

RESEARCH

Open Access



Structural variants linked to Alzheimer's disease and other common age-related clinical and neuropathologic traits

Ricardo A. Vialle^{1*}, Katia de Paiva Lopes¹, Yan Li¹, Bernard Ng¹, Julie A. Schneider¹, Aron S. Buchman¹, Yanling Wang¹, Jose M. Farfel¹, Lisa L. Barnes¹, Aliza P. Wingo^{2,3}, Thomas S. Wingo⁴, Nicholas T. Seyfried⁵, Philip L. De Jager⁶, Chris Gaiteri^{1,7}, Shinya Tasaki¹ and David A. Bennett¹

Abstract

Background Alzheimer's disease (AD) is a complex neurodegenerative disorder with substantial genetic influence. While genome-wide association studies (GWAS) have identified numerous risk loci for late-onset AD (LOAD), the functional mechanisms underlying most of these associations remain unresolved. Large genomic rearrangements, known as structural variants (SVs), represent a promising avenue for elucidating such mechanisms within some of these loci.

Methods By leveraging data from two ongoing cohort studies of aging and dementia, the Religious Orders Study and Rush Memory and Aging Project (ROS/MAP), we performed genome-wide association analysis testing 20,205 common SVs from 1088 participants with whole genome sequencing (WGS) data. A range of Alzheimer's disease and other common age-related clinical and neuropathologic traits were examined.

Results First, we mapped SVs across 81 AD risk loci and discovered 22 SVs in linkage disequilibrium (LD) with GWAS lead variants and directly associated with the phenotypes tested. The strongest association was a deletion of an *Alu* element in the 3'UTR of the *TMEM106B* gene, in high LD with the respective AD GWAS locus and associated with multiple AD and AD-related disorders (ADRD) phenotypes, including tangles density, TDP-43, and cognitive resilience. The deletion of this element was also linked to lower TMEM106B protein abundance. We also found a 22-kb deletion associated with depression in ROS/MAP and bearing similar association patterns as GWAS SNPs at the *IQCK* locus. In addition, we leveraged our catalog of SV-GWAS to replicate and characterize independent findings in SV-based GWAS for AD and five other neurodegenerative diseases. Among these findings, we highlight the replication of genome-wide significant SVs for progressive supranuclear palsy (PSP), including markers for the 17q21.31 *MAPT* locus inversion and a 1483-bp deletion at the *CYP2A13* locus, along with other suggestive associations, such as a 994-bp duplication in the *LMNTD1* locus, suggestively linked to AD and a 3958-bp deletion at the *DOCK5* locus linked to Lewy body disease (LBD) ($P = 3.36 \times 10^{-4}$).

Conclusions While still limited in sample size, this study highlights the utility of including analysis of SVs for elucidating mechanisms underlying GWAS loci and provides a valuable resource for the characterization of the effects of SVs in neurodegenerative disease pathogenesis.

Keywords Alzheimer's disease, Structural variants, GWAS, WGS, Cognitive decline

*Correspondence:

Ricardo A. Vialle

ricardo_a_vialle@rush.edu

Full list of author information is available at the end of the article



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Background

Alzheimer's disease (AD) is a complex neurodegenerative disorder heavily influenced by genetics. While recent genome-wide association studies (GWAS) have been successful in identifying numerous single-nucleotide polymorphisms (SNPs) in risk loci for late-onset AD (LOAD) [1–6], the functional mechanisms underlying these associations often remain unresolved. Large genomic rearrangements, known as structural variants (SVs), represent an understudied class of genetic variation that is often not discoverable using genotyping assays but is recently garnering attention with the popularization of whole-genome sequencing data, which allows the detection of such variants at a large scale with considerable confidence [7, 8]. SVs can directly disrupt gene function or influence regulatory mechanisms, potentially accounting for causal relationships at these loci [9, 10].

Although a few SVs have already been linked to AD, with the most well-known being rare duplications of the *APP* gene causally linked to the early onset form of the disease [11–15], our knowledge of the impact of common SVs in LOAD is mostly limited to a few copy number variants (CNVs), usually detected using SNP arrays or PCR assays and identified in small sample-size studies that have weak replication [16]. A more recent effort led by the Alzheimer's Disease Sequencing Project (ADSP) tried to link common frequency SVs to AD but did not find genome-wide significant results, even with a sample size of over 12,000 individuals [17]. For other neurodegenerative diseases, the role of SVs is usually clearer. For example, in frontotemporal dementia (FTD), repeat expansions in *C9orf72* and the 17q21.31 *MAPT* locus inversion are linked to disease risk [18–20]. The same inversion is also a major genetic risk factor for progressive supranuclear palsy (PSP) [18, 21], and it is associated with Parkinson's disease (PD) [22, 23]. Additionally, a recent study found a common deletion in the gene *TPCNI* associated with Lewy body dementia (LBD) in a locus also associated with AD GWAS [1, 20]. These findings highlight the fact that mapping SVs across multiple diseases and related clinical and neuropathological traits can be beneficial to understanding their role in complex diseases.

