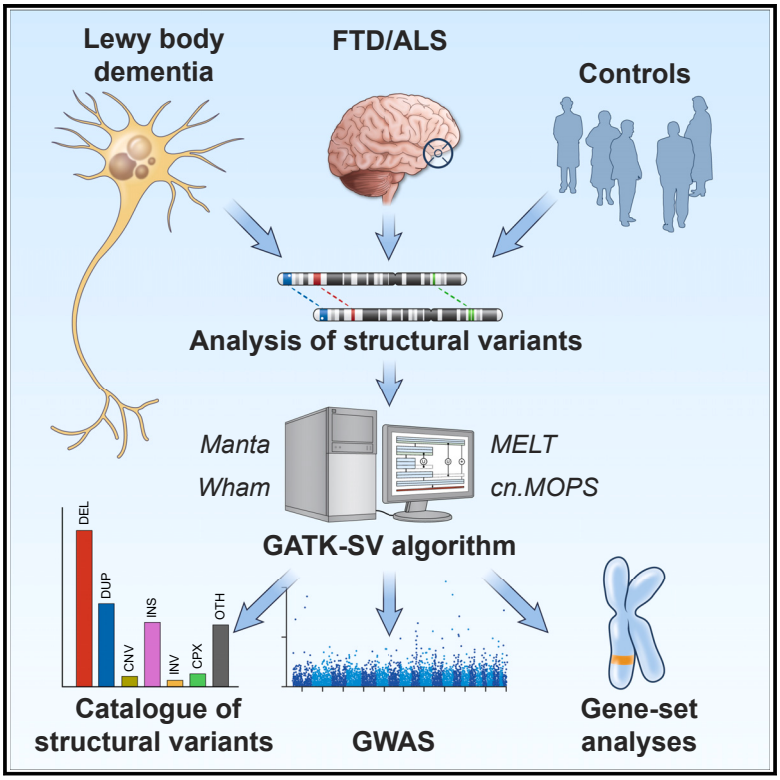
Cell Genomics

Resource

Genome-wide structural variant analysis identifies risk loci for non-Alzheimer's dementias

Graphical abstract



Highlights

- Structural variants were called in the genomes of non-Alzheimer's dementias
- Discovery of *TPCN1* as a novel risk locus for Lewy body dementia
- Structural variants at *C9orf72* and *MAPT* were present in frontotemporal dementia
- Gene-set analyses identified likely pathogenic rare structural variants

Authors

Karri Kaivola, Ruth Chia, Jinhui Ding, ..., J. Raphael Gibbs, Bryan J. Traynor, Sonja W. Scholz

Correspondence

sonja.scholz@nih.gov

In brief

This article from Kaivola, Chia, and Ding et al. describes the identification, characterization, and analysis of structural variants in genome data from patients with the non-Alzheimer's dementias Lewy body dementia (LBD) and frontotemporal dementia (FTD)/amyotrophic lateral sclerosis (ALS).

Kaivola et al., 2023, Cell Genomics *3*, 100316
June 14, 2023
https://doi.org/10.1016/j.xgen.2023.100316

CellPress

Cell Genomics

CellPress
OPEN ACCESS

Resource

Genome-wide structural variant analysis identifies risk loci for non-Alzheimer's dementias

Karri Kaivola,[1,53] Ruth Chia,[2,53] Jinhui Ding,[3,53] Memoona Rasheed,[2] Masashi Fujita,[4] Vilas Menon,[4] Ronald L. Walton,[5] Ryan L. Collins,[6,7,8] Kimberley Billingsley,[9,10] Harrison Brand,[6,7,8] Michael Talkowski,[6,7,11] Xuefang Zhao,[6,7] Ramita Dewan,[2] Ali Stark,[2] Anindita Ray,[1] Sultana Solaiman,[1] Pilar Alvarez Jerez,[9,10] Laksh Malik,[10] Ted M. Dawson,[12,13,14,15] Liana S. Rosenthal,[12] Marilyn S. Albert,[12] Olga Pletnikova,[16,17] Juan C. Troncoso,[17] Mario Masellis,[18,19,20,21] Julia Keith,[22] Sandra E. Black,[19,23,24,20,21] Luigi Ferrucci,[25] Susan M. Resnick,[26]

(Author list continued on next page)

[1]Neurodegenerative Diseases Research Unit, National Institute of Neurological Disorders and Stroke, Bethesda, MD, USA
[2]Neuromuscular Diseases Research Section, Laboratory of Neurogenetics, National Institute on Aging, Bethesda, MD, USA
[3]Computational Biology Group, Laboratory of Neurogenetics, National Institute on Aging, Bethesda, MD, USA
[4]Center for Translational & Computational Neuroimmunology, Department of Neurology, Columbia University Irving Medical Center and the Taub Institute for Research on Alzheimer's Disease and the Aging Brain, New York, NY, USA
[5]Department of Neuroscience, Mayo Clinic, 4500 San Pablo Road South, Jacksonville, FL, USA
[6]Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA
[7]Program in Medical and Population Genetics, Broad Institute of Massachusetts Institute of Technology (M.I.T.), Cambridge, MA, USA
[8]Division of Medical Sciences and Department of Medicine, Harvard Medical School, Boston, MA, USA
[9]Laboratory of Neurogenetics, National Institute on Aging, Bethesda, MD, USA
[10]Centre for Alzheimer's and Related Dementias, National Institute on Aging, Bethesda, MD, USA
[11]Department of Neurology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA
[12]Department of Neurology, Johns Hopkins University Medical Center, Baltimore, MD, USA
[13]Neuroregeneration and Stem Cell Programs, Institute of Cell Engineering, Johns Hopkins University Medical Center, Baltimore, MD, USA
[14]Department of Pharmacology and Molecular Science, Johns Hopkins University Medical Center, Baltimore, MD, USA
[15]Solomon H. Snyder Department of Neuroscience, Johns Hopkins University Medical Center, Baltimore, MD, USA
[16]Department of Pathology and Anatomical Sciences, Jacobs School of Medicine and Biomedical Sciences, University of Buffalo, Buffalo, NY, USA
[17]Department of Pathology (Neuropathology), Johns Hopkins University Medical Center, Baltimore, MD, USA
[18]Cognitive & Movement Disorders Clinic, Sunnybrook Health Sciences Centre, University of Toronto, 1 King's College Circle, Room 2374, Toronto, ON, Canada
[19]Division of Neurology, Department of Medicine, University of Toronto, Toronto, ON, Canada
[20]Hurvitz Brain Sciences Research Program, Sunnybrook Research Institute, University of Toronto, 2075 Bayview Avenue, Toronto, ON, Canada
[21]LC Campbell Cognitive Neurology Research Unit, Sunnybrook Research Institute, University of Toronto, 2075 Bayview Avenue, Toronto, ON, Canada

(Affiliations continued on next page)

SUMMARY

We characterized the role of structural variants, a largely unexplored type of genetic variation, in two non-Alzheimer's dementias, namely Lewy body dementia (LBD) and frontotemporal dementia (FTD)/amyotrophic lateral sclerosis (ALS). To do this, we applied an advanced structural variant calling pipeline (GATK-SV) to short-read whole-genome sequence data from 5,213 European-ancestry cases and 4,132 controls. We discovered, replicated, and validated a deletion in *TPCN1* as a novel risk locus for LBD and detected the known structural variants at the *C9orf72* and *MAPT* loci as associated with FTD/ALS. We also identified rare pathogenic structural variants in both LBD and FTD/ALS. Finally, we assembled a catalog of structural variants that can be mined for new insights into the pathogenesis of these understudied forms of dementia.

INTRODUCTION

Structural variants are duplicated, deleted, inserted, inverted, or translocated DNA segments measuring at least 50 base pairs in length by current definition. This class of genomic variation is a major source of genetic diversity in the human genome that contributes to phenotypic heterogeneity.[1,2] Not surprisingly, structural variants have been implicated in the pathogenesis of complex neurological disorders. For example, *APP* duplications,[3,4] *SNCA* multiplications,[5,6] and recombination events at

Toshiko Tanaka,[25] The American Genome Center, International LBD Genomics Consortium, International ALS/FTD Consortium, PROSPECT Consortium, Eric Topol,[27] Ali Torkamani,[27] Pentti Tienari,[28,29] Tatiana M. Foroud,[30] Bernardino Ghetti,[31] John E. Landers,[32] Mina Ryten,[33,34,35] Huw R. Morris,[36,37] John A. Hardy,[36,37,38,39] Letizia Mazzini,[40] Sandra D'Alfonso,[41] Cristina Moglia,[42,43] Andrea Calvo,[42,43] Geidy E. Serrano,[44] Thomas G. Beach,[44] Tanis Ferman,[45] Neill R. Graff-Radford,[46] Bradley F. Boeve,[47] Zbigniew K. Wszolek,[48] Dennis W. Dickson,[5] Adriano Chiò,[42,43,48] David A. Bennett,[49] Philip L. De Jager,[4] Owen A. Ross,[5,54] Clifton L. Dalgard,[50,51,54] J. Raphael Gibbs,[3,54] Bryan J. Traynor,[2,12,52,54] and Sonja W. Scholz[1,12,54,55,*]

[22]Department of Anatomical Pathology, Sunnybrook Health Sciences Centre, University of Toronto, 1 King's College Circle, Room 2374, Toronto, ON, Canada
[23]Institute of Medical Science, Faculty of Medicine, University of Toronto, 1 King's College Circle, Room 2374, Toronto, ON, Canada
[24]Heart and Stroke Foundation Canadian Partnership for Stroke Recovery, Sunnybrook Health Sciences Centre, University of Toronto, 1 King's College Circle, Room 2374, Toronto, ON, Canada
[25]Longitudinal Studies Section, National Institute on Aging, Baltimore, MD, USA
[26]Laboratory of Behavioral Neuroscience, National Institute on Aging, Baltimore, MD, USA
[27]Scripps Research Translational Institute, Scripps Research, La Jolla, CA, USA
[28]Translational Immunology, Research Programs Unit, University of Helsinki, Helsinki, Finland
[29]Department of Neurology, Helsinki University Hospital, Helsinki, Finland
[30]Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN, USA
[31]Department of Pathology and Laboratory Medicine, Indiana University School of Medicine, Indianapolis, IN, USA
[32]Department of Neurology, University of Massachusetts Medical School, Worcester, MA, USA
[33]Department of Genetics and Genomic Medicine Research & Teaching, UCL GOS Institute of Child Health, University College London, London, UK
[34]Department of Neurodegenerative Disease, Queen Square Institute of Neurology, University College London, London, UK
[35]NIHR Great Ormond Street Hospital Biomedical Research Centre, University College London, London, UK
[36]Department of Clinical and Movement Neurosciences, UCL Queen Square Institute of Neurology, University College London, London, UK
[37]UCL Movement Disorders Centre, University College London, London, UK
[38]UK Dementia Research Institute, Department of Neurogenerative Disease and Reta Lila Weston Institute, London, UK
[39]Institute of Advanced Study, The Hong Kong University of Science and Technology, Hong Kong SAR, China
[40]Maggiore della Carita University Hospital, Novara, Italy
[41]Department of Health Sciences, University of Eastern Piedmont, Novara, Italy
[42]Rita Levi Montalcini Department of Neuroscience, University of Turin, Turin, Italy
[43]Azienda Ospedaliero Universitaria Citta, della Salute e della Scienza, Corso Bramante, 88, Turin, Italy
[44]Civin Laboratory for Neuropathology, Banner Sun Health Research Institute, Sun City, AZ, USA
[45]Department of Psychiatry and Psychology, Mayo Clinic, 4500 San Pablo Road South, Jacksonville, FL, USA
[46]Department of Neurology, Mayo Clinic, 4500 San Pablo Road South, Jacksonville, FL, USA
[47]Center for Sleep Medicine, Mayo Clinic, Rochester, MN, USA
[48]Institute of Cognitive Sciences and Technologies, C.N.R., Via S. Martino della Battaglia, 44, Rome, Italy
[49]Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, IL, USA
[50]Department of Anatomy, Physiology and Genetics, Uniformed Services University of the Health Sciences, Bethesda, MD, USA
[51]The American Genome Center, Collaborative Health Initiative Research Program, Uniformed Services University of the Health Sciences, Bethesda, MD, USA
[52]RNA Therapeutics Laboratory, Therapeutics Development Branch, National Center for Advancing Translational Sciences, Rockville, MD, USA
[53]These authors contributed equally
[54]These authors contributed equally
[55]Lead contact
*Correspondence: sonja.scholz@nih.gov
https://doi.org/10.1016/j.xgen.2023.100316

the *GBA* locus[7] are rare causes of Lewy body dementia (LBD), the second most common form of neurodegenerative dementia after Alzheimer's disease. The discovery of repeat expansions in *C9orf72* unified frontotemporal dementia (FTD) and amyotrophic lateral sclerosis (ALS) into the same clinical spectrum.[8–10] These findings highlight the pathogenic role of structural variants in disease and suggest that it may account for at least some of the unexplained heritability in age-related neurological conditions.

