsuperscript{32,33}, Minna Oinas\textsuperscript{34}, Nigel J. Cairns\textsuperscript{35}, John C. Morris\textsuperscript{35}, Glenda M. Halliday\textsuperscript{36,37,38}, Viviana M. Van Deerlin\textsuperscript{39}, John Q. Trojanowski\textsuperscript{39}, Maurizio Grassano\textsuperscript{40}, Andrea Calvo\textsuperscript{40,41}, Gabriele Mora\textsuperscript{42}, Antonio Canosa\textsuperscript{40,41}, Gianluca Floris\textsuperscript{43}, Ryan C. Bohannan\textsuperscript{44}, Francesca Brett\textsuperscript{45}, Ziv Gan-Or\textsuperscript{46}, Joshua T. Geiger\textsuperscript{2}, Anni Moore\textsuperscript{10}, Patrick May\textsuperscript{47}, Rejko Krüger\textsuperscript{47,48,49}, David S. Goldstein\textsuperscript{50}, Grisel Lopez\textsuperscript{51}, Nahid Tayebi\textsuperscript{51}, Ellen Sidransky\textsuperscript{52}, The American Genome Center\textsuperscript{53}, Lucy Norcliffe-Kaufmann\textsuperscript{54}, Jose-Alberto Palma\textsuperscript{55}, Horacio Kaufmann\textsuperscript{56}, Vikram G. Shakkottai\textsuperscript{57}, Matthew Perkins\textsuperscript{58}, Kathy L. Newell\textsuperscript{59}, Thomas Gassey\textsuperscript{60}, Claudia Schulte\textsuperscript{56}, Francesco Landi\textsuperscript{57}, Erika Salvi\textsuperscript{61}, Daniele Cusi\textsuperscript{62}, Eliezer Masliah\textsuperscript{63}, Ronald C. Kim\textsuperscript{64}, Chad A. Caraway\textsuperscript{65}, Edwin S. Monuki\textsuperscript{66}, Maura Brunetti\textsuperscript{67}, Ted M. Dawson\textsuperscript{68,69,70}, Liana S. Rosenthal\textsuperscript{71}, Marilyn S. Albert\textsuperscript{72}, Olga Pletnikova\textsuperscript{73,74}, Juan C. Troncoso\textsuperscript{75}, Margaret E. Flanagan\textsuperscript{76,77}, Qinwen Mao\textsuperscript{78,79,80}, Eileen H. Bigio\textsuperscript{81,82}, Eloy Rodriguez-Rodriguez\textsuperscript{83}, Jon Infante\textsuperscript{84}, Carmen Lage\textsuperscript{85}, Isabel Gonzalez-Aramburu\textsuperscript{86}, Pascual Sanchez-Juan\textsuperscript{87}, Bernardino Ghetti\textsuperscript{88}, Julia Keith\textsuperscript{89}, Sandra E. Black\textsuperscript{90,91}, Mario Masellis\textsuperscript{92,93}, Ekaterina Rogueva\textsuperscript{94}, Charles Duyckaerts\textsuperscript{95,96}, Alexis Brice\textsuperscript{97,98}, Suzanne Lesage\textsuperscript{99}, Georgia Xiromerisiou\textsuperscript{100,101}, Matthew J. Barrett\textsuperscript{102}, Bensin S. Tilley\textsuperscript{103}, Steve Gentleman\textsuperscript{104,105}, Giancarlo Logroscino\textsuperscript{106,107}, Geidy E. Serrano\textsuperscript{108}, Thomas G. Beach\textsuperscript{109}, Ian G. McKeith\textsuperscript{110}, Alan J. Thomas\textsuperscript{111}, Johannes Attems\textsuperscript{112}, Christopher M. Morris\textsuperscript{113}, Laura Palmer\textsuperscript{114}, Seth Love\textsuperscript{115}, Claire Troakes\textsuperscript{116}, Safa Al-Sarraj\textsuperscript{117}, Angela K. Hodges\textsuperscript{118}, Dag Aarsland\textsuperscript{119,120,121}, Gregory Klein\textsuperscript{122}, Scott M. Kaiser\textsuperscript{123}, Randy Woltjer\textsuperscript{124}, Pau Pastor\textsuperscript{125}, Lynn M. Bekris\textsuperscript{126}, James B. Leverenz\textsuperscript{127}, Lilah M. Besser\textsuperscript{128}, Amanda Kuzma\textsuperscript{129}, Alan E. Renton\textsuperscript{130}, Alison Goate\textsuperscript{131}, David A. Bennett\textsuperscript{132}, Clemens R. Scherzer\textsuperscript{133}, Huw R. Morris\textsuperscript{134}, David Raffaele