Previously, we reported that SVs have an impact on many molecular phenotypes in the human brain, including over 300 SVs that were also in LD with GWAS traits [24]. Building on this foundation, here we extend our findings in many ways by testing almost 20,000 common SVs identified through whole-genome sequencing (WGS) in 1088 individuals for association with a comprehensive set of common age-related clinical and neuropathologic traits. We aimed to assess the role of SVs systematically, focusing on their potential contribution in relation to

more powerful GWAS studies. We leveraged data from two ongoing cohort studies of aging and dementia: the Religious Orders Study and the Rush Memory and Aging Project (ROS/MAP). These cohorts benefit from having a comprehensive set of harmonized deeply phenotyped measures, including clinical diagnoses of Alzheimer's dementia and mild cognitive impairment (MCI); multiple measurements of cognition (e.g., cognitive decline and resilience); diagnosis of major depressive disorder (MDD) and depressive symptoms; indices of motor function; neuropathological evaluations of β -amyloid load, neurofibrillary tangles density, Lewy bodies, and TDP-43; and indices of cerebrovascular diseases (CVD) such as cerebral amyloid angiopathy (CAA), gross and microinfarcts, atherosclerosis, and arteriolosclerosis.

This study offers a valuable resource for elucidating the role of SVs in AD and ADRD, providing insights into the functional mechanisms underlying GWAS signals and potential novel targets for further investigation. By integrating WGS-derived SV data with richly phenotyped cohorts, we aim to bridge the gap between genetic associations and their biological impact, highlighting the utility of SVs in unraveling complex disease mechanisms.

Methods

Study participants

The study uses data from 529 participants from the Religious Orders Study (ROS) [25, 26] and 559 participants from the Rush Memory and Aging Project (MAP) [25, 27], two longitudinal cohort studies. ROS and MAP enroll participants free of known dementia. Participants agree to annual clinical evaluations and to donate their brains at death. ROS was initiated in 1994 and enrolls older Catholic priests, nuns, and brothers from nearly 40 groups located in 12 US states [25, 28]. By the time the samples were sent for sequencing (end of 2017), 1437 individuals had completed their baseline assessment. The MAP was established in 1997 and recruits older men and women from retirement communities and individual householders in the greater Chicago area [25, 27–29]. As of the end of 2017, 1967 participants had completed their baseline evaluation. ROS and MAP are studies conducted by the same team of investigators sharing a common core of measures and procedures, allowing direct comparison between variables and efficient data merging for combined analyses. The follow-up rate for surviving participants surpasses 90%. Both studies were approved by a Rush University Medical Center Institutional Review Board. Each participant provided written informed consent at enrollment and signed the Uniform Anatomical Gift Act. The whole genome sequencing data analyzed in this manuscript was limited to subjects with autopsy data [30]. Genotyped structural variant calls were available

on 1088 non-Latino white subjects from the ROS/MAP cohort studies [24]. Detailed characteristics of each cohort are presented in Additional file 1: Table S1.

Alzheimer's dementia and cognitive function

Standardized cognitive and clinical assessments are conducted annually by examiners unaware of previous data. The clinical diagnosis for dementia follows the directives provided by the joint working group of the National Institute of Neurological and Communicative Disorders and Stroke and the AD and Related Disorders Association, as described [31, 32]. Mild cognitive impairment (MCI) was defined as individuals assessed by the neuropsychologist as cognitively impaired but not diagnosed as having dementia by the examining physician, as previously outlined [33]. Persons without dementia or MCI were designated as having no cognitive impairment (NCI) as described [33]. A final consensus cognitive diagnosis is determined by a neurologist with proficiency in dementia after reviewing select clinical information after death without knowledge of any postmortem data [34]. For the current association analysis, two binary statuses were used: Alzheimer's dementia vs no dementia (MCI + NCI) and with cognitive impairment (AD + MCI) vs no cognitive impairment (NCI).