To date, identifying the structural variants underlying disease has relied on candidate-gene studies rather than unbiased genome-wide assessments. More recently, however, the study of structural variants has gained momentum due to the improved availability of whole-genome sequence datasets and modernized detection algorithms.[11–13] These efforts are enhanced by the publication of structural variant catalogs, providing crucial insights into the frequency and heterogeneity of this understudied variant type in the human population.[11,14–16]

Here, we applied these advancements to systematically map genomic structural variants in large cohorts of patients diagnosed with non-Alzheimer's dementias (LBD and FTD/ALS) and in unaffected controls.[12,13] We exploited this resource to perform genome-wide association studies (GWASs) to identify common structural variants acting as risk modulators and identify rare, pathogenic structural variants in neurodegenerative disease genes. Our work discovered structural variants associated with risk of developing these fatal neurological conditions,

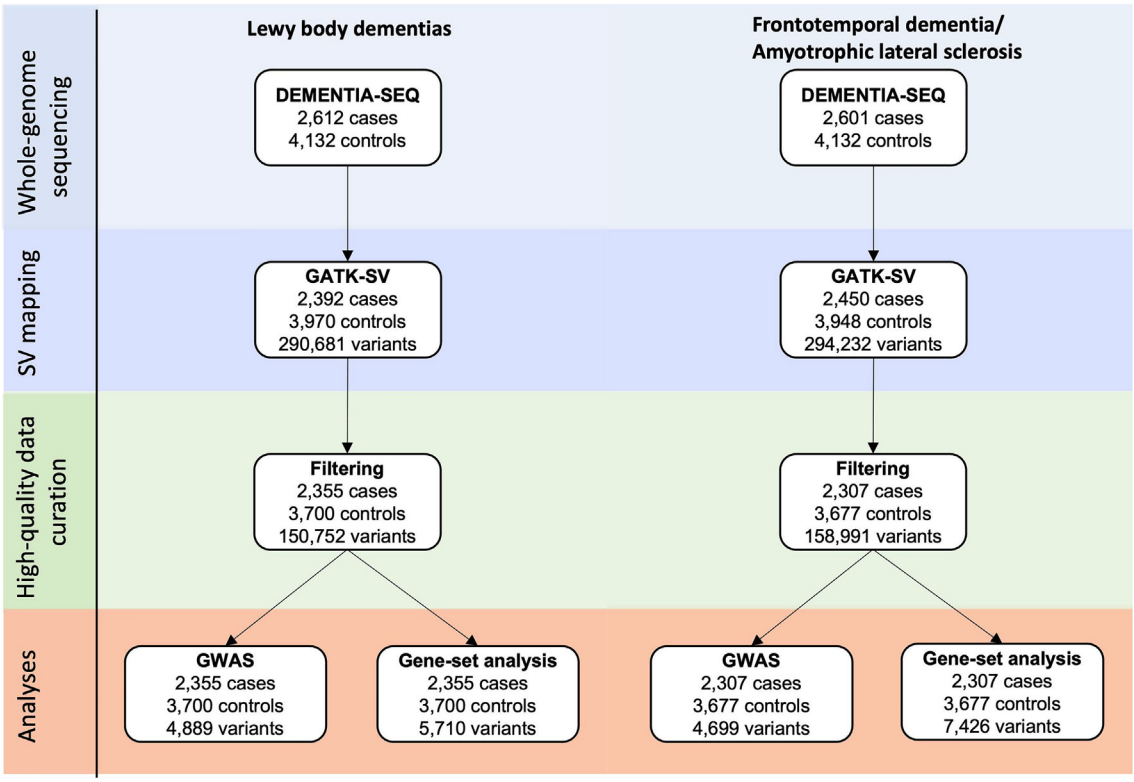


Figure 1. Study flowchart

Schematic illustration of the analytical workflow. The same controls were used in the analysis of LBD and FTD/ALS.

confirming the critical role of this variant class in neurodegeneration. Building on our work, we compiled an interactive resource that can be investigated by other researchers for new insights into non-Alzheimer's dementias.

RESULTS

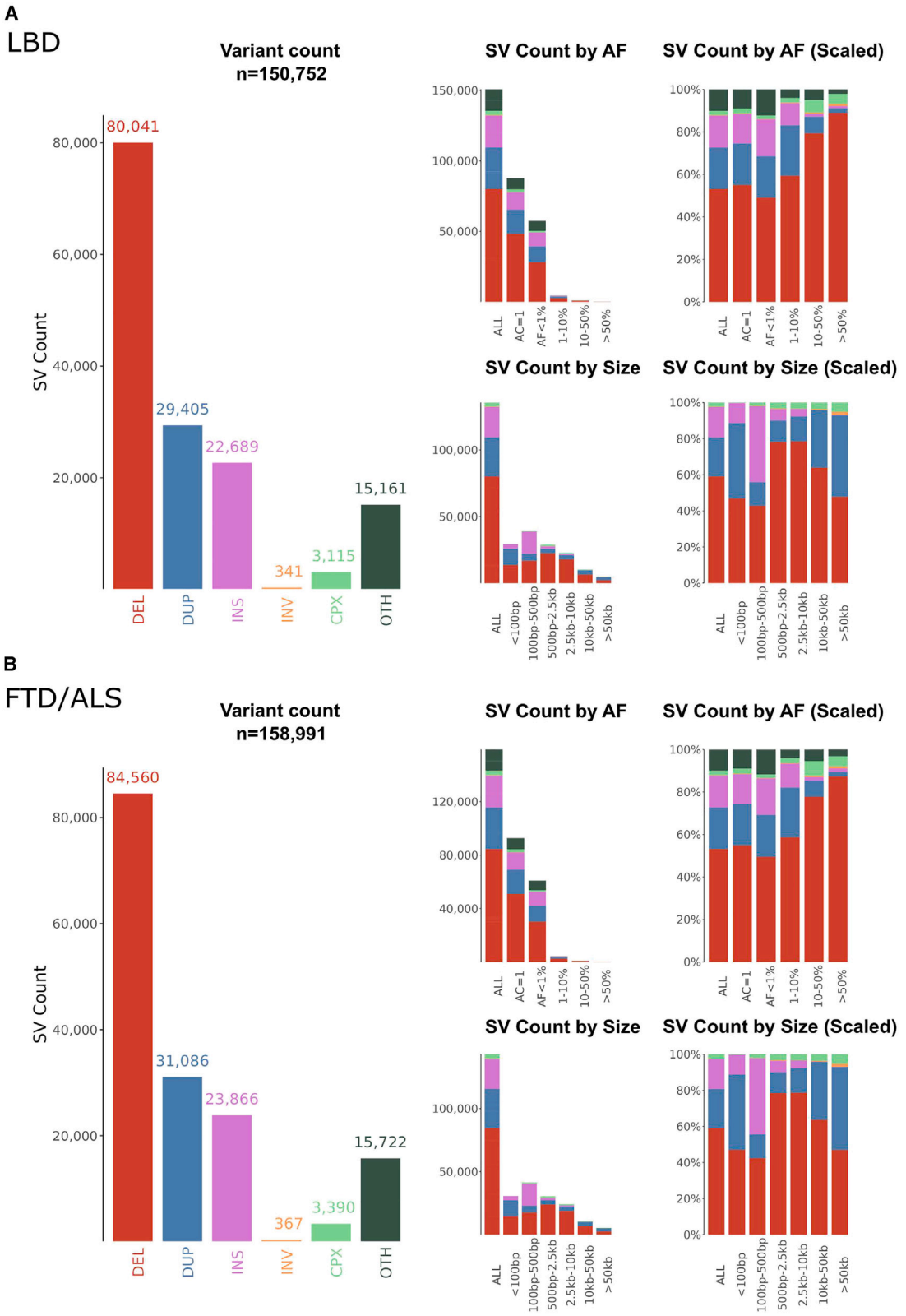
Structural variant identification in non-Alzheimer's dementias

We mapped structural variants in whole-genome sequence data obtained from patients diagnosed with LBD (n = 2,612), FTD/ALS (n = 2,601), and unaffected individuals (n = 4,132) using the Genome Analysis Toolkit's Structural Variant (GATK-SV) pipeline.[11–13] This method improves the identification of structural variants in Illumina short-read whole-genome sequencing data by applying machine learning to combine information from five detection algorithms. Figure 1 summarizes the study workflow.

After applying stringent quality control and filtering steps, there were 150,752 structural variants in the 2,355 LBD cases and 3,700 controls available for the association analysis. Similarly, there were 158,991 structural variants in the 2,307 FTD/ALS cases and 3,677 controls. In the LBD case-control cohort, there were on average 895 autosomal structural variants per participant. In the FTD/ALS case-control cohort, there were 865 structural variants on average per participant (Figure 2 and Figures S1–S3 for descriptive statistics of filtered variants and Figures S4–S6 for summaries of unfiltered variants). The median structural variant length was 467 base pairs in the LBD case-control cohort and 479 base pairs in the FTD/ALS case-control cohort. As expected, the most common structural variants were deletions (53.1% of all structural variants in the LBD case-control cohort and 53.2% in the FTD/ALS case-control cohort) or duplications (19.5% and 19.6%). Further, most of the structural variants were rare (minor allele frequency [MAF] <1%; 96.3% of the observed structural variants in the LBD case-control cohort and 96.6% in the FTD/ALS cohort). The structural variant type, size, and allele frequency did not markedly differ between the cases and controls for LBD or FTD/ALS (Table S1).

VALIDATION OF STRUCTURAL VARIANTS USING LONG-READ SEQUENCING

We compared structural variants mapped from short-read genome sequence data using the GATK-SV pipeline to long-read Nanopore sequencing data generated for 20 samples. This analysis was performed by applying Truvari bench,[17] an algorithm that correlates structural variants based on type, size, and location. The validation rates of the raw data were consistent with prior observations and varied according to structural variant type, allele frequency, and quality.[18–20] After the application of stringent quality filters, the highest concordance estimate was observed for deletions (mean validation rate = 84.3% and genotype concordance = 94.3%), whereas duplications had the



(legend on next page)

lowest concordance rate (50.2% and 89.2%; Table S2). Common structural variants (MAF $\geq$ 1%) were more likely to be confirmed by long-read sequencing (n = 5,101 structural variants identified in the 20 participants; the validation rates were 84.7%, 52.5%, 61.6%, and 53.2% for deletions, duplications, insertions, and inversions, respectively). The validation rates for rare structural variants (MAF < 1%) were not as robust (n = 1,917 rare structural variants identified by Nanopore long-read sequencing in the 20 samples; the validation rates were 79.5%, 43.3%, 61.4%, and 0.0% for deletions, duplications, insertions, and inversions).

Genome-wide association studies of common structural variants in LBD and FTD/ALS

After quality control filtering, we performed GWASs on common (i.e., MAF $\geq$ 1%) high-quality variants separately for the LBD and FTD-ALS case-control cohorts. In total, 4,899 structural variants were tested for association with LBD, and 4,699 variants were tested for FTD/ALS.

In the FTD/ALS structural variant GWAS, we identified two statistically significant signals (Table 1; Figure 3). The first structural variant was located on chromosome 9p21, corresponding to the *C9orf72* hexanucleotide repeat expansion ($p = 4.99 \times 10^{-18}$, odds ratio [OR] = 14.47, 95% confidence interval [CI] = 7.90–26.49). The second association signal was due to the 673-kb complex, common structural variant in the *MAPT* locus on the long arm of chromosome 17 that is part of the H2 haplotype ($p = 3.48 \times 10^{-6}$, OR = 0.77, 95% CI = 0.68–0.86). Both loci are known risk loci for FTD/ALS,[10,21] representing an internal validation of the GATK-SV pipeline.

In the LBD structural variant GWAS, we identified a novel locus on chromosome 12q24.13 that reached Bonferroni-corrected statistical significance ($p = 9.2 \times 10^{-6}$, OR = 1.43, 95% CI = 1.22–1.67; Figure 4A). This association signal was driven by a 309-base-pair deletion that we validated using long-read sequencing (95.0% allele concordance rate; Table 1; Figure 4B). The deletion was located within intron 2 of *TPCN1* (NM_017901.6; Figure 4C), a gene encoding the two-pore segment channel 1 that is highly expressed in neurons and glia. The same locus had previously been reported in an Alzheimer's disease GWAS.[22] To further explore this relationship, we generated a beta-beta plot of the *TPCN1* locus to compare the association signals in LBD (obtained from our previously reported single-nucleotide variant GWAS[12]) with a published Alzheimer's disease GWAS.[22] This showed that the association signals at the *TPCN1* locus were identical (Figure 5A). However, no such relationship was detected between the *TPCN1* locus association in LBD versus a recent Parkinson's disease GWAS[23] (Figure 5B).

Replication of *TPCN1* deletion in LBD

We identified the *TPCN1* deletion in a replication cohort consisting of an additional 555 LBD cases and 274 control subjects. The MAF was 6.9% in these cases and 4.9% in the control samples, similar to what was observed in the discovery cohort. Furthermore, the *TPCN1* deletion was significantly associated with LBD in this cohort (p = 0.024, OR = 2.46, 95% CI = 1.12–5.36), representing an independent replication of the novel risk locus. Meta-analysis of the discovery and replication cohorts showed the same direction of effect ($p = 1.64 \times 10^{-6}$, OR = 1.46, 95% CI = 1.25–1.70). The *TPCN1* deletion associated with an increased risk of LBD was tagged by the single-nucleotide polymorphism rs6489896 (chr12:113,281,983:T>C; $r^2 = 0.95$, $D' = 0.98$), and a statistically significant association was also observed with this variant in the replication dataset (p = 0.024, OR = 2.46, 95% CI = 1.13–5.37).