Ferrari\textsuperscript{135}, Diego Albani\textsuperscript{136}, Stuart Pickering-Brown\textsuperscript{137}, Kelley Faber\textsuperscript{138,139}, Walter A. Kukull\textsuperscript{140}, Estrella Morenas-Rodriguez\textsuperscript{141}, Alberto Lleó\textsuperscript{142,143}, Juan Fortea\textsuperscript{144,145}, Daniel Alcolea\textsuperscript{146,147}, Jordi Clarimon\textsuperscript{148}, Mike A. Nalls\textsuperscript{149,150,151}, Luigi Ferrucci\textsuperscript{152}, Susan M. Resnick\textsuperscript{153}, Toshiko Tanaka\textsuperscript{154}, Tatiana M. Foroud\textsuperscript{155}, Neill R. Graff-Radford\textsuperscript{156}, Zbigniew K. Wszolek\textsuperscript{157}, Tanis Ferman\textsuperscript{158}, Bradley F. Boeve\textsuperscript{159}, John A. Hardy\textsuperscript{160,161,162}, Eric J. Topol\textsuperscript{163,164}, Ali Torkamani\textsuperscript{165}, Andrew B. Singleton\textsuperscript{166,167}, Mina Ryten\textsuperscript{168}, Dennis W. Dickson\textsuperscript{169}, Adriano Chiò\textsuperscript{170,171}, Owen A. Ross\textsuperscript{172,173}, J. Raphael Gibbs\textsuperscript{174,175}, Clifton L. Dalgard\textsuperscript{176,177,178}, Bryan J. Traynor\textsuperscript{179,180,181}, and Sonja W. Scholz\textsuperscript{182,183,184,185,186}}
Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN, USA. 110National Alzheimer’s Coordinating Center, Department of Epidemiology, University of Washington, Seattle, WA, USA. 111Biomedizinisches Centrum, Biochemie, Ludwig-Maximilians-Universität München & Deutsches Zentrum für Neurodegenerative Erkrankungen, Munich, Germany. 112Sant Pau Biomedical Research Institute, Hospital de la Santa Creu i Sant Pau, Universitat Autònoma de Barcelona, Barcelona, Spain. 113The Network Center for Biomedical Research in Neurodegenerative Diseases, Madrid, Spain. 114Laboratory of Neurogenetics, National Institute on Aging, Bethesda, MD, USA. 115Data Tecnica International, Glen Echo, MD, USA. 116Center for Alzheimer’s and Related Dementias, National Institute on Aging, Bethesda, MD, USA. 117Longitudinal Studies Section, National Institute on Aging, Baltimore, MD, USA. 118Laboratory of Behavioral Neuroscience, National Institute on Aging, Baltimore, MD, USA. 119Department of Neurology, Mayo Clinic, Jacksonville, FL, USA. 120Department of Psychiatry and Psychology, Mayo Clinic, Jacksonville, FL, USA. 121Department of Neurology, Mayo Clinic, Rochester, MN, USA. 122UK Dementia Research Institute of UCL, UCL Institute of Neurology, University College London, London, UK. 123UCL Movement Disorders Centre, University College London, London, UK. 124Institute for Advanced Study, The Hong Kong University of Science and Technology, Hong Kong SAR, China. 125Scripps Research Translational Institute, Scripps Research, La Jolla, CA, USA. 126Institute of Cognitive Sciences and Technologies, C.N.R., Rome, Italy. 127Department of Clinical Genomics, Mayo Clinic, Jacksonville, FL, USA. 128Department of Anatomy, Physiology & Genetics, Uniformed Services University of the Health Sciences, Bethesda, MD, USA. 129The American Genome Center, Uniformed Services University of the Health Sciences, Bethesda, MD, USA. 130These authors contributed equally: Ruth Chia, Marya S. Sabir. 131These authors jointly supervised: Adriano Chiò, Owen A. Ross, J. Raphael Gibbs, Clifton L. Dalgard, Bryan J. Traynor, Sonja W. Scholz. *A list of authors and their affiliations appears at the end of the paper. ✉ e-mail: sonja.scholz@nih.gov