Quantitative measurements of cognitive function were measured yearly. Cognitive evaluations comprise 19 cognitive performance tests that are common to both studies. A Mini-Mental State Examination is employed for descriptive reasons, whereas the Complex Ideational Material from the Boston Diagnostic Aphasia Examination is solely utilized for diagnostic classification. The remaining 17 tests are merged into a comprehensive metric of global cognition. The scores for each test were converted to a composite score [35]. For the association analysis, measurements were considered proximal to death.

Estimated slopes for global cognition were also included as a measure of cognitive decline. The random slope of global cognition is calculated to the individualized projected pace of change in the global cognition variable across time. This projection is generated through a linear mixed-effects model, with global cognition serving as the longitudinal outcome. The model adjusts for age at baseline, sex, and years of education [36, 37]. Additionally, as a measure of cognitive resilience, another random slope of global cognition is calculated, controlling for demographics and neuropathologies. The linear mixed-effects model generates the projection using global cognition as the outcome while adjusting for age at baseline, sex, and years of education as demographics, and global AD pathology burden, β -amyloid, PHF tau tangles, gross chronic cerebral infarctions, chronic microinfarctions,

Lewy body disease, TDP-43, hippocampal sclerosis, cerebral amyloid angiopathy, cerebral atherosclerosis, and arteriolosclerosis as pathology factors [38].

Parkinsonism, frailty, and motor function

A global Parkinsonian summary score was also tested. The global Parkinsonian summary score is a composite measure of Parkinsonian signs. It is calculated as the average of four separate domains based on a 26-item modified version of the motor portion of the United Parkinson's Disease Rating Scale (mUPDRS). These domains include bradykinesia, gait, rigidity, and tremor, and they are administered by a trained nurse clinician. These measures are highly reliable and reproducible in both men and women across various cohorts and have been modified to be more applicable to individuals without Parkinson's disease and easier for non-physicians to administer and score [39].

In addition, a total of two motor-related indices were evaluated. A measure of frailty and a composite measure of global motor function. Each of the indices is described below. Frailty is defined as multiple system weaknesses. A continuous composite measure of frailty is based on four components: grip strength, timed walking, body composition (BMI), and fatigue. The raw scores for each component are converted into z-scores using the mean and standard deviation values from all participants at baseline [40, 41]. The global motor function combines multiple motor tests, including the Purdue Pegboard Test, finger-tapping test, the time and number of steps to cover a distance of 8 feet, the time and number of steps for 360-degree turn, leg and toe stand, grip strength, and pinch strength. Each test's performance score is converted to a score based on the mean score of all participants at baseline, and then the scores are averaged together to create the composite measure. This measure provides a comprehensive assessment of motor and gait function in individuals [42].

Depression and depressive symptoms

Two measures of depression were evaluated: a binary status of clinical diagnosis of major depressive disorder (MDD) and a quantitative score of depressive symptoms. The clinical diagnosis of major depressive disorder was made by an examining physician at each evaluation. Diagnosis was based on criteria of the Diagnostic and Statistical Manual of Mental Disorders (DSM-III-R), a clinical interview with the participant, and a review of responses to questions adapted from the Diagnostic Interview Schedule [43]. A binary variable was tested classifying probably or highly probable vs. possible or not present MDD into the presence or absence of MDD. Depressive symptoms were assessed using a modified 10-item

version of the Center for Epidemiologic Studies Depression Scale (CES-D) [44–46]. Participants were asked whether they experienced each of the ten symptoms frequently in the past week. An overall score was obtained by aggregating the number of symptoms reported.

Neuropathological evaluations

Systematic assessment of various neurodegenerative and cerebrovascular conditions, including pathological diagnosis of Alzheimer's disease, Lewy bodies, LATE, hippocampal sclerosis, chronic macroscopic infarcts and microinfarcts, cerebral amyloid angiopathy, atherosclerosis, and arteriolosclerosis were performed as previously reported [38, 47]. The assessments were conducted by examiners who were blinded to all clinical data. A summary of each variable used in the current study is described below:

NIA-Reagan diagnosis of AD: The NIA-Reagan diagnosis of Alzheimer's disease was measured based on a set of consensus recommendations for diagnosing the disease after death, taking into account the presence of both neurofibrillary tangles (Braak) and neuritic plaques (CERAD). This criteria was modified because the neuropathological evaluation is carried out without knowledge of the patient's clinical information, including a dementia diagnosis. Thus, the level of Alzheimer's disease pathology is determined by a neuropathologist [28]. In our association analysis, we utilized a dichotomized variable where individuals with an intermediate or high likelihood fulfill the criteria for a pathological diagnosis of Alzheimer's disease.