Functional effect of the chromosome 12q24.13 deletion

In the Genotype-Tissue Expression (GTEx Analysis Research v8) database,[24,25] rs6489896 was an expression quantitative trait locus (eQTL) for the *RITA1* gene in the brain, with decreased expression levels in all of the evaluated brain regions (e.g., cortex $p = 7.3 \times 10^{-11}$, normalized effect size = −0.41). A single-nucleus, cell-type-level RNA sequencing dataset based on 424 brains of the Religious Orders Study/Memory and Aging Project (ROS/MAP) collection confirmed these findings[26]; only *RITA1* gene expression in excitatory neurons had a statistically significant association with rs6489896 ($p = 9.80 \times 10^{-8}$, beta = −0.35; Figure 5C; Table S4). While these observations suggest that the deletion exerts a functional effect on the neighboring *RITA1* gene, the location of the structural variant within intron 2 of *TPCN1* means that we cannot exclude the possibility that the deletion is also influencing that gene. In that regard, it is interesting to note that rs6489896 was a splicing quantitative trait locus for *TPCN1* in GTEx with the strongest association observed in subcutaneous adipose tissue ($p = 9.80 \times 10^{-40}$, normalized effect value −1.20). A similar effect was not observed in various brain regions, although this may reflect the small sample size limiting the power to detect such effects.

Identification of rare, disease-causing structural variants in LBD and FTD/ALS

Our cohort was not sufficiently powered to perform gene-burden analysis, as most of the genes did not have more than a single exonic structural variant. Nevertheless, to explore the role of rare structural variants in the pathogenesis of non-Alzheimer's dementias, we examined their occurrence in 50 genes previously implicated in familial neurodegenerative diseases (see Table S5 for a list of the genes). Our analysis identified 83 and 81 exonic structural variants that mapped to these established neurodegenerative disease genes in the LBD and FTD/ALS cohorts, respectively (Table S6). Among these, we found a number of structural variants that are known disease-causing mutations. These included an *SNCA* duplication and an *OPTN* deletion, each detected in a single LBD case.

Figure 2. Descriptive statistics of the structural variants in the study cohorts

Descriptive statistics of high-quality subset of structural variants mapped by the GATK-SV pipeline in the (A) LBD case-control cohort and (B) FTD/ALS case-control cohort. AF, allele frequency; DEL, deletions; DUP, duplications; INS, insertions; INV, inversions; CPX, complex variant; OTH, other variant; SV, structural variant.

Table 1. Significantly associated structural variants in the LBD and FTD/ALS GWAS analyses

	Chr	Position	Gene	Type	Length (bp)	MAF cases	MAF controls	p value	OR (95% CI)
LBD	12	113,245,316	*TPCN1*	deletion	309	0.075	0.052	9.10×10^{-6}	1.43 (1.22–1.67)
FTD/ALS	9	27,573,524	*C9orf72*	unresolved[a]	NA	0.032	0.0018	4.99×10^{-18}	14.47 (7.90–26.49)
FTD/ALS	17	45,603,799	*MAPT*	complex inversion	673,211	0.17	0.19	3.48×10^{-6}	0.77 (0.68–0.86)

The Bonferroni threshold of significance was 1.02×10^{-5} (= 0.05/4,889 structural variants with an MAF $\geq$ 1%) for the LBD GWAS and 1.06×10^{-5} (=0.05/4,699) for the FTD/ALS GWAS. Chromosome positions are displayed according to reference genome build hg38. Chr, chromosome; LBD, Lewy body dementia; FTD, frontotemporal dementia; ALS, amyotrophic lateral sclerosis; bp, base pairs; MAF, minor allele frequency; OR, odds ratio; CI, confidence interval; NA, not applicable.

[a]Confirmed to refer to the hexanucleotide repeat expansion in the *C9orf72* gene.

In the FTD/ALS case-control cohort, we identified a *LRRK2* duplication in one case with a clinical diagnosis of non-fluent variant primary progressive aphasia. This form of dementia is most commonly a manifestation of a four-repeat tauopathy upon pathological evaluation.[27] Mutations in *LRRK2* have been identified as rare causes of four-repeat tauopathies,[3,28] suggesting that this *LRRK2* duplication could be disease causing. Additionally, we identified a heterozygous deletion of the promoter and the first two exons of *CHCHD10* in one patient with FTD and motor neuron disease. Heterozygous missense mutations in *CHCHD10*, especially in exon 2, are a known cause of FTD/ALS,[29–31] and a loss-of-function mechanism has been proposed.[32] We also found a deletion of the last exon and 3′-untranslated region of *FIG4* in one patient diagnosed with FTD. Heterozygous missense mutations in *FIG4*, including in exon 23, have been linked to FTD/ALS,[33] and our case provides supportive evidence of their pathogenicity. Table 2 summarizes the clinical and pathological features of these cases.

Resource for exploring structural variants in non-Alzheimer's dementias

The structural variant dataset generated in this study is among the largest published to date in neurological disorders. As such, it represents a valuable resource that can be exploited for further research. We prioritized making our structural variant data easily accessible and user friendly. The raw sequence data are available in dbGaP (study accession phs001963.v1.p1), and details of the individual structural variants are provided in Tables S7 and S8. Moreover, we developed an online resource that enables researchers to explore the structural variant data.

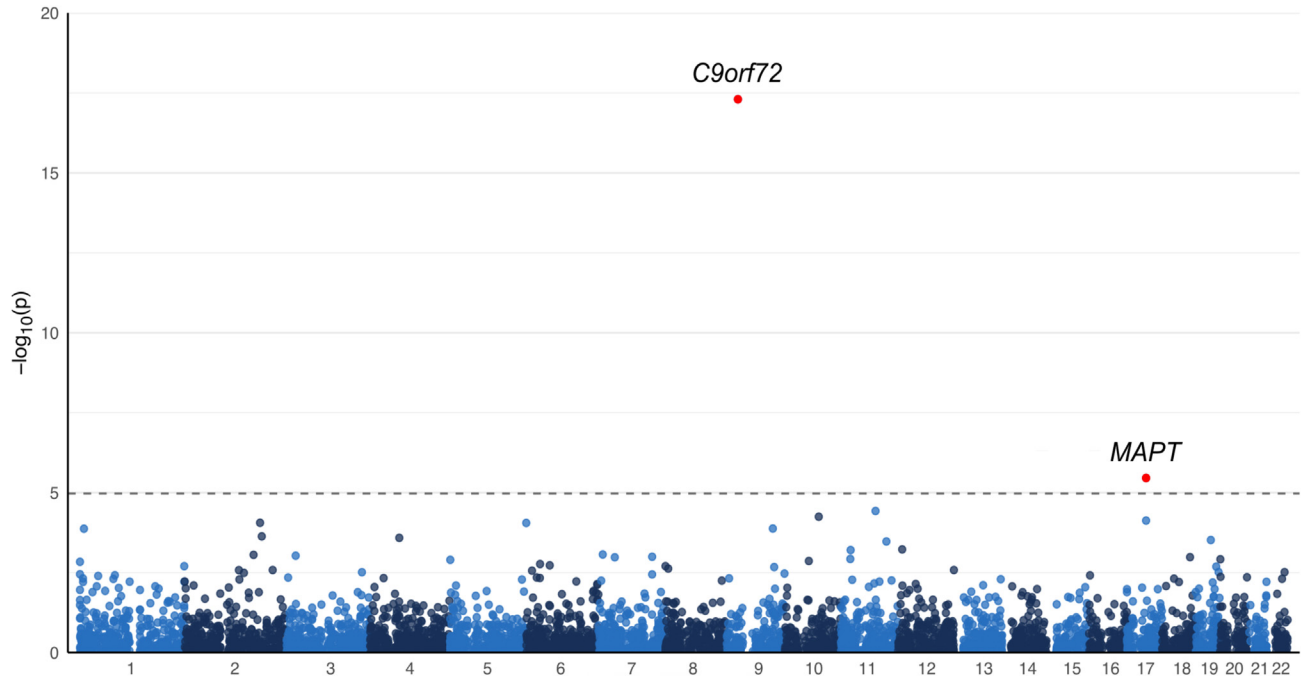
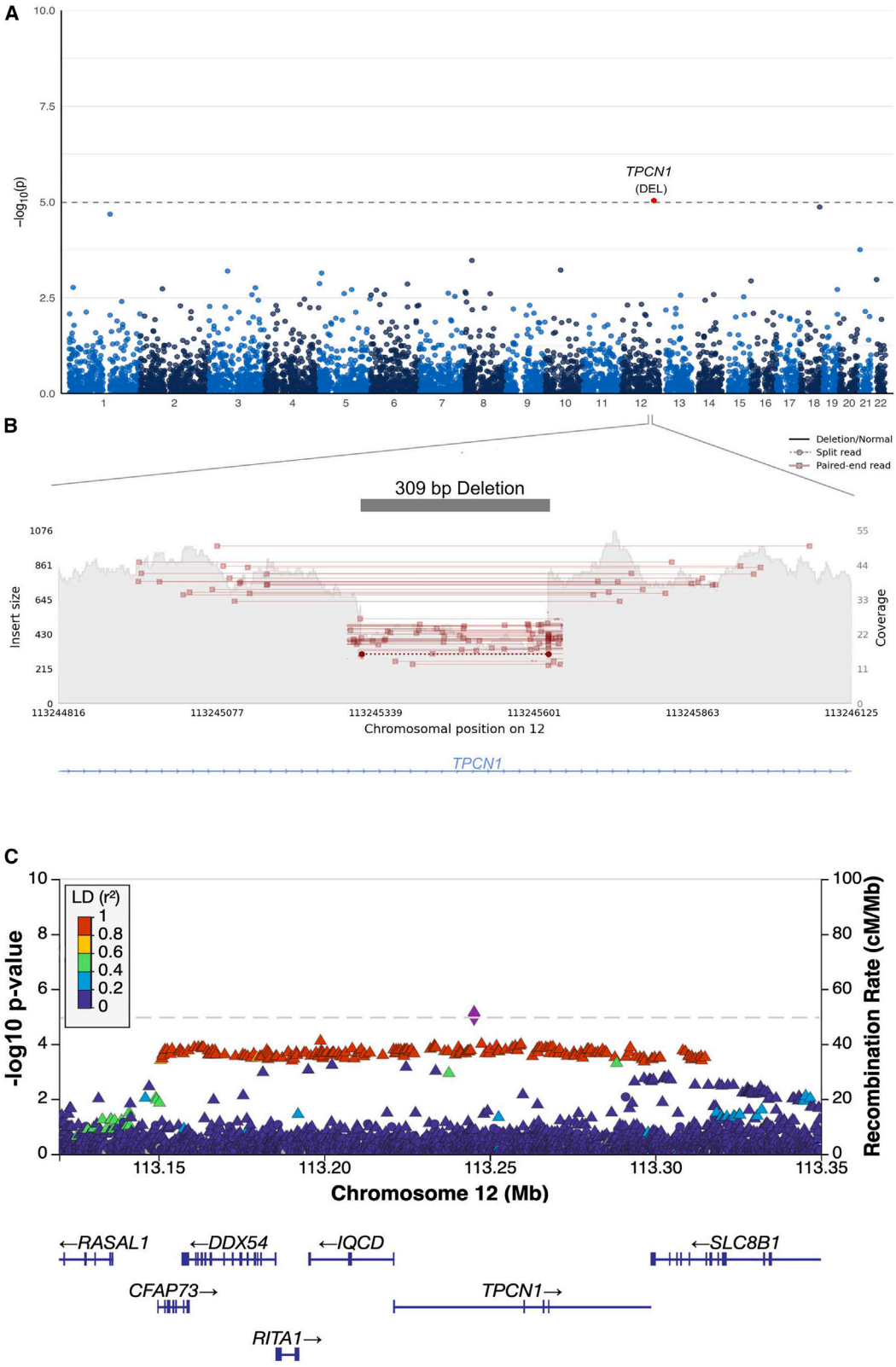


Figure 3. Genome-wide representation of the common structural variants in the FTD/ALS case-control cohort

Manhattan plot depicting the GWAS results (n = 2,307 FTD and 3,677 controls, and 4,699 structural variants with MAF $\geq$ 1%). The annotations show the break-end variant on chromosome 9, corresponding to the *C9orf72* hexanucleotide repeat expansion, and a 673-kb complex inversion that is part of the *MAPT* H2 haplotype.



(legend on next page)

For example, the data can be filtered by several parameters, including location, gene name, structural variant type, overlapping genetic region, structural variant quality, case-control allele frequency, and combined annotation-dependent depletion (CADD) structural variant score (https://ndru-ndrs-lng-nih.shinyapps.io/non_ad_dementias_sv_app/; Figure 6A). The online resource also allows visualization of structural variants in a genomic context. Moreover, we created BigBed files of structural variants that can be viewed, e.g., in Ensembl and UCSC genome browsers as custom tracks (Figure 6B). This allows the structural variants to be explored with other available tracks, such as regulatory regions, known pathological variants, and gnomAD structural variants. We have made our programming code publicly available so that other researchers can apply it and modify it to match their needs (https://github.com/ruthchia/Structural_variant_analysis-LBD-FTD; https://doi.org/10.5281/zenodo.7796321).

DISCUSSION

In this study, we investigated the role of structural variants in LBD and FTD/ALS, two forms of non-Alzheimer's dementia. We applied a multi-algorithm pipeline to identify, characterize, and analyze structural variants and to create a resource.[11] We conducted GWASs and neurodegenerative disease gene-set evaluations to elucidate the role of these variants in complex diseases. In this process, we identified a common structural variant in *TPCN1* that modulates disease risk, a previously known pathogenic variant in *C9orf72*, a common complex variant at the *MAPT* locus, and likely pathogenic, rare variants in several other neurodegenerative disease genes.