The American Genome Center

Anthony R. Sotis129, Coralie Viollet8,9, Gauthaman Sukumar129, Camille Alba129, Nathaniel Lott129, Elisa McGrath Martinez129, Meila Tuck129, Jatinder Singh129, Dagmar Bacikova129, Xijun Zhang129, Daniel N. Hupalo129, Adelani Adeleye129, Matthew D. Wilkerson129, Harvey B. Pollard129 and Clifton L. Dalgard128,129

The American Genome Center

Anthony R. Sotis129, Coralie Viollet8,9, Gauthaman Sukumar129, Camille Alba129, Nathaniel Lott129, Elisa McGrath Martinez129, Meila Tuck129, Jatinder Singh129, Dagmar Bacikova129, Xijun Zhang129, Daniel N. Hupalo129, Adelani Adeleye129, Matthew D. Wilkerson129, Harvey B. Pollard129 and Clifton L. Dalgard128,129
Methods

Cohort description and study design. A total of 5,154 participants of European ancestry (2,981 LBD cases and 2,173 neurologically healthy controls) were recruited across 17 European and 27 North American sites/consortia to create a genomic resource for LBD research. In addition to the resource genomes, we obtained convenience control genomes from (1) the Welderly cohort (n = 1,202), a cohort of healthy, aged European-ancestry individuals recruited in the United States and (2) European-ancestry control genomes generated by the National Institute on Aging and the Accelerating Medicine Partnership—Parkinson’s Disease Initiative (www.amp-edu.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 notable neurodegenerative disease on histopathological examination.

Patients with LBD were diagnosed with pathologically definite or clinically probable disease according to consensus criteria. The case cohort included 1,789 (69.0%) autopsy-confirmed LBD cases and 802 (31.0%) patients with clinically probable LBD. Among LBD cases 63.4% were male, as is typical for the LBD patient population. The demographic characteristics of the cohorts are summarized in Supplementary Table 5. The approximate institutional review boards of participating institutions approved the study (nos. 03-AG-N329 and 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 genomic DNA samples was performed using the PicoGreen dsDNA assay (Thermo Fisher). PCR-free, paired-end libraries were constructed by automated liquid handlers using 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 and 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. The GatkH3D reference genome 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 (2), was implemented in a 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 GATK Best Practices and using another Broad Institute workflow for joint discovery and Variant Quality Score Recalibration. Both Broad workflows for whole-genome sequence processing and joint discovery are publicly available (https://github.com/gatk/workflows/broad-prod-refs-germline-variant-indexes). 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 onVerifyBamID freemix metric); (2) an excessive heterozygosity rate (>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 = 0.8), (6) non-European ancestry based on principal components analysis when compared to the HapMap 3 Genome Reference Panel (Extended Data Fig. 9a) and (7) samples that were related (defined as having pi-hat > 0.125).

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

Followed applicable quality control filters, 6,651 samples remained available for analysis. Extended Data Fig. 10 shows quality control metrics.

Statistical analysis for single-variant association. We performed a GWAS in LBD (n = 2,591 cases and n = 4,027 controls) using logistic regression in PLINK (v.2.0) with a minor allele frequency threshold of >1% based on allele frequency estimates in the LBD case cohort. We used the step function in the R MASS package to determine the minimum number of principal components (generated from the first independent single-nucleotide variants required to correct for population substructure). The first two principal components in our study cohorts compared to the HapMap3 Genomic Resource Panel are shown in Extended Data Fig. 9a. Based on this analysis, we incorporated sex, age and five principal components (PC1, PC3, PC4, PC5 and PC7) as covariates in our model. Quantile–quantile plots revealed minimal residual deviation from the expected null distribution, further supported by a sample-size-adjusted, genome-wide inflation factor λ_{min} 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 the addition of each respective index variant to the covariates (Extended Data Fig. 3).