Burden of neuritic and diffuse plaques and neurofibrillary tangles: Neuritic and diffuse plaques were identified through microscopic examination of silver-stained slides from five specific brain regions. The index count in each region is scaled by dividing it by its corresponding standard deviation. The scaled regional measures are then averaged to obtain a summary measure for the plaque burden. The five regions examined are the mid-frontal cortex, mid-temporal cortex, inferior parietal cortex, entorhinal cortex, and mid-hippocampus CA1. In addition, a global score of AD pathology was also tested as a quantitative measure derived from the counts of three silver-stained measures of AD pathologies: neuritic plaques, diffuse plaques, and neurofibrillary tangles. Measures were made in the same five regions, and each regional count was scaled by dividing by the corresponding standard deviation. The average of the three measured was used [48, 49].

β -amyloid load: β -amyloid protein is identified by molecularly specific immunohistochemistry and quantified by image analysis in 8 brain regions. The percent area of the cortex occupied by β -amyloid is calculated, and a mean score is determined from 4 or more regions [50].

PHFtau tangles density: Neuronal tangles are identified with phosphorylated Tau protein antibodies (AT8, Innogenetics, San Ramon, CA, USA; 1:1000) [50] and quantified in 8 brain regions using systematic sampling by stereology to determine cortical density. A mean tangles score is then calculated from 4 or more regions [50].

TDP-43: TDP-43 cytoplasmic inclusions in neurons and glia are assessed for each of the eight brain regions and scored based on four stages of TDP-43 distribution (ranging from none to involvement of all eight regions). A dichotomized presence or absence is used in these analyses [51].

Lewy body disease: A pathologic diagnosis of Lewy body (LB) disease is determined based on four stages of distribution of α -synuclein in the brain. Brain tissue samples from multiple regions are evaluated with α -synuclein immunostaining [52]. The McKeith criteria were modified to assess the presence of LB in different categories: not present, nigral-predominant, limbic-type, and neocortical-type. A dichotomized version of this variable is used, referring to Lewy bodies present or absent [52].

Arteriolosclerosis: Arteriolosclerosis refers to histological changes observed in small brain vessels during aging, including intimal deterioration, smooth muscle degeneration, and fibrohyalinotic thickening that narrows the vascular lumen. The severity of arteriolosclerosis is evaluated as none, mild, moderate, or severe [53].

Cerebral atherosclerosis: The severity of large vessel cerebral atherosclerosis is visually assessed by examining several arteries and their proximal branches in the circle of Willis. The rating was based on the extent of involvement, including the number of arteries affected and the degree of occlusion. A semiquantitative scale is used: none or possible, mild, moderate, and severe [54].

Cerebral amyloid angiopathy: A semiquantitative summary of cerebral amyloid angiopathy (CAA) pathology in 4 neocortical regions is calculated using paraffin-embedded sections that were immunostained for β -amyloid using one of three monoclonal anti-human antibodies: 4G8 (1:9000; Covance Labs, Madison, WI), 6F/3D (1:50; Dako North America Inc., Carpinteria, CA), and 10D5 (1:600; Elan Pharmaceuticals, San Francisco, CA) [55]. Meningeal

and parenchymal vessels are assessed for β -amyloid deposition and scored from 0 to 4 based on the extent of circumferential deposition for each region. The CAA score for each region is the maximum of the meningeal and parenchymal CAA scores, which are then averaged across regions to summarize as a continuous measure of CAA pathology [55].

Cerebral infarctions—gross-chronic: Neuropathologic evaluations are performed to determine the presence of one or more gross chronic cerebral infarctions. The evaluations are blinded to clinical data and reviewed by a board-certified neuropathologist. The examination documents the age (acute/subacute/chronic), size, and location (side and region) of infarcts visible to the naked eye on fixed slabs. All visible and suspected macroscopic infarcts are dissected for histologic confirmation. A value is one or more gross chronic infarctions vs none [56, 57].