In the FTD/ALS structural variant GWAS, we identified a breakend mutation at chromosome 9p21.2, corresponding to the *C9orf72* hexanucleotide repeat expansion, and a 673-kb complex inversion related to the *MAPT* H2 haplotype. Both are structural variants known to be associated with FTD/ALS, demonstrating the ability of the analysis pipeline to capture this type of disease-associated genomic variation.

In the LBD structural variant GWAS, we identified, validated, and replicated a significant association with a 309-base-pair deletion that overlaps intron 2 of the canonical transcript of *TPCN1*. *TPCN1* encodes a voltage-dependent calcium channel that is expressed throughout the body[34] and located at the endolysosomal membranes.[35–37] Interestingly, TPCN1 has been reported to be essential for long-term potentiation in hippocampal neurons,[38] and *Tpcn1* knockout mice have been shown to have impaired memory and spatial learning.[39] Furthermore, loss of presenilins decreases lysosomal calcium stores, which, in turn, alters TPCN1 levels by decreasing the functional dimer form.[40]

The potential role of the *TPCN1* locus in dementia is supported by a recent GWAS in Alzheimer's disease that reported a suggestive association between the intronic *TPCN1* variant rs6489896 and Alzheimer's disease.[22] This rs6489896 variant tagged the 309-base-pair *TPCN1* deletion in our data, and there was a near-perfect correlation between the haplotypes in the Alzheimer's disease GWAS and our study (Figure 5A). These data indicate that the *TPCN1* deletion is not specific to LBD. The suggestive association between *TPCN1* and Alzheimer's disease could be due to the well-established shared etiology between Alzheimer's disease and LBD, or a subpopulation of LBD-variant Alzheimer's disease patients. Interestingly, the *TPCN1* locus did not correlate with Parkinson's disease (Figure 5B). Furthermore, functional genomic evaluations based on bulk RNA sequencing and single-nuclear RNA sequencing data showed that the *TPCN1* deletion-tagging rs6489896 variant strongly influences the expression of *RITA1*, a gene located centromeric to *TPCN1* on the long arm of chromosome 12. RITA1 modulates Notch-signaling,[41] although its exact functions remain elusive. Based on these preliminary data, we cannot rule out that other genes or a combination of genes within the *TPCN1* locus may also influence disease risk.

We identified several rare structural variants of interest in our analysis of the neurodegenerative disease genes. *SNCA* duplications are an established rare cause of Parkinson's disease and LBD,[42] while *OPTN* deletions are a known cause of FTD[43] and ALS,[44] indicating that the observed structural variants are disease causing. Furthermore, optineurin, the protein encoded by *OPTN*, is deposited in Lewy bodies,[45] the pathological hallmark of LBD, suggesting a molecular link between this pathogenic mutation and the LBD pathogenesis.

In conclusion, we used a multi-algorithm pipeline to create a resource of structural variants in LBD and FTD/ALS. We identified and validated common and rare structural variants driving disease risk in these neurological disorders and presented a resource that can be mined for further insights into this type of genomic variation. Our study highlighted the utility of using short-read whole-genome sequencing for structural variant discovery and demonstrated the value of studying this type of genomic variation to understand the underlying pathogenesis of non-Alzheimer's dementias.

LIMITATIONS OF THE STUDY

The main limitations of our study stem from the inherent difficulty of calling structural variants from short-read whole-genome sequencing data. As such, the validation rate of structural variants detected in short-read sequencing data is not ideal. To mitigate this problem, we used a multi-algorithm pipeline, GATK-SV,[11] to create consensus structural variant calls and focused on a subset of high-quality structural variants in our analyses. Overall, the mean number of structural variant sites, the distribution of structural variant types, and the

Figure 4. Genome-wide representation of the common structural variants in LBD

Figure360▷ For a Figure360 author presentation of this figure, see https://doi.org/10.1016/j.xgen.2023.100316.

(A) Manhattan plot depicting the GWAS results (n = 2,355 LBD cases and 3,700 controls; 4,889 structural variants with MAF > 1%).

(B) An illustration of the *TPCN1* deletion in a heterozygous carrier. Paired-end reads that did not span the deletion have been omitted for clarity.

(C) The *TPCN1* deletion (purple diamond) is the top association signal in the LBD analysis and shows strong linkage disequilibrium with nearby single-nucleotide variants.

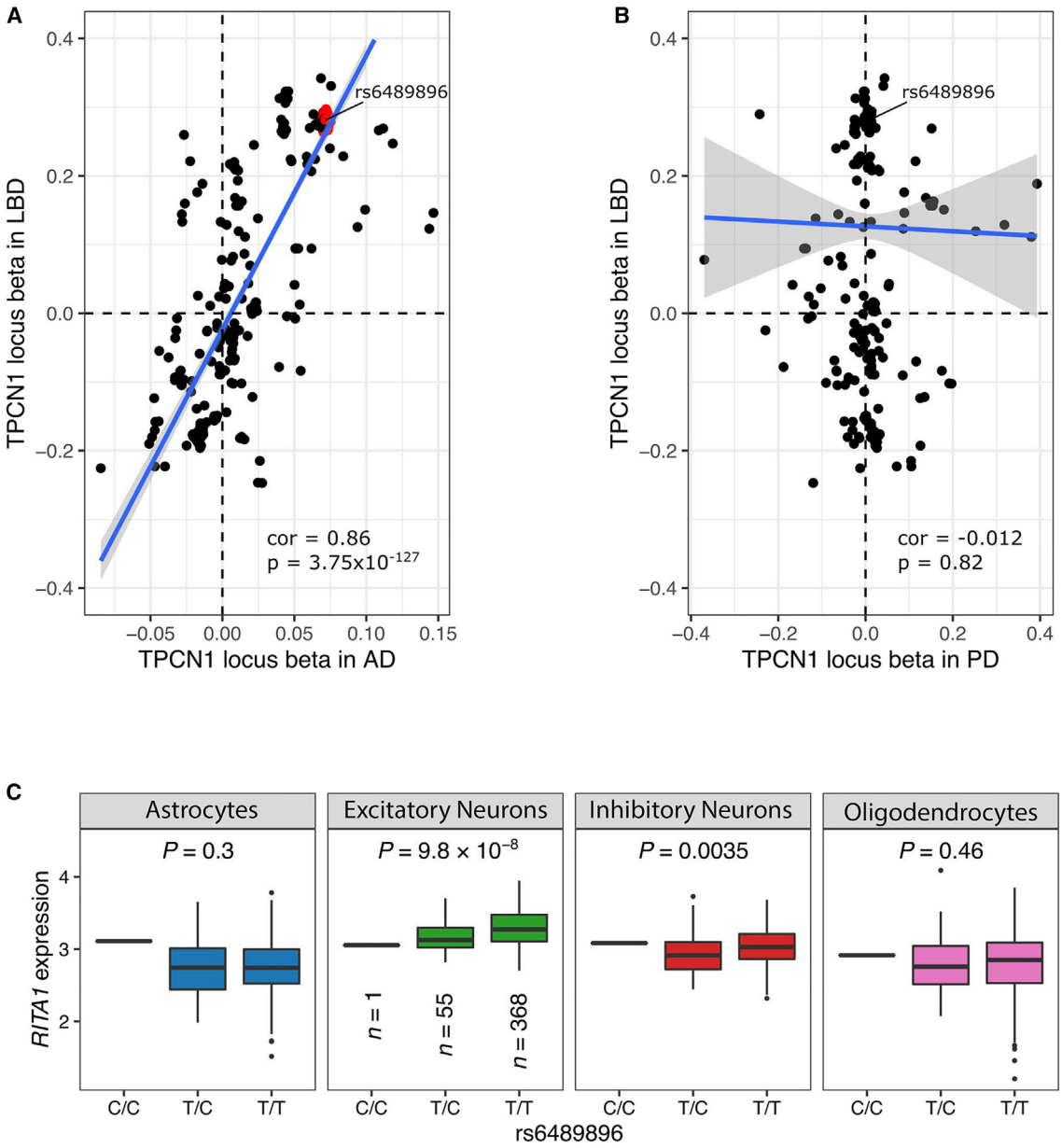


Figure 5. Genomic fine-mapping of the *TPCN1* locus

(A) Scatterplot showing the beta regression coefficients of common variants (MAF ≥ 1%) at the *TPCN1* locus in the Chia et al. LBD GWAS (y axis)[12] and Bellenguez et al. Alzheimer's disease GWAS (x axis).[22]

(B) Beta-beta plot comparing the *TPCN1* locus variants in the Chia et al. LBD GWAS (y axis)[12] with the Nalls et al. Parkinson's disease GWAS results (excluding 23&Me data; x axis).[23]

(C) Single-nucleus RNA sequence analysis of the *TPCN1* deletion-tagging SNP rs6489896 identified a significant *cis*-eQTL for *RITA1* in excitatory neurons.

proportion of structural variant calls in Hardy-Weinberg equilibrium are within the expected limits. Moreover, we found good structural variant mapping precision and genotype concordance between short-read sequencing and long-read sequencing data (Table S2). These findings indicate that our structural variant calls are robust. However, there are no accepted filtering standards following the GATK-SV pipeline, and direct comparison with other studies employing different filters can be complex (Table S3). It is worth noting that breakpoint locations in structural variant calls are rarely exact. This issue also holds true when using a multi-algorithm pipeline that collapses sufficiently overlapping variants into a single structural variant. Therefore, structural variants where the functional impact is due to breakpoint location should be

Table 2. Pathogenic, rare structural variants in the LBD and FTD/ALS case-control cohorts

	Chr	Position	Type	Length (bp)	MAF cases	MAF controls	Overlapping regions	Patient characteristic
SNCA	4	89,123,177	duplication	1,347,138	0.00021	0	whole gene	pathologically diagnosed LBD
OPTN	10	12,982,796	deletion	314,105	0.00021	0	whole gene	pathologically diagnosed LBD
FIG4	6	109,822,689	deletion	8,753	0.00021	0	exon 23, 3′ UTR	non-fluent variant PPA
LRRK2	12	40,073,033	duplication	327,105	0.00021	0	whole gene	non-fluent variant PPA
CHCHD10	22	23,767,017	deletion	2,653	0.00021	0	promoter, exons 1–2	FTD (with motor neuron disease)

Chromosome positions are displayed according to reference genome build hg38. PPA, primary progressive aphasia.

interpreted cautiously. Moreover, structural variants mapped to low-complexity repeat regions require validation since these regions are technically problematic to sequence, and the current human reference genome (GRCh38) can be incomplete at these loci. Another challenge when studying structural variants is the assessment of pathogenicity and unraveling the mechanisms by which they disrupt neuronal function. As previous data are scarce and the consequences of structural variants are not well understood, cell biology studies, especially those integrating other multi-omic data, are needed to fully understand the consequences of this mutation class.