For the LBD GWAS replication analysis, we obtained genotyping array data from a independent, nonoverlapping, European-ancestry LBD case-control cohorts, totaling 970 LBD cases and 8,928 controls, as described elsewhere. The data were cleaned by applying the same sample- and variant-level quality control steps used in the discovery genomes. We imputed the data against the NHLBI TOPMed imputation reference panel under default settings with Eagle (5.9.2) (6599388-T) with the pathological changes of Alzheimer’s disease. Neuritic plaque staining information, as assessed by the B Braak method, was available for 700 pathologically confirmed LBD cases while neurofibrillary tangle pathology staging, as assessed by the Braak method, was available for 1,459 confirmed LBD cases. Association testing between risk alleles and the semiquantitative neuritic plaque and neurofibrillary tangle burden was performed using Fisher’s exact test.

Colocalization analyses. Coloc (v.4.0.1) was used to evaluate the probability of LBD loci and QTL sharing a single causal variant. This tool incorporates a Bayesian statistical framework that computes posterior probabilities for five hypotheses: namely, there is no association with either trait (hypothetic 0, H0); an associated LBD variant exists but no associated eQTL variant (H1); there is an associated eQTL variant but no associated LBD variant (H2); there is an association with an eQTL and LBD risk variant, but they are two independent variants (H3); and there is a shared associated LBD variant and eQTL variant within the analyzed region (H4). Cis-eQTL were derived from eQTLGen (n = 31,684 individuals, accessed 20 February 2020) and PsychoENCODE (n = 1,387 individuals, accessed 20 February 2020) (ref. 44). 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 × 10−4). Colocalization was run using the default priors p0 = 10−4 and p1 = 10−8, while prior p2 was set to p2 = 5 × 10−4 (ref. 44). Loci with PHU > 0.90 were considered colocalized. All colocalizations were subjected to a second test to explore the robustness of our conclusions to changes in prior p0, (that is, the probability that a given variant affects both traits).

Cell-type and tissue-specificity measures. To determine the specificity of a given tissue type, we counted the number of tissue types with a specificity value of 0.95 or greater from two independent gene expression datasets: (1) bulk-tissue RNA-seq of 53 human tissues from GTEx v8 (ref. 45) and (2) human single-nucleus RNA-seq of the middle temporal gyrus from the Allen Institute for Brain Science (n = 7 cell types). The specificity values for GTEx were generated using modified code from a previous publication. Expression of tissues was averaged by organ (except in the case of brain: n = 35 tissues in total). Cell-type and tissue-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. The specificity values for both datasets and the code used to generate these values are available at https://github.com/RHReynolds/MarkerGenes.

Heritability analysis. Narrow-sense heritability (h2), a measure of additive genetic variance, was calculated using genome-based restricted maximum likelihood– linkage disequilibrium and minor allele frequency stratified (GREML-LDMS) to determine how much of the genetic liability for LBD is explained by common genetic variants. This analysis included unrelated individuals (pi-hat = 0.125, n = 2,591 LBD cases and n = 4,027 controls) and autosomal variants with MAF > 1%. The analysis was adjusted for sex, age and five principal components (PC1, PC3, PC4, PC5 and 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 SKAT-O analysis of missense mutations to determine the difference in the aggregate burden of rare coding variants between LBD cases and controls. This analysis was performed in RVTESTS (v.2.1.0) using default parameters after annotation of variants in ANNOVAR (v.2018-04/16) (ref. 46). 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 MAC ≥2 as filters. The covariates used in this analysis included sex, age and five principal components (PC1, PC3, PC4, PC5 and 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, allele dosages were counted across Alzheimer’s disease or Parkinson’s disease loci per sample (that is, 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 s.d. = 1 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 and 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. Polygenic risk score was computed by summing the risk alleles associated with LBD that had been weighted by the effect size estimated by performing a GWAS in pathologically confirmed LBD cases and controls. This workflow identified the optimum P value threshold (1 × 10⁻¹⁰ in our dataset) for variant selection, allowing for the inclusion of variants that failed to reach genome-wide significance but that, nonetheless, contributed to disease risk. After exclusion of variants lacking an rs-identifier, the remaining 122 variants were ranked based on their PRSice-2 P values with the addition of genes GBA, BDNF, VPS13A and TMEM175 to the top five positions. The list was then analyzed for pathway enrichment using the gProfiler toolkit (v.0.1.8). We defined genes involved in the pathways and gene sets using the following databases: (1) GO, (2) Kyoto Encyclopedia of Genes and Genomes, (3) Reactome and (4) WikiPathways. Significant pathways and gene lists with either a single gene or containing >1,000 genes were discarded. Significance was defined as P < 0.05. The gProfiler 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 separation of 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.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability
Individual-level sequence data for the resource genomes have been deposited at dbGaP (accession no. 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 QT data are available at resource.psychencode.org. Bulk-tissue RNA-seq data (GTEx v.8) are available at the Genotype-Tissue Expression consortium portal (https://www.gtexportal.org/home/). Human single-nucleus RNA-seq 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://gatk-broadinstitute.github.io/gatk-workflows/broad-prod-wgs-germline-snp-indels.html). Contamination rates were estimated using VerifyBamID (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 (v.3.5.2; https://www.r-project.org/). Human single-nucleus RNA-seq 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.