Cerebral infarctions—micro-chronic: Neuropathologic evaluations are performed to determine the presence of one or more chronic microinfarcts, which are chronic microscopic infarctions. The evaluations are blinded to clinical data and reviewed by a board-certified neuropathologist. At least nine regions in one hemisphere are examined for microinfarcts on 6 μ m paraffin-embedded sections stained with hematoxylin/eosin. The examination includes six cortical regions (mid-frontal, middle temporal, entorhinal, hippocampal, inferior parietal, and anterior cingulate cortices), two subcortical regions (anterior basal ganglia, thalamus), and midbrain. Age (acute/subacute/chronic) and location (side and region) of microinfarcts are recorded. A value of 0 indicates no chronic microinfarcts, while a value of 1 indicates the presence of one or more chronic microinfarcts [56].

WGS data and variant calling

Whole-genome sequencing (WGS) data were previously generated from DNA samples from blood or cortex tissues [30]. Briefly, libraries were sequenced on an Illumina HiSeq X sequencer using 2 \times 150 bp cycles. Single nucleotide variant and small indel discovery and genotyping were performed utilizing an NYGC automated pipeline, which included alignment to the GRCh37 human reference using the Burrows-Wheeler Aligner and processing using the GATK best-practices workflow. The workflow included marking duplicate reads, local realignment around indels, and using Genome Analysis Toolkit (GATK) base quality score recalibration.

Structural variant discovery and genotyping were also previously described. A combination of seven different

software tools, including DELLY [58], LUMPY [59], Manta [60], BreakDancer [61], CNVnator [62], BreakSeq [63], and MELT [64], was applied to identify SVs in each sample. The variants were then merged at the individual level using SURVIVOR [65] using the following criteria for each SV type: DEL, all Manta calls plus any SVs with support of other 2 different tools; INS all Manta and BreakSeq calls; DUP/INV/TRA any SVs with support of 2 different tools. Merging of SVs was performed requiring 1000 bp maximum distance between breakpoints. Genotyping was performed for each SV type separately using smooove [66]. Mobile elements were genotyped separately via specific designed functions in the MELT pipeline. For harmonization, we used a 70% reciprocal overlap criteria (considering their breakpoint positions and lengths) to convert insertions into MEIs. ROS and MAP were jointly genotyped, resulting in a final set of 72,348 SVs mapped in 1,106 individuals after quality control. Details about the specific SV pipeline steps are described in the original publication [24]. A further level of QC was used to remove any related individuals with kinship scores higher than 0.0442 using KING [67], resulting in a final data set of 1,088 individuals.

Linkage disequilibrium with GWAS variants

Linkage disequilibrium between SNPs and SVs was also previously generated [24]. A joint call set with 8,566,510 SNPs and 72,348 SVs was used to calculate LD in terms of R^2 for all SVs using PLINK and considering a window of 5 Mb.

Single variant association analysis of SVs and AD/ADRD traits

For the association analysis, the initial 72,348 SVs mapped were filtered with a Hardy–Weinberg equilibrium (HWE) P -value lower than 10^{-6} and minor allele counts (MAC) greater than 10 in each cohort. Resulting in 20,022 SVs in ROS and 20,078 SVs in MAP. We performed genome-wide association analysis testing for each common SV with 16 quantitative and eight binary AD/ADRD phenotypes. The variables measuring β -amyloid tangles density, neurofibrillary tangles burden, and neuritic and diffuse plaque burden were squared root transformed before the association analysis.

Analysis was performed using SAIGEgds (Scalable and Accurate Implementation of GEneralized mixed model) [68, 69]. The method uses the saddlepoint approximation to calibrate the distribution of score test statistics and state-of-the-art optimization strategies to reduce computational costs. SAIGE can analyze large-scale data while controlling for unbalanced case–control ratios and sample relatedness, making it applicable to GWAS for thousands of phenotypes by large biobanks. The

method was tested on UK Biobank data and demonstrated its efficiency in analyzing extensive sample data. The SAIGE method involves two main steps: (1) fitting the null logistic (for binary traits) or linear (for quantitative traits) mixed model to estimate model parameters using the average information restricted maximum likelihood algorithm, and (2) testing for associations between each genetic variant and phenotype using the saddlepoint approximation method on score test statistics. Several optimization strategies have been applied to make fitting the null logistic mixed model practical for large data sets, such as using the raw genotypes as input and the preconditioned conjugate gradient method to solve linear systems iteratively. These optimizations make SAIGE computationally efficient and applicable to GWAS for thousands of phenotypes by large biobanks.

Here, we performed separate genome-wide scans using 529 participants from ROS and 559 participants from MAP. All tests were controlled by age at death, sex, years of education, five genetic principal components, and the genetic correlation matrix (modeled as a random effect). To combine consistent genetic effects across both studies, a meta-analysis was conducted using METASOFT v.2.0.1 [70]. Effect sizes and standard errors of each SV-trait pair were used as input. We carried out a random-effects meta-analysis using the RE2 model, optimized to detect associations under heterogeneity.