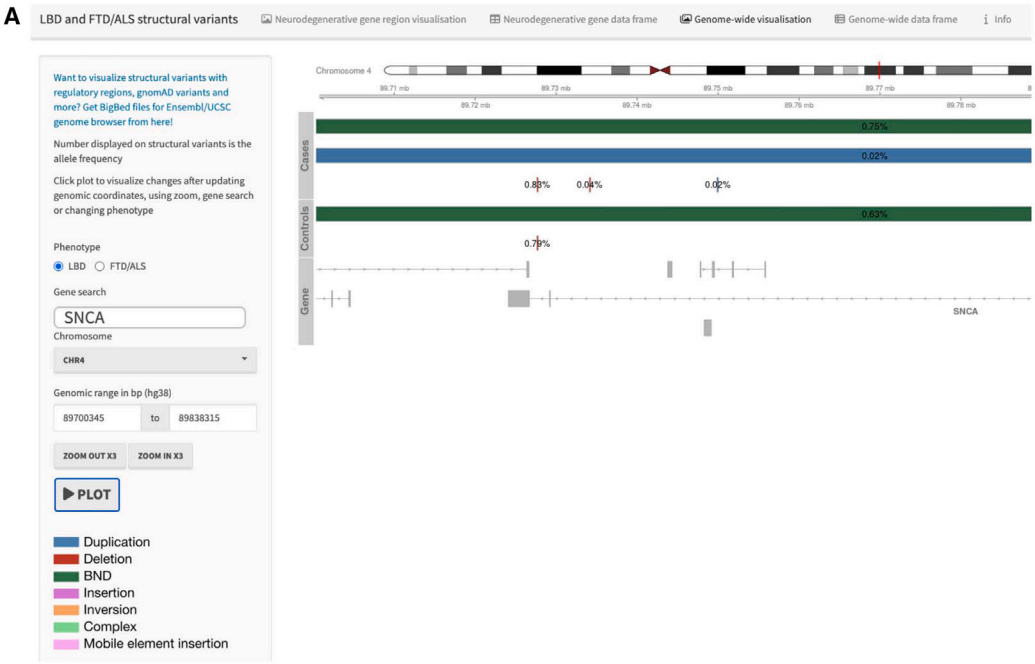
CONSORTIA

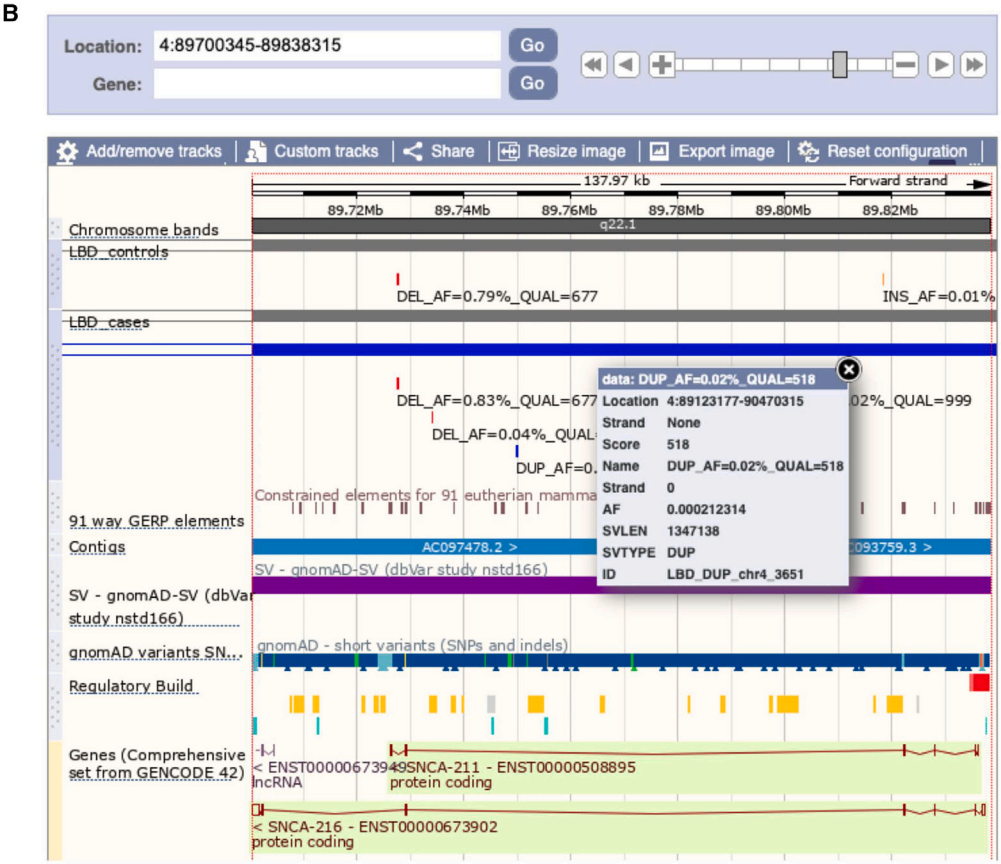
The members of The American Genome Center (TAGC) are Anthony R. Soltis, Coralie Viollet, Gauthaman Sukumar, Camille Alba, Nathaniel Lott, Elisa McGrath Martinez, Meila Tuck, Jatinder Singh, Dagmar Bacikova, Xijun Zhang, Daniel N. Hupalo, Adelani Adeleye, Matthew D. Wilkerson, Harvey B. Pollard, and Clifton L. Dalgard. The members of the International LBD Genomics Consortium are Sandra E. Black, Ziv Gan-Or, Julia Keith, Mario Masellis, Ekaterina Rogaeva, Alexis Brice, Suzanne Lesage, Georgia Xiromerisiou, Andrea Calvo, Antonio Canosa, Adriano Chio, Giancarlo Logroscino, Gabriele Mora, Reijko Krüger, Patrick May, Daniel Alcolea, Jordi Clarimon, Juan Fortea, Isabel Gonzalez-Aramburu, Jon Infante, Carmen Lage, Alberto Lleó, Pau Pastor, Pascual Sanchez-Juan, Francesca Brett, Dag Aarsland, Safa Al-Sarraj, Johannes Attems, Steve Gentleman, John A. Hardy, Angela K. Hodges, Seth Love, Ian G. McKeith, Christopher M. Morris, Huw R. Morris, Laura Palmer, Stuart Pickering-Brown, Mina Ryten, Alan J. Thomas, Claire Troakes, Marilyn S. Albert, Matthew J. Barrett, Thomas G. Beach, Lynn M. Bekris, David A. Bennett, Bradley F. Boeve, Clifton L. Dalgard, Ted M. Dawson, Dennis W. Dickson, Kelley Faber, Tanis Ferman, Luigi Ferrucci, Margaret E. Flanagan, Tatiana M. Foroud, Bernardino Ghetti, J. Raphael Gibbs, Alison Goate, David S. Goldstein, Neill R. Graff-Radford, Horacio Kaufmann, Walter A. Kukull, James B. Leverenz, Grisel Lopez, Qinwen Mao, Eliezer Masliah, Edwin Monuki, Kathy L. Newell, Jose-Alberto Palma, Matthew Perkins, Olga Pletnikova, Alan E. Renton, Susan M. Resnick, Liana S. Rosenthal, Owen A. Ross, Clemens R. Scherzer, Geidy E. Serrano, Vikram G. Shakkottai, Ellen Sidransky, Toshiko Tanaka, Nahid Tayebi, Eric Topol, Ali Torkamani, Juan C. Troncoso, Randy Woltjer, Zbigniew K. Wszolek, and Sonja W. Scholz. The members of the International ALS/FTD Consortium are Robert H. Baloh, Robert Bowser, Alexis Brice, James Broach, William Camu, Adriano Chiò, John Cooper-Knock, Carsten Drepper, Vivian E. Drory, Travis L. Dunckley, Eva Feldman, Pietro Fratta, Glenn Gerhard, Summer B. Gibson, Jonathan D. Glass, John A. Hardy, Matthew B. Harms, Terry D. Heiman-Patterson, Lilja Jansson, Janine Kirby, Justin Kwan, Hannu Laaksovirta, John E. Landers, Francesco Landi, Isabelle Le Ber, Serge Lumbroso, Daniel JL. MacGowan, Nicholas J. Maragakis, Kevin Mouzat, Liisa Myllykangas, Richard W. Orrell, Lyle W. Ostrow, Roger Pamphlett, Erik Pioro, Stefan M. Pulst, John M. Ravits, Wim Robberecht, Ekaterina Rogaeva, Jeffrey D. Rothstein, Michael Sendtner, Pamela J. Shaw, Katie C. Sidle, Zachary Simmons, Thor Stein, David J. Stone, Pentti J. Tienari, Bryan J. Traynor, Juan C. Troncoso, Miko Valori, Philip Van Damme, Vivianna M. Van Deerlin, Ludo Van Den Bosch, and Lorne Zinman. See Document S1 for consortium member affiliations.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Patient and control cohorts
- METHOD DETAILS
  - Short-read whole-genome sequencing
  - Structural variant mapping
  - Sample quality control and sample batching
  - Structural variant evidence collection
  - Structural variant discovery
  - Downstream filtering
  - Data curation
  - Validation of structural variants using nanopore long-read sequencing
  - Structural variant evaluation in the LBD replication cohort
  - Neurodegenerative disease gene analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Genome-wide association studies
  - Brain expression analysis





(legend on next page)

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.xgen.2023.100316.

ACKNOWLEDGMENTS

The authors are grateful to all patients, their family members, and caregivers, as well as all healthy participants, for making this study possible. This research was supported in part by the Intramural Research Program of the US National Institutes of Health (National Institute on Aging, National Institute of Neurological Disorders and Stroke; project numbers 1ZIAAG000935 [PI B.J.T.], 1ZIANS003154 [PI S.W.S.]). K.K. was supported by grants from The Päivikki and Sakari Sohlberg Foundation and the Finnish Parkinson Foundation. We are grateful to the Banner Sun Health Research Institute Brain and Body Donation Program of Sun City, Arizona, for the provision of human biological materials. The Brain and Body Donation Program is supported by the National Institute of Neurological Disorders and Stroke (U24 NS072026, National Brain and Tissue Resource for Parkinson's Disease and Related Disorders), the National Institute on Aging (P30 AG19610 and P30 AG072980, Arizona Alzheimer's Disease Center), the Arizona Department of Health Services (contract 211002, Arizona Alzheimer's Research Center), the Arizona Biomedical Research Commission (contracts 4001, 0011, 05-901, and 1001 to the Arizona Parkinson's Disease Consortium), and the Michael J. Fox Foundation for Parkinson's Research. The study used tissue samples and data from the Johns Hopkins Morris K. Udall Center of Excellence for Parkinson's Disease Research (NIH P50 NS38377). We thank the members of the Laboratory of Neurogenetics (NIH) for their collegial support and technical assistance. We thank members of the North American Brain Expression Consortium (NABEC) for providing samples derived from brain tissue. Brain tissue for the NABEC cohort was obtained from the Baltimore Longitudinal Study on Aging at the Johns Hopkins School of Medicine, the NICHD Brain and Tissue Bank for Developmental Disorders at the University of Maryland, the Banner Sun Health Research Institute Brain and Body Donation Program, and the University of Kentucky Alzheimer's Disease Center Brain Bank. We thank the UK Brain Expression Consortium (UKBEC) and the Northwestern University Brain Bank for providing DNA or tissue samples. The ROS/MAP study was supported by the National Institute on Aging (RF1 AG057473, U01 AG061356). This work utilized the computational resources of the NIH HPC Biowulf cluster USA (http://hpc.nih.gov). A complete list of acknowledgments is given in the supplemental information.

AUTHOR CONTRIBUTIONS

Conceptualization, S.W.S., B.J.T., K.B., and J.R.G.; methodology, S.W.S., B.J.T., C.L.D., R.C., K.B., M.R., A.R., and S.S.; software, J.D., J.R.G., R.L.C., H.T., M.B., X.Z., and K.K.; formal analysis, K.K., R.C., J.D., R.L.C., J.R.G., V.M., M.F., D.A.B., and P.L.D.; investigation, K.K., M.R., R.C., C.L.D., R.L.W., R.L.C., A.R., S.S., K.B., P.A.J., L.M., T.M.D., L.S.R., M.S.A., O.P., J.C.T., M.M., J.K., S.E.B., L.F., S.M.R., The American Genome Center; resources, International LBD Genomics Consortium, International ALS/FTD Consortium, E.T., A.T., T.M.F., J.E.L., L.M., S.D., C.M., L.M., S.D., A.C., G.E.S., T.G.B., P.T., T.F., B.G., N.R.G.-R., B.F.B., Z.K.W., D.W.D., A.C., D.A.B., P.L.D., O.A.R., C.L.D., J.R.G., B.J.T., and S.W.S.; data curation, K.K., R.C., J.D., and M.R.; writing – original draft, K.K.; writing – review & editing, all authors; visualization, K.K., J.D., and R.C.; supervision, O.A.R., C.L.D., J.R.G., S.W.S., and B.J.T.; funding acquisition, S.W.S. and B.J.T.

DECLARATION OF INTERESTS

S.W.S. serves on the Scientific Advisory Council of the Lewy Body Dementia Association and the Multiple System Atrophy Coalition. S.W.S. and B.J.T. receive research support from Cerevel Therapeutics. B.J.T. holds patents on the clinical testing and therapeutic implications of the *C9orf72* repeat expansion. H.R.M. is employed by the University College London. In the last 12 months, he reports paid consultancy from Roche and Amylyx; lecture fees/honoraria from BMJ, Kyowa Kirin, and Movement Disorders Society; and research grants from Parkinson's UK, Cure Parkinson's Trust, PSP Association, Medical Research Council, and the Michael J Fox Foundation. H.R.M. is a co-applicant on a patent application related to *C9orf72* "method for diagnosing a neurodegenerative disease" (PCT/GB2012/052140).

Received: January 23, 2023
Revised: March 21, 2023
Accepted: April 6, 2023
Published: May 4, 2023

REFERENCES

1. Ho, S.S., Urban, A.E., and Mills, R.E. (2020). Structural variation in the sequencing era. Nat. Rev. Genet. *21*, 171–189. https://doi.org/10.1038/s41576-019-0180-9.
2. Wheeler, D.A., Srinivasan, M., Egholm, M., Shen, Y., Chen, L., McGuire, A., He, W., Chen, Y.J., Makhijani, V., Roth, G.T., et al. (2008). The complete genome of an individual by massively parallel DNA sequencing. Nature *452*, 872–876. https://doi.org/10.1038/nature06884.
3. Blauwendraat, C., Pletnikova, O., Geiger, J.T., Murphy, N.A., Abramzon, Y., Rudow, G., Mamais, A., Sabir, M.S., Crain, B., Ahmed, S., et al. (2019). Genetic analysis of neurodegenerative diseases in a pathology cohort. Neurobiol. Aging *76*, 214.e1–214214.e9. https://doi.org/10.1016/j.neurobiolaging.2018.11.007.
4. Guyant-Marechal, I., Berger, E., Laquerrière, A., Rovelet-Lecrux, A., Vienet, G., Frebourg, T., Rumbach, L., Campion, D., and Hannequin, D. (2008). Intrafamilial diversity of phenotype associated with app duplication. Neurology *71*, 1925–1926. https://doi.org/10.1212/01.wnl.0000339400.64213.56.
5. Farrer, M., Kachergus, J., Forno, L., Lincoln, S., Wang, D.S., Hulihan, M., Maraganore, D., Gwinn-Hardy, K., Wszolek, Z., Dickson, D., and Langston, J.W. (2004). Comparison of kindreds with parkinsonism and alpha-synuclein genomic multiplications. Ann. Neurol. *55*, 174–179. https://doi.org/10.1002/ana.10846.
6. Obi, T., Nishioka, K., Ross, O.A., Terada, T., Yamazaki, K., Sugiura, A., Takanashi, M., Mizoguchi, K., Mori, H., Mizuno, Y., and Hattori, N. (2008). Clinicopathologic study of a SNCA gene duplication patient with Parkinson disease and dementia. Neurology *70*, 238–241. https://doi.org/10.1212/01.wnl.0000299387.59159.db.
7. Toffoli, M., Chen, X., Sedlazeck, F.J., Lee, C.Y., Mullin, S., Higgins, A., Koletsi, S., Garcia-Segura, M.E., Sammler, E., Scholz, S.W., et al. (2022). Comprehensive short and long read sequencing analysis for the Gaucher and Parkinson's disease-associated GBA gene. Commun. Biol. *5*, 670. https://doi.org/10.1038/s42003-022-03610-7.
8. DeJesus-Hernandez, M., Mackenzie, I.R., Boeve, B.F., Boxer, A.L., Baker, M., Rutherford, N.J., Nicholson, A.M., Finch, N.A., Flynn, H., Adamson, J., et al. (2011). Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. Neuron *72*, 245–256. https://doi.org/10.1016/j.neuron.2011.09.011.
9. Majounie, E., Renton, A.E., Mok, K., Dopper, E.G.P., Waite, A., Rollinson, S., Chiò, A., Restagno, G., Nicolaou, N., Simon-Sanchez, J., et al. (2012). Frequency of the C9orf72 hexanucleotide repeat expansion in patients with amyotrophic lateral sclerosis and frontotemporal dementia: a

Figure 6. Interactive catalog of structural variants in non-Alzheimer's dementias

(A) Screen shot of the online portal that lets the user filter structural variants and display genomic regions. For example, this image shows a large duplication encompassing the entire *SNCA* gene in an LBD case.