References
Erikson, G. A. et al. Whole-genome sequencing of a healthy aging cohort. Cell 165, 1002–1011 (2016).
Emre, M. et al. Clinical diagnostic criteria for dementia associated with Parkinson’s disease. Mov. Disord. 22, 1689–1707 (2007).
Savica, R. et al. Incidence of dementia with Lewy bodies and Parkinson’s disease dementia. JAMA Neurol. 70, 1396–1402 (2013).
Regier, A. A. et al. Functional equivalence of genome sequencing pipelines enables harmonized variant calling across human genetics projects. Nat. Commun. 9, 4038 (2018).
Van der Auwera, G. A. 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).
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).
International HapMap Consortium et al. Integrating common and rare genetic variation in diverse human populations. Nature 467, 52–58 (2010).
Schneider, V. A. et al. Evaluation of GRCh38 and de novo haploid genome assemblies demonstrates the enduring quality of the reference assembly. Genome Biol. 27, 849–864 (2017).
Karczewski, K. J. et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature 581, 434–443 (2020).
Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 7, 8 (2015).
Venable, W. N. & Ripley, B. D. Modern Applied Statistics with S (Springer, 2002).
Das, S. et al. Next-generation genotype imputation service and methods. Nat. Genet. 48, 1284–1287 (2016).
Fuchsberger, C., Abecasis, G. R. & Hinds, D. A. minicM2: faster genotype imputation. Bioinformatics 31, 782–784 (2015).
Talbot, D. et al. Sequencing of 53,831 diverse genomes from the NHBLI TOPMed Program. Preprint at bioRxiv https://doi.org/10.1101/563866 (2019).
Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 26, 2190–2191 (2010).
Mirra, S. S. 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).
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).
Giambartolomei, C. et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet. 10, e1004383 (2014).
Wallace, C. Eliciting priors and relating the single causal variant assumption in colocalisation analyses. PLoS Genet. 16, e1008720 (2020).
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).
Skene, N. G. & Grant, S. G. Identification of vulnerable cell types in major brain disorders using single cell transcriptomes and expression weighted cellular type enrichment. Front. Neurosci. 10, 16 (2016).
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).
Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 38, e164 (2010).
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. T.G.B. 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 variant analysis, with contributions from M.B.M., M.D.-F. and C.B. M.S.S. performed the heritability analysis. S.-C. performed the genetic risk score analysis. S.S.-A. performed and supervised the study. M.P., K.L.N., E.M.-R., A.K.H., D. Aarsland, G.K., S.M.K., R.W., P.P., L.M. Bekris, J.B.L., L.M. performed validation experiments. R.H.R., M.R., M.G., A. Calvo, G.M., A. Canosa, G.E., R.C.B., EB., Z.G.-O., P.M., R.K., D.S.G., G.L., N.T., E.S., I.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., E.P., J.C.T., M.E.E., Q.M., E.H.B., E.-R.-R., J.J., 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.I.B., R.S.T., S.G., G.L., G.E.S., T.G.B., 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.E., D. Albani, S.P., R.E., K.A.W.K., E.M.-R., A.L., J.F., D. Alcolea, J.C., L.F., S.M.R., T.T., T.M.E., 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., I.P., A.B.S., J.A.H., O.A., I.C., L.S.H., K. Marder, A.L., P. St.G.-H., E.L., K. Morgan, T.L., T.T.W., Z.I., D.G., I.S., P.T., L.M., M.O., N.I.C., J.C.M., G.M.H., V.M.V.D., J.Q.T., T.G., C.S., E.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.