Proteomics data

In our analysis, we utilized ROS/MAP proteomics data to link the impact of SVs with protein abundances. The methods for data generation were previously published in detail [71–74]. Briefly, frozen DLPFC tissue samples underwent homogenization, followed by the quantification of protein concentrations. Isobaric TMT peptide labels were subsequently introduced and fractionated by high pH. These fractions were then subjected to liquid chromatography-mass spectrometry, and the resulting spectra were cross-referenced with the UniProt database. After quality control to eliminate technical confounders, 10,030 proteins from 971 individuals were available for downstream analysis [71].

Comparison with SV-GWAS studies

We obtained full summary stats for SV-GWAS performed for Alzheimer's disease (AD) [17], Parkinson's disease (PD) [23], progressive supranuclear palsy (PSP) [21], Lewy body dementia (LBD) [20], and frontotemporal dementia/amyotrophic lateral sclerosis (FTD/ALS) [20]. All these studies had their SV calls mapped to the GRCh38 reference genome. Therefore, we lifted genomic coordinates from GRCh38 to GRCh37 using the “liftOver” function from the R package “rtracklayer” [75] and

the corresponding chain file (hg38ToHg19). To match SVs across studies, we considered SVs with a reciprocal overlap of 80% using “bedtools intersect” [76] and removed overlapping SVs with discordant SV types (i.e., DELs vs INSSs). Specifically for GWAS for AD and PSP, insertion lengths were not reported in the respective final reported results. For these cases (INS only), we matched SVs by their breakpoint position using “bedtools closest” while considering a threshold of 100 bp. The resulting matched SVs were further compared in terms of MAF, showing high concordance (Additional file 1: Fig. S1). For replication, we considered only matching SVs that were tested with ROS/MAP phenotypes and reached a threshold of nominal $P \leq 0.05$ in at least one of the phenotypes tested.

Results

Characteristics of the samples

Genotyped structural variant calls were obtained on 1088 non-Latino white subjects from the ROS/MAP cohort studies [25]. In both studies, participants underwent annual clinical evaluations and donated their brains at death. The mean (SD) age at enrollment and age at death across the participants used in this study was 80.9 (6.8) and 89.0 (6.4) years, respectively, with an average (SD) follow-up period of 7.2 (4.9) years. Of all participants, 43.7% had a diagnosis of Alzheimer's dementia at death, and nearly two-thirds had pathologic AD confirmed post-mortem. A decline in cognition was observed, with the Mini-Mental State Examination (MMSE) score decreasing from 28 (IQR 26–29) at baseline to 25 (IQR 15–28) proximate to death. TDP-43 pathology extending beyond the amygdala was observed in just over a third of brains, and 13% presented Lewy bodies in nigra and/or cortex regions. Cerebrovascular diseases, including macroscopic infarcts and microinfarcts, were observed in more than a third and a quarter, respectively, and moderate to severe amyloid angiopathy, atherosclerosis, and arteriolosclerosis were observed in about a third of the brains.

Genome-wide SV association scans with clinical and neuropathologic phenotypes

Genome-wide association scans of structural variants were performed on a range of clinical and neuropathologic phenotypes covering multiple clinical and pathologic variables related to aging and dementia. Twenty-four phenotypes were analyzed, including clinical diagnosis of AD, MCI, and MDD; depressive symptomatology; measurements of cognitive function, motor function, frailty, and parkinsonian scores; and neuropathologies, such as β -amyloid, tangles, TDP-43, Lewy bodies, and multiple cerebrovascular diseases

indices. SV discovering and genotyping pipelines were obtained from combined ROS/MAP samples as previously described [24]. SVs were classified into different classes of variation, including deletions (DEL), insertions (INS), duplications (DUP), inversions (INV), complex rearrangements (CPX), and three classes of mobile element insertions (MEI), *Alu*, SVA, and LINE1. A total of 72,348 SVs were initially mapped. For the association analysis, variants with Hardy–Weinberg equilibrium (HWE) P -value lower than 10^{-6} were removed. Single variant association tests were applied across 16 quantitative and eight binary traits (Additional file 1: Table S2) for 20,022 SVs in ROS and 20,078 SVs in MAP with minor allele count (MAC) greater than 10 in each cohort. Scans were performed using the tool SAIGEgds (Scalable