(B) Visualization of structural variants using BigBed files in the Ensembl genome browsers as custom tracks (https://www.ensembl.org).

cross-sectional study. Lancet Neurol. *11*, 323–330. https://doi.org/10.1016/S1474-4422(12)70043-1.

10. Renton, A.E., Majounie, E., Waite, A., Simón-Sánchez, J., Rollinson, S., Gibbs, J.R., Schymick, J.C., Laaksovirta, H., van Swieten, J.C., Myllykangas, L., et al. (2011). A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. Neuron *72*, 257–268. https://doi.org/10.1016/j.neuron.2011.09.010.
11. Abel, H.J., Larson, D.E., Regier, A.A., Chiang, C., Das, I., Kanchi, K.L., Layer, R.M., Neale, B.M., Salerno, W.J., Reeves, C., et al. (2020). Mapping and characterization of structural variation in 17,795 human genomes. Nature *583*, 83–89. https://doi.org/10.1038/s41586-020-2371-0.
12. Chia, R., Sabir, M.S., Bandres-Ciga, S., Saez-Atienzar, S., Reynolds, R.H., Gustavsson, E., Walton, R.L., Ahmed, S., Viollet, C., Ding, J., et al. (2021). Genome sequencing analysis identifies new loci associated with Lewy body dementia and provides insights into its genetic architecture. Nat. Genet. *53*, 294–303. https://doi.org/10.1038/s41588-021-00785-3.
13. Dewan, R., Chia, R., Ding, J., Hickman, R.A., Stein, T.D., Abramzon, Y., Ahmed, S., Sabir, M.S., Portley, M.K., Tucci, A., et al. (2021). Pathogenic huntingtin repeat expansions in patients with frontotemporal dementia and amyotrophic lateral sclerosis. Neuron *109*, 448–460.e4. https://doi.org/10.1016/j.neuron.2020.11.005.
14. Jakobsson, M., Scholz, S.W., Scheet, P., Gibbs, J.R., VanLiere, J.M., Fung, H.C., Szpiech, Z.A., Degnan, J.H., Wang, K., Guerreiro, R., et al. (2008). Genotype, haplotype and copy-number variation in worldwide human populations. Nature *451*, 998–1003. https://doi.org/10.1038/nature06742.
15. Li, Y., Roberts, N.D., Wala, J.A., Shapira, O., Schumacher, S.E., Kumar, K., Khurana, E., Waszak, S., Korbel, J.O., Haber, J.E., et al. (2020). Patterns of somatic structural variation in human cancer genomes. Nature *578*, 112–121. https://doi.org/10.1038/s41586-019-1913-9.
16. Sudmant, P.H., Rausch, T., Gardner, E.J., Handsaker, R.E., Abyzov, A., Huddleston, J., Zhang, Y., Ye, K., Jun, G., Fritz, M.H.Y., et al. (2015). An integrated map of structural variation in 2,504 human genomes. Nature *526*, 75–81. https://doi.org/10.1038/nature15394.
17. English, A.C., Menon, V.K., Gibbs, R.A., Metcalf, G.A., and Sedlazeck, F.J. (2022). Truvari: refined structural variant comparison preserves allelic diversity. Preprint at bioRxiv. https://doi.org/10.1101/2022.02.21.481353.
18. Byrska-Bishop, M., Evani, U.S., Zhao, X., Basile, A.O., Abel, H.J., Regier, A.A., Corvelo, A., Clarke, W.E., Musunuri, R., Nagulapalli, K., et al. (2022). High-coverage whole-genome sequencing of the expanded 1000 Genomes Project cohort including 602 trios. Cell *185*, 3426–3440.e19. https://doi.org/10.1016/j.cell.2022.08.004.
19. Vialle, R.A., de Paiva Lopes, K., Bennett, D.A., Crary, J.F., and Raj, T. (2022). Integrating whole-genome sequencing with multi-omic data reveals the impact of structural variants on gene regulation in the human brain. Nat. Neurosci. *25*, 504–514. https://doi.org/10.1038/s41593-022-01031-7.
20. Billingsley, K.J., Ding, J., Jerez, P.A., Illarianova, A., Grenn, F.P., Makarious, M.B., Moore, A., Vitale, D., Reed, X., Hernandez, D., et al. (2022). Genome-wide analysis of structural variants in Parkinson's disease using short-read sequencing data. BioRxiv. https://doi.org/10.1101/2022.08.22.504867.
21. Hutton, M., Lendon, C.L., Rizzu, P., Baker, M., Froelich, S., Houlden, H., Pickering-Brown, S., Chakraverty, S., Isaacs, A., Grover, A., et al. (1998). Association of missense and 5'-splice-site mutations in tau with the inherited dementia FTDP-17. Nature *393*, 702–705. https://doi.org/10.1038/31508.
22. Bellenguez, C., Küçükali, F., Jansen, I.E., Kleineidam, L., Moreno-Grau, S., Amin, N., Naj, A.C., Campos-Martin, R., Grenier-Boley, B., Andrade, V., et al. (2022). New insights into the genetic etiology of Alzheimer's disease and related dementias. Nat. Genet. *54*, 412–436. https://doi.org/10.1038/s41588-022-01024-z.
23. Nalls, M.A., Blauwendraat, C., Vallerga, C.L., Heilbron, K., Bandres-Ciga, S., Chang, D., Tan, M., Kia, D.A., Noyce, A.J., Xue, A., et al. (2019). 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. https://doi.org/10.1016/S1474-4422(19)30320-5.
24. GTEx Consortium (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science *369*, 1318–1330. https://doi.org/10.1126/science.aaz1776.
25. Eraslan, G., Drokhlyansky, E., Anand, S., Fiskin, E., Subramanian, A., Slyper, M., Wang, J., Van Wittenberghe, N., Rouhana, J.M., Waldman, J., et al. (2022). Single-nucleus cross-tissue molecular reference maps toward understanding disease gene function. Science *376*, eabl4290. https://doi.org/10.1126/science.abl4290.
26. Fujita, M., Gao, Z., Zeng, L., McCabe, C., White, C.C., Ng, B., Green, G.S., Rozenblatt-Rosen, O., Phillips, D., Amir-Zilberstein, L., et al. (2022). Cell-subtype specific effects of genetic variation in the aging and Alzheimer cortex. bioRxiv. https://doi.org/10.1101/2022.11.07.515446.
27. Grossman, M. (2010). Primary progressive aphasia: clinicopathological correlations. Nat. Rev. Neurol. *6*, 88–97. https://doi.org/10.1038/nrneurol.2009.216.
28. Rajput, A., Dickson, D.W., Robinson, C.A., Ross, O.A., Dächsel, J.C., Lincoln, S.J., Cobb, S.A., Rajput, M.L., and Farrer, M.J. (2006). Parkinsonism, Lrrk2 G2019S, and tau neuropathology. Neurology *67*, 1506–1508. https://doi.org/10.1212/01.wnl.0000240220.33950.0c.
29. Bannwarth, S., Ait-El-Mkadem, S., Chaussenot, A., Genin, E.C., Lacas-Gervais, S., Fragaki, K., Berg-Alonso, L., Kageyama, Y., Serre, V., Moore, D.G., et al. (2014). A mitochondrial origin for frontotemporal dementia and amyotrophic lateral sclerosis through CHCHD10 involvement. Brain *137*, 2329–2345. https://doi.org/10.1093/brain/awu138.
30. Dols-Icardo, O., Nebot, I., Gorostidi, A., Ortega-Cubero, S., Hernández, I., Rojas-García, R., García-Redondo, A., Povedano, M., Lladó, A., Álvarez, V., et al. (2015). Analysis of the CHCHD10 gene in patients with frontotemporal dementia and amyotrophic lateral sclerosis from Spain. Brain *138*, e400. https://doi.org/10.1093/brain/awv175.
31. Müller, K., Andersen, P.M., Hübers, A., Marroquin, N., Volk, A.E., Danzer, K.M., Meitinger, T., Ludolph, A.C., Strom, T.M., and Weishaupt, J.H. (2014). Two novel mutations in conserved codons indicate that CHCHD10 is a gene associated with motor neuron disease. Brain *137*, e309. https://doi.org/10.1093/brain/awu227.
32. Woo, J.A.A., Liu, T., Trotter, C., Fang, C.C., De Narvaez, E., LePochat, P., Maslar, D., Bukhari, A., Zhao, X., Deonarine, A., et al. (2017). Loss of function CHCHD10 mutations in cytoplasmic TDP-43 accumulation and synaptic integrity. Nat. Commun. *8*, 15558. https://doi.org/10.1038/ncomms15558.
33. Chow, C.Y., Landers, J.E., Bergren, S.K., Sapp, P.C., Grant, A.E., Jones, J.M., Everett, L., Lenk, G.M., McKenna-Yasek, D.M., Weisman, L.S., et al. (2009). Deleterious variants of FIG4, a phosphoinositide phosphatase, in patients with ALS. Am. J. Hum. Genet. *84*, 85–88. https://doi.org/10.1016/j.ajhg.2008.12.010.
34. Ishibashi, K., Suzuki, M., and Imai, M. (2000). Molecular cloning of a novel form (two-repeat) protein related to voltage-gated sodium and calcium channels. Biochem. Biophys. Res. Commun. *270*, 370–376. https://doi.org/10.1006/bbrc.2000.2435.
35. Cang, C., Bekele, B., and Ren, D. (2014). The voltage-gated sodium channel TPC1 confers endolysosomal excitability. Nat. Chem. Biol. *10*, 463–469. https://doi.org/10.1038/nchembio.1522.
36. She, J., Guo, J., Chen, Q., Zeng, W., Jiang, Y., and Bai, X.C. (2018). Structural insights into the voltage and phospholipid activation of the mammalian TPC1 channel. Nature *556*, 130–134. https://doi.org/10.1038/nature26139.
37. Wang, X., Zhang, X., Dong, X.P., Samie, M., Li, X., Cheng, X., Goschka, A., Shen, D., Zhou, Y., Harlow, J., et al. (2012). TPC proteins are phosphoinositide- activated sodium-selective ion channels in endosomes and lysosomes. Cell *151*, 372–383. https://doi.org/10.1016/j.cell.2012.08.036.

38. Foster, W.J., Taylor, H.B.C., Padamsey, Z., Jeans, A.F., Galione, A., and Emptage, N.J. (2018). Hippocampal mGluR1-dependent long-term potentiation requires NAADP-mediated acidic store Ca(2+) signaling. Sci. Signal. *11*, eaat9093. https://doi.org/10.1126/scisignal.aat9093.

39. Mallmann, R.T., and Klugbauer, N. (2020). Genetic inactivation of two-pore channel 1 impairs spatial learning and memory. Behav. Genet. *50*, 401–410. https://doi.org/10.1007/s10519-020-10011-1.

40. Neely Kayala, K.M., Dickinson, G.D., Minassian, A., Walls, K.C., Green, K.N., and Laferla, F.M. (2012). Presenilin-null cells have altered two-pore calcium channel expression and lysosomal calcium: implications for lysosomal function. Brain Res. *1489*, 8–16. https://doi.org/10.1016/j.brainres.2012.10.036.

41. Wacker, S.A., Alvarado, C., von Wichert, G., Knippschild, U., Wiedenmann, J., Clauss, K., Nienhaus, G.U., Hameister, H., Baumann, B., Borggrefe, T., et al. (2011). RITA, a novel modulator of Notch signalling, acts via nuclear export of RBP-J. EMBO J. *30*, 43–56. https://doi.org/10.1038/emboj.2010.289.

42. Fuchs, J., Nilsson, C., Kachergus, J., Munz, M., Larsson, E.M., Schüle, B., Langston, J.W., Middleton, F.A., Ross, O.A., Hulihan, M., et al. (2007). Phenotypic variation in a large Swedish pedigree due to SNCA duplication and triplication. Neurology *68*, 916–922. https://doi.org/10.1212/01.wnl.0000254458.17630.c5.

43. Pottier, C., Bieniek, K.F., Finch, N., van de Vorst, M., Baker, M., Perkersen, R., Brown, P., Ravenscroft, T., van Blitterswijk, M., Nicholson, A.M., et al. (2015). Whole-genome sequencing reveals important role for TBK1 and OPTN mutations in frontotemporal lobar degeneration without motor neuron disease. Acta Neuropathol. *130*, 77–92. https://doi.org/10.1007/s00401-015-1436-x.