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 and 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 coapplicant on a patent application related to C9or72—Method for diagnosing a neurodegenerative disease (no. 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 Parkinsonism Disease. H.K. is Editor-in-Chief of Clinical Autonomic Research, serves as PI of a clinical trial sponsored by Biogen MA, Inc. (TRACK SMA, no. S19-01846) and 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 for 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 an 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 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 coprincipal investigator on Biogen, Inc. (no. 22PD2001), Biohaven Pharmaceuticals, Inc. (nos. BHV4157-206 and BHV3241-301) and Neuraly, 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 Parkinsonism 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.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41588-021-00785-3. Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41588-021-00785-3.

Correspondence and requests for materials should be addressed to S.W.S.

Peer review information Nature Genetics thanks Sudha Seshadri and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permissions information is available at www.nature.com/reprints.
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 | See next page for caption.
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 -log10 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 H0–H4 hypotheses across varying p12 priors are shown. A dashed vertical line indicates the value of p12 used in the initial analysis (p12 = 5 x 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 $-\log_{10}$ 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 ($5 \times 10^{-8}$). 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 \( \lambda_{1000} \) 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.
Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- Confirmed
- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
- Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted
- Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen’s d, Pearson’s r), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection

Whole-genome sequencing was performed using the Illumina X10 sequencer (v.2.5 chemistry) using 150 bp, paired-end cycles.

Data analysis

(1) 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-snp-indels).
(2) Contamination rates were assessed using VerifyBamID (v.1.1.3) (https://genome.sph.umich.edu/wiki/VerifyBamID).
(3) 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/).
(4) 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).
(5) Imputation was performed using Minimac4 on data phased by Eagle (v.2.4) (https://github.com/poruloh/Eagle).
(6) Meta-analysis was performed using METAL (v.2018-08-28) (https://genome.sph.umich.edu/wiki/METAL).
(7) Heritability analysis was performed using GRML-LDMS in GCTA (v.1.26.0) (https://cnsgenomics.com/software/gcta).
(8) 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/).
(9) Genetic risk score analyses were performed in PLINK 1.9 (v.1.9.0-beta4.4) (https://www.cog-genomics.org/plink).  
(10) LD 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) (https://github.com/chr1swallace/coloc).
(11) Colocalization analyses were performed in R-3.2 using the packages coloc (v.4.0.1) (https://github.com/chr1swallace/coloc).
(12) 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).
(13) Polygenic risk scores were constructed using PRSice-2 (v.2.2.1) (https://www.prsice.info).
(14) Pathway enrichment analysis was performed using the R package gprofiler2 (v.0.2.0) (https://cran.r-project.org/web/packages/
Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

(1) The individual-level sequence data for the resource genomes have been deposited at dbGaP (accession number: phs001963.v1.p1 NIA DementiaSeq): https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001963.v1.p1
(2) The GWAS summary statistics have been deposited in the GWAS catalog (study accession #: GCST90001390): https://www.ebi.ac.uk/gwas/home.
(3) eQTLGen data are openly available at https://www.eqtlgen.org/cis-eqtls.html.
(4) PsychENCODE QTL data are openly available at http://resource.psychencode.org/.
(5) Bulk-tissue RNA sequence data (GTEx version 8) are openly available at the Genotype-Tissue Expression consortium portal (https://www.gtexportal.org/home/).
(6) Human single-nucleus RNA sequence data are openly available at the Allen Institute for Brain Science portal (portal.brain-map.org/atlas-and-data/rnaseq/human-mtg/smart-seq).
(7) Specificity values for the Allen Institute for Brain Science and GTEx data and the code used to generate these values are openly available at: https://github.com.RHReynolds/MarkerGenes.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences  ☐ Behavioural & social sciences  ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size
After quality control filters were applied, there were 6,651 samples available for analysis (2,591 cases, and 4,027 controls). For replication, we used a dataset from 970 Lewy body dementia cases and 8,928 healthy individuals. We determined that a sample of this size has 98% power to detect an association of a variant, assuming an odds ratio of 1.25 and a minor allele frequency of 36% under an additive model. The odds ratio and minor allele frequency values were chosen based on a variant that achieved genome-wide significance in the discovery cohort (BIN1).

Data exclusions
Genomes were excluded from the analysis for the following pre-established reasons:
(1) a high contamination rate (>5% based on VerifyBamID freemix metric),
(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 (Supplementary Fig. 8),
and (7) samples that were related (defined as having a pi-hat > 0.125).

Replication
The replication cohort included 970 cases with Lewy body dementia and 8,928 controls. All participants were of European ancestry. We confirm that all attempts at replication in this cohort were successful.

Randomization
Randomization is not used in studies involving case-control genome-wide association studies. Furthermore, it is not possible to randomize the disease status of the individuals involved in the study. The case samples were diagnosed with LBD according to clinical and neuropathological criteria before their inclusion in the study. The control subjects were diagnosed as healthy and without neurological disease before their inclusion in the study. A randomization to test our data's significance was not used, as it was not applicable in this setting.

Blinding
Blinding is not used in studies involving case-control genome-wide association studies. Furthermore, it was not possible to blind the analyst of the disease status of the samples. The case samples were diagnosed with LBD according to clinical and neuropathological consensus criteria before their inclusion in the study. The control samples were designated as healthy before their inclusion in the study.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.
Materials & experimental systems

| Involved in the study |
|-----------------------|
| Antibodies            |
| Eukaryotic cell lines |
| Palaeontology and archaeology |
| Animals and other organisms |
| Human research participants |
| Clinical data |
| Dual use research of concern |

Methods

| Involved in the study |
|-----------------------|
| ChIP-seq |
| Flow cytometry |
| MRI-based neuroimaging |

Human research participants

Population characteristics

All study participants were of European ancestry. Demographic characteristics are described in Supplementary Table 5. Characteristics of the replication cohort have been described elsewhere (PMID: 30798004, PMID: 29263008).

Recruitment

Lewy body dementia cases were diagnosed according to established consensus criteria. Similarly, the control cohort participants were screened to ensure that they were healthy. This approach minimized self-selection bias and ensured that each participant was assigned the correct case/control status in the analysis. Our sample selection was limited to a single racial/ethnic category, namely European-ancestry individuals. This selection criteria may have introduced a bias in the study in that our reported results are mainly relevant for this racial/ethnic category. Limiting to a single racial/ethnic category is the standard approach in case-control genome-wide association studies. The impact of this on our results is outlined in the limitations paragraph of the discussion.

Ethics oversight

The participating institute whose institutional review board approved the protocol is the National Institute on Aging, Bethesda, Maryland, USA (study number: 03-AG-N329).

Clinical data

Clinical trial registration

NCT02014246

Study protocol

Study protocol is available upon request.

Data collection

Participants were recruited across 17 European and 27 North American centers with neurology expertise. Samples were acquired for this study between 01/01/2017 - 09/30/2018. Data were collected from each participant, including demographic information, case/control status, and disease characteristics (for Lewy body dementia cases).

Outcomes

The primary outcome measure of this study is the identification of pathogenic genetic variants that are causative for the movement disorder or dementia that the patient has been diagnosed with. These disease-causing variants are often inherited. The secondary outcome measure of this study is the identification of genetic variants that alter susceptibility/risk for the movement disorder or dementia that the patient has been diagnosed with. These genetic risk factors are associated with disease that can be apparently sporadic in nature. Identification of risk variants followed a case-control genome-wide association study design. As is appropriate for such studies, we assessed the outcome based on the variants that achieved or surpassed genome-wide significance, set at a p-value threshold of 5.0 x 10E-8. This threshold is standard in studies of this type. The threshold is described in the manuscript.