44. Iida, A., Hosono, N., Sano, M., Kamei, T., Oshima, S., Tokuda, T., Nakajima, M., Kubo, M., Nakamura, Y., and Ikegawa, S. (2012). Novel deletion mutations of OPTN in amyotrophic lateral sclerosis in Japanese. Neurobiol. Aging *33*, 1843.e19–1843.e24. https://doi.org/10.1016/j.neurobiolaging.2011.12.037.

45. Osawa, T., Mizuno, Y., Fujita, Y., Takatama, M., Nakazato, Y., and Okamoto, K. (2011). Optineurin in neurodegenerative diseases. Neuropathology *31*, 569–574. https://doi.org/10.1111/j.1440-1789.2011.01199.x.

46. Collins, R.L., Brand, H., Karczewski, K.J., Zhao, X., Alföldi, J., Francioli, L.C., Khera, A.V., Lowther, C., Gauthier, L.D., Wang, H., et al. (2020). A structural variation reference for medical and population genetics. Nature *581*, 444–451. https://doi.org/10.1038/s41586-020-2287-8.

47. Chen, X., Schulz-Trieglaff, O., Shaw, R., Barnes, B., Schlesinger, F., Källberg, M., Cox, A.J., Kruglyak, S., and Saunders, C.T. (2016). Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. Bioinformatics *32*, 1220–1222. https://doi.org/10.1093/bioinformatics/btv710.

48. Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics *34*, 3094–3100. https://doi.org/10.1093/bioinformatics/bty191.

49. Li, H. (2021). New strategies to improve minimap2 alignment accuracy. Bioinformatics *37*, 4572–4574. https://doi.org/10.1093/bioinformatics/btab705.

50. Sedlazeck, F.J., Rescheneder, P., Smolka, M., Fang, H., Nattestad, M., von Haeseler, A., and Schatz, M.C. (2018). Accurate detection of complex structural variations using single-molecule sequencing. Nat. Methods *15*, 461–468. https://doi.org/10.1038/s41592-018-0001-7.

51. Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience *4*, 7. https://doi.org/10.1186/s13742-015-0047-8.

52. Abraham, G., Qiu, Y., and Inouye, M. (2017). FlashPCA2: principal component analysis of Biobank-scale genotype datasets. Bioinformatics *33*, 2776–2778. https://doi.org/10.1093/bioinformatics/btx299.

53. Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., and Li, H. (2021). Twelve years of SAMtools and BCFtools. GigaScience *10*, giab008. https://doi.org/10.1093/gigascience/giab008.

54. Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat. Biotechnol. *29*, 24–26. https://doi.org/10.1038/nbt.1754.

55. Willer, C.J., Li, Y., and Abecasis, G.R. (2010). METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics *26*, 2190–2191. https://doi.org/10.1093/bioinformatics/btq340.

56. Kleinert, P., and Kircher, M. (2022). A framework to score the effects of structural variants in health and disease. Genome Res. *32*, 766–777. https://doi.org/10.1101/gr.275995.121.

57. Boughton, A.P., Welch, R.P., Flickinger, M., VandeHaar, P., Taliun, D., Abecasis, G.R., and Boehnke, M. (2021). LocusZoom.js: interactive and embeddable visualization of genetic association study results. Bioinformatics *37*, 3017–3018.

58. Belyeu, J.R., Chowdhury, M., Brown, J., Pedersen, B.S., Cormier, M.J., Quinlan, A.R., and Layer, R.M. (2021). Samplot: a platform for structural variant visual validation and automated filtering. Genome Biol. *22*, 161.

59. Emre, M., Aarsland, D., Brown, R., Burn, D.J., Duyckaerts, C., Mizuno, Y., Broe, G.A., Cummings, J., Dickson, D.W., Gauthier, S., et al. (2007). Clinical diagnostic criteria for dementia associated with Parkinson's disease. Mov. Disord. *22*, 1689–1707, quiz 1837. https://doi.org/10.1002/mds.21507.

60. McKeith, I.G., Boeve, B.F., Dickson, D.W., Halliday, G., Taylor, J.P., Weintraub, D., Aarsland, D., Galvin, J., Attems, J., Ballard, C.G., et al. (2017). Diagnosis and management of dementia with Lewy bodies: fourth consensus report of the DLB Consortium. Neurology *89*, 88–100. https://doi.org/10.1212/WNL.0000000000004058.

61. Neary, D., Snowden, J.S., Gustafson, L., Passant, U., Stuss, D., Black, S., Freedman, M., Kertesz, A., Robert, P.H., Albert, M., et al. (1998). Frontotemporal lobar degeneration: a consensus on clinical diagnostic criteria. Neurology *51*, 1546–1554. https://doi.org/10.1212/wnl.51.6.1546.

62. Höglinger, G.U., Respondek, G., Stamelou, M., Kurz, C., Josephs, K.A., Lang, A.E., Mollenhauer, B., Müller, U., Nilsson, C., Whitwell, J.L., et al. (2017). Clinical diagnosis of progressive supranuclear palsy: the movement disorder society criteria. Mov. Disord. *32*, 853–864. https://doi.org/10.1002/mds.26987.

63. Brooks, B.R., Miller, R.G., Swash, M., and Munsat, T.L.; World Federation of Neurology Research Group on Motor Neuron Diseases (2000). El Escorial revisited: revised criteria for the diagnosis of amyotrophic lateral sclerosis. Amyotroph Lateral Scler. *1*, 293–299. https://doi.org/10.1080/146608200300079536.

64. Erikson, G.A., Bodian, D.L., Rueda, M., Molparia, B., Scott, E.R., Scott-Van Zeeland, A.A., Topol, S.E., Wineinger, N.E., Niederhuber, J.E., Topol, E.J., and Torkamani, A. (2016). Whole-genome sequencing of a healthy aging cohort. Cell *165*, 1002–1011. https://doi.org/10.1016/j.cell.2016.03.022.

65. Werling, D.M., Brand, H., An, J.Y., Stone, M.R., Zhu, L., Glessner, J.T., Collins, R.L., Dong, S., Layer, R.M., Markenscoff-Papadimitriou, E., et al. (2018). An analytical framework for whole-genome sequence association studies and its implications for autism spectrum disorder. Nat. Genet. *50*, 727–736. https://doi.org/10.1038/s41588-018-0107-y.

66. Gardner, E.J., Lam, V.K., Harris, D.N., Chuang, N.T., Scott, E.C., Pittard, W.S., Mills, R.E., and Devine, S.E.; 1000 Genomes Project Consortium (2017). the mobile element locator tool (MELT): population-scale mobile element discovery and biology. Genome Res. *27*, 1916–1929. https://doi.org/10.1101/gr.218032.116.

67. Kronenberg, Z.N., Osborne, E.J., Cone, K.R., Kennedy, B.J., Domyan, E.T., Shapiro, M.D., Elde, N.C., and Yandell, M. (2015). Wham: identifying structural variants of biological consequence. PLoS Comput. Biol. *11*, e1004572. https://doi.org/10.1371/journal.pcbi.1004572.

68. Klambauer, G., Schwarzbauer, K., Mayr, A., Clevert, D.A., Mitterecker, A., Bodenhofer, U., and Hochreiter, S. (2012). cn.MOPS: mixture of Poissons for discovering copy number variations in next-generation sequencing data with a low false discovery rate. Nucleic Acids Res. *40*, e69. https://doi.org/10.1093/nar/gks003.

69. Van der Auwera, G.A., and O'Connor, B.D. (2020). Genomics in the Cloud: Using Docker, GATK, and WDL in Terra (O'Reilly Media).

70. Smolka, M., Paulin, L.F., Grochowski, C.M., Mahmoud, M., Behera, S., Gandhi, M., Hong, K., Pehlivan, D., Scholz, S.W., Carvalho, C.M.B., et al. (2022). Comprehensive structural variant detection: from mosaic to population-level. Preprint at bioRxiv. https://doi.org/10.1101/2022.04.04.487055.

71. Karolchik, D., Hinrichs, A.S., Furey, T.S., Roskin, K.M., Sugnet, C.W., Haussler, D., and Kent, W.J. (2004). The UCSC Table Browser data retrieval tool. Nucleic Acids Res. *32*, D493–D496. https://doi.org/10.1093/nar/gkh103.

72. Hahne, F., and Ivanek, R. (2016). Visualizing genomic data using Gviz and bioconductor. Methods Mol. Biol. *1418*, 335–351. https://doi.org/10.1007/978-1-4939-3578-9_16.

73. Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., 3rd, Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive integration of single-cell data. Cell *177*, 1888–1902.e21. https://doi.org/10.1016/j.cell.2019.05.031.

74. Shabalin, A.A. (2012). Matrix eQTL: ultra fast eQTL analysis via large matrix operations. Bioinformatics *28*, 1353–1358. https://doi.org/10.1093/bioinformatics/bts163.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Human cerebral brain tissue and/or whole blood	A comprehensive list of study sites that provided biospecimens is listed in the supplemental information of this paper	N/A
Critical commercial assays		
Neuro Consortium v1.1 genotyping array	Illumina	https://www.illumina.com/science/consortia/human-consortia/neuro-consortium.html
Nanobind Tissue Big DNA Kit	Circulomics	https://www.circulomics.com/
NEBNext® Companion Module	New England Biolabs	E7180L
AMPure XP beads	Beckman Coulter	A63382
Ligation Sequencing Kit	Oxford Nanopore technologies	SQK-LSK109
Flow cell wash kit	Oxford Nanopore technologies	EXP-WSH004
Deposited data		
Human reference genome NCBI build 38, GRCh38	Genome Reference Consortium	https://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/human/
Gene region annotation data	Ensembl Biomart	https://www.ensembl.org/biomart/martview/
Gene region annotation data	UCSC TableBrowser	https://genome.ucsc.edu/cgi-bin/hgTables
GeneHancer v.5.1.0 data	GeneCards	https://www.genecards.org/
DEMENTIA-SEQ whole-genome sequencing data	dbGaP	https://www.ncbi.nlm.nih.gov/gap/Study Accession: phs001963.v2.p1
Gene expression data	GTeX V8	https://gtexportal.org/home/
Structural variant data and programming code	This study	https://github.com/ruthchia/Structural_variant_analysis-LBD-FTD https://doi.org/10.5281/zenodo.7796321
Structural variants as BigBed files for custom tracks in Ensembl/UCSC genome browser	This study	https://github.com/ruthchia/Structural_variant_analysis-LBD-FTD
Online resource for structural variant exploration	This study	https://ndru-ndrs-lng-nih.shinyapps.io/non_ad_dementias_sv_app/
Software and algorithms		
GATK-SV structural variant calling pipeline (versions gatk-sv-v1_2020.6.7 to gatk-sv-0.13-beta.2021.3.15)	Collins R.L. et al.[46]	https://github.com/broadinstitute/gatk-sv
Manta v.1.6.0	Chen X. et al.[47]	https://github.com/Illumina/manta
Guppy v.5.0.12	Oxford Nanopore technologies	https://github.com/nanoporetech/megalodon
MinKNOW v.22.08.6	Oxford Nanopore technologies	https://nanoporetech.com/
MiniMap2	Li H.[48,49]	https://github.com/lh3/minimap2
Sniffles2	Sedlazeck F.J. et al.[50]	https://github.com/fritzsedlazeck/Sniffles
PLINK v.1.9 and PLINK v.2.3	Chang C.C. et al.[51]	https://www.cog-genomics.org/plink/2.0/
flashPCA v.2	Abraham G. et al.[52]	https://github.com/gabraham/flashpca
Samtools v.1.16.1	Danecek P. et al.[53]	https://samtools.github.io/bcftools/bcftools.html

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Python 3.7. with packages pandas v. 1.3.5 and numpy v.1.21.6	NA	www.python.org
Integrative Genomics Viewer	Robinson J.T. et al.[54]	https://software.broadinstitute.org/software/igv/
R v.4.1.3 with packages tidyverse v.1.3.2, stats v.4.1.3, data.table v.1.14.6, gmodels v.2.18.1.1, readxl v.1.4.1, Gviz v.1.38.4, AnnotationHub v.3.2.2, ggplot2 v.3.4.0, Seurat v.4, Matrix-eQTL v.2.3, shiny v.1.7.3, shinywidgets v.0.7.5, datamods v.1.4.0, shinythemes v.1.2.0, shinysky v.0.1.3	R core team	https://www.r-project.org/
Truvari v.3.5.0	English A.C. et al.[17]	https://github.com/ACEnglish/truvari
METAL 2020-05-05	Willer C.J. et al.[55]	http://csg.sph.umich.edu/abecasis/metal/
CADD-SV v.1.1	Kleinert P. et al.[56]	https://cadd-sv.bihealth.org/
CellRanger v.6.0.0	10x Genomics	https://support.10xgenomics.com/single-cell-gene-expression/software/overview/welcome
LocusZoom v.0.12	Boughton A.P. et al.[57]	http://locuszoom.org/
Samplot June 18, 2021	Belyeu J.R. et al.[58]	https://github.com/ryanlayer/samplot

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Sonja W. Scholz (sonja.scholz@nih.gov).

Materials availability

This study did not generate new unique reagents. Short-read whole-genome sequencing data have been deposited in dbGaP (accession code phs001963.v2.p1).

Data and code availability

All original code and genome browser data tracks on structural variants have been deposited in GitHub: https://github.com/ruthchia/Structural_variant_analysis-LBD-FTD and Zenodo: https://doi.org/10.5281/zenodo.7796321. The GATK-SV pipeline used to map structural variants is available from GitHub: https://github.com/broadinstitute/gatk-sv. The structural variant resource app is available at https://ndru-ndrs-lng-nih.shinyapps.io/non_ad_dementias_sv_app/. Annotated structural variant calls in neurodegenerative disease genes are summarized in Table S6.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Patient and control cohorts

A total of 2,612 LBD cases, 2,601 FTD/ALS cases, and 4,132 unaffected controls were included in the study. These cohorts have been described elsewhere.[12,13] Briefly, LBD patients were diagnosed with pathologically definite (69.05% of the cohort) or clinically probable disease (30.95%) according to the McKeith and Emre consensus criteria.[59,60] These consensus criteria guide optimal methods to establish the clinical and pathological diagnosis of LBD, including diagnostic biomarkers. The FTD/ALS cohort included 1,377 patients diagnosed with FTD spectrum disorders, including the known subtypes of behavioral variant FTD, primary progressive aphasia, and progressive supranuclear palsy, and 1,065 patients diagnosed with ALS. Patients with FTD were diagnosed according to the Neary criteria[61] or the Movement Disorders Society criteria for progressive supranuclear palsy.[62] These criteria define core measures and several supportive and exclusion criteria for establishing a diagnosis of FTD. Patients with ALS were diagnosed according to the revised El Escorial criteria.[63] These criteria classify patients according to the level of diagnostic certainty and have been shown to be specific to the diagnosis of ALS.

The control participants included convenience control genomes obtained from the Wellderly cohort (n = 1,202 healthy, aged European-ancestry individuals recruited in the United States)[64] and the Accelerating Medicine Partnership – Parkinson's Disease Initiative (n = 1,016; www.amp-pd.org). The control cohorts were selected based on a lack of evidence of cognitive decline in their clinical history and the absence of neurological deficits on neurological examination. Pathologically confirmed control individuals (n = 605)

had no evidence of notable neurodegenerative disease on histopathological examination. The appropriate institutional review boards of the participating institutions approved the study (study identification numbers 03-AG-N329 and NCT02014246), and informed consent was obtained from all subjects or their surrogate decision-makers according to the Declaration of Helsinki.

The LBD replication cohort consisted of 667 independent cases with a clinical or neuropathological diagnosis of LBD and 274 neuropathologically unaffected control subjects. All samples were of European ancestry. The demographic characteristics of the study cohorts are summarized in Table S9.

METHOD DETAILS

Short-read whole-genome sequencing

The genomes were generated using PCR-free library preparations followed by 150 base-pair, paired-end sequencing on an Illumina HiSeq X Ten sequencing platform (version 2.5 chemistry, Illumina), as described elsewhere.[12,13] The average sequencing read-depth was 35x, and the mean coverage per genome was 36.3 (95% CI = 29.3–43.3).

Structural variant mapping

We used the Broad Institute's GATK-SV pipeline (in cohort mode) following default settings for structural variant discovery and downstream filtering.[46,65] This pipeline leverages five algorithms to call structural variants: (1) Manta,[47] (2) the Mobile Element Locator Tool (MELT),[66] (3) WHole-genome Alignment Metrics (WHAM),[67] (4) the Mixture Of PoissonS model for CNV detection (cn.MOPS),[68] and (5) GATK gCNV.[69] The LBD cases (n = 2,612) and the control subjects (n = 4,132) were called together, and the FTD/ALS cases (n = 2,601) and the same control subjects (n = 4,132) were independently called together using the GATK-SV pipeline. This approach was chosen to ensure accurate identification of rare structural variants in the case cohorts.

Sample quality control and sample batching

We excluded 239 samples from the LBD case-control cohort (representing 3.5% of the total cohort) and 259 samples (3.9%) from the FTD/ALS case-control cohort for failing at least one of the following quality control metrics: median sequencing coverage in 100 base-pair bins, dosage bias score δ, autosomal ploidy spread, Z score of outlier 1 megabase bins, chimera rate, pairwise alignment rate, read length, library contamination, ambiguous sex genotypes, and discordance between inferred and reported sex. We kept samples with non-canonical sex chromosome configurations in their batches and removed all non-autosomal structural variant calls.

Of note, the LBD and FTD cohorts used the same initial set of control samples. However, there was a difference of 23 control samples after the GATK-SV pipeline was applied due to the initial quality control steps in the pipeline. The thresholds for sample exclusion are not fixed. Instead, they are based on values derived from the analyzed samples, including metrics for ancestry and coverage. This filtering approach led to minor differences in which control samples were excluded when they were called separately with the LBD cases and FTD cases.

Following quality control, we subdivided the filtered samples into 16 batches. We ranked and binned the samples based on sex, median 100 base-pair binned coverage, and dosage bias scores so that samples of corresponding quality were batched together.

Structural variant evidence collection

The structural variant evidence of individual samples collected in this step included: raw structural variant/copy number variant calls and raw evidence-binned read counts, split reads, discordant read-pairs, and single nucleotide polymorphism B-allele frequency from five different structural variant algorithms: Manta (v1.4),[47] MELT (v2.2.0),[66] Wham (v1.7),[67] cn.MOPS (v1.20.1),[68] and GATK-gCNV.[69] One control sample in the LBD case-control cohort and two FTD/ALS cases had a high number of improper pairs in MELT and were excluded. We then aggregated the structural variant calls from the five algorithms and standardized them to meet the specifications required for the structural variant discovery pipeline.

Structural variant discovery

Structural variant discovery consisted of first clustering structural variants across all batches for each structural variant calling algorithm. Then, low-quality variants and samples with outlying variant numbers were removed. The filtered variants were combined for each variant calling algorithm across batches and genotypes. After assigning variant probability scores, all probable variants were re-genotyped and all variants from all variant calling algorithms across all batches were combined and complex variants were resolved.

Downstream filtering

Initial GATK-SV filtering steps included: (1) minGQ filtering with a 1% false discovery rate threshold, (2) FilterOutlierSamples, (3) BatchEffect, and (4) FilterCleanupQualRecalibration. In total, we filtered out 160 outliers in the LBD cohort and 76 in the FTD/ALS cohort. The final call set included 290,681 structural variants in the LBD cohort (n = 2,392 cases and n = 3,970 controls) and 294,232 variants in the FTD/ALS cohort (n = 2,450 cases and n = 3,948 controls).

Data curation

To create a high-quality subset of structural variants and samples for subsequent genetic analyses, we further filtered structural variants after the GATK-SV pipeline. We included structural variants with ''PASS'', ''MULTIALLELIC'' or ''UNRESOLVED'' filter. We set genotypes with a genotype quality (GQ) < 300 to missing and included variants with a genotyping rate >95% and Hardy-Weinberg equilibrium p value > $1x10^{-6}$ (mid-p adjustment). We excluded individuals with missing information and individuals who had failed previous single-nucleotide variant/indel calling.[12] This resulted in 2,355 LBD cases and 3,700 controls with 150,752 structural variants, and 2,307 FTD/ALS cases and 3,677 controls with 158,991 structural variants. These samples and variants underwent principal component analysis based on single-nucleotide polymorphism data.[12] For GWASes, we additionally filtered out structural variants with MAF <1% in cases, QUAL <500, variants within 100 kb of telomeres, variants within 2.5 Mb of centromeres, and variants within the variable, diversity, joining recombinant regions.

Validation of structural variants using nanopore long-read sequencing

We performed long-read Nanopore genome sequencing in 20 samples that had also been sequenced on the short-read Illumina HiSeq X10 platform. High molecular weight DNA extraction and long-read sequencing were performed on an Oxford Nanopore PromethION platform, as per the manufacturer's instructions. The tissue and library preparation steps followed a protocol described elsewhere: https://www.protocols.io/view/processing-human-frontal-cortex-brain-tissue-for-p-kxygxzmmov8j/v1. Briefly, the Nanobind Tissue Big DNA Kit (Circulomics) was used to extract DNA from cerebellar tissue on a KingFisher Apex instrument (Thermo Fisher Scientific, USA). The DNA was sheared to a target size of 30 kb on a Megaruptor 3 instrument (Diagenode, BE) and quantified using a Tapestation 4200 (Agilent Technologies). Next, the NEBNext Companion Module (New England Biolabs, USA) was used for DNA and end-repair, followed by AMPure XP bead purification (Beckman Coulter) and adaptor ligation (Ligation Sequencing Kit; Oxford Nanopore Technologies). PromethION R.9.4.1 flow cells were primed and loaded with 400 ng of the DNA libraries. One sample was loaded on each flow cell and flushed using the wash kit (Oxford Nanopore Technologies). Libraries were reloaded every 24 h to maximize the data output, with a total run time of 72 h. We performed base calling using Guppy (v.5.0.12; Oxford Nanopore Technologies) and MinKNOW (version 22.08.6; Oxford Nanopore Technologies) on a PromethION compute device. The reads were assembled to the reference genome (GRCh38) using MiniMap2[48,49] and structural variants were called with sniffles2.[70]

We used the *T. bench* algorithm to study the structural variant validation rate per sample and per structural variant type with different MAF thresholds (all, <1%, ≥1%).[17] A structural variant was validated if (1) the structural variant type matched, (2) breakpoint distance was within 200 base pairs, and (3) the reciprocal size overlap was >70% between the short-read and long-read structural variant calls.

We validated the top GWAS hits and the structural variants of interest in the neurodegenerative diseases genes using long-read sequencing (Nanopore), and genotyping array-based CNV calling by visualization of B-allele frequencies and log-R-ratio (Illumina, Neuro Consortium array v1.1). If DNA was exhausted in sequencing, we validated structural variant calls visually using Integrative Genomics Viewer.[54] Validation data are presented in the Supplementary Materials.

Structural variant evaluation in the LBD replication cohort

We used the Manta algorithm to detect the structural variants within the *TPCN1* locus on chromosome 12q24.13 in the 667 LBD cases and the 274 controls making up the replication cohort.[47] This analysis used default settings and focused on the region defined by the *TPCN1* deletion plus 1,000 flanking base-pairs on either side. Result files were merged with *bcftools*,[53] and missing genotypes were set to reference homozygotes. We used flashPCA (v.2)[52] based on single nucleotide polymorphism data to calculate the principal components. We then used the 'step()' function, as implemented in the R (v.4.0.3) 'stats' package (https://www.R-project.org/), to select the appropriate covariates from age, sex, and principal components 1–10 to include in the association model. Next, we excluded related samples and samples with missing covariates. This left 555 LBD cases and 274 controls available for the association analysis. Logistic regression on LBD status and deletion genotype, adjusted for age and principal components 1–5, was performed using PLINK2.[51] The age of 144 samples was expressed as a 10-year interval (e.g., 60–69 years), and the middle point of age (e.g., 65) was used for these subjects. We performed a meta-analysis of the discovery and replication cohorts using the inverse-variant weighted method implemented in METAL.[55]

Neurodegenerative disease gene analysis

We extracted structural variants that were located within 50 neurodegenerative disease genes (+/− 1Mb, Table S5). To annotate structural variants that overlapped these genes, we retrieved the Ensembl canonical transcript gene regions, exon ranges, and gene strand information using the UCSC Table Browser (GRCh38, GENCODE V38).[71] We defined promoter regions as the 1kb flanking region upstream of the transcription start site on the transcribed strand. We used PLINK2 (v.2.3)[51] to calculate structural variant allele frequencies in the cases and controls. To annotate structural variants, we used CADD-SV (v.1.1) with default settings to calculate scores for duplications, deletions, and insertions (including mobile element insertions).[56]

To help explore structural variants within neurodegenerative disease gene regions, we created an interactive app using the R packages *shiny*, *shinywidgets*, and *datamods*. Prerendered images for the interactive app were created using the *Gviz* package.[72]

QUANTIFICATION AND STATISTICAL ANALYSIS

Genome-wide association studies

We used flashPCA (v.2)[52] to calculate the principal components for the LBD case-control and the FTD/ALS case-control cohorts. We then used the 'step()' function, as implemented in the R (v.4.0.3) 'stats' package (https://www.R-project.org/), to select the appropriate covariates for each association analysis. In the LBD GWAS, we adjusted for age, sex, and principal components 1, 3, 4, 7, and 8. In the FTD/ALS GWAS, we adjusted for age, sex, and principal components 1, 2, 3, and 7. We used logistic Firth hybrid regression implemented in PLINK2[51] to perform the genome-wide association analyses. The Bonferroni threshold of significance was 1.02×10^{-5} (=0.05/4,889 structural variants with a MAF $\geq$1%) for the LBD GWAS and 1.06×10^{-5} (= 0.05/4,699) for the FTD/ALS GWAS.

Brain expression analysis

We examined the expression of *TPCN1* and neighboring genes in a single-nucleus RNA-seq dataset from the Religious Orders Study/Memory and Aging Project (ROS/MAP) cohort.[26] These data were prepared from 424 dorsolateral prefrontal cortexes of individuals of advanced age using the 10x Genomics Single Cell 3′ kit, as described elsewhere.[26] Sequencing reads were processed, and the unique molecule identifier (UMI) count matrix was generated using Cell Ranger software (v.6.0.0, 10x Genomics). The classification of the cell types was performed by clustering cells by gene expression using the R package Seurat (v.4).[73] The ''pseudobulk'' gene expression matrix was constructed by aggregating UMI counts of the same cell type of the same donor and normalizing them to the log2 counts per million reads mapped (CPM) values. Genotyping was performed by whole-genome sequencing followed by GATK processing. Mapping of *cis*-eQTL was performed using Matrix-eQTL (v.2.3) for single nucleotide polymorphisms within 1Mb of the transcription start sites.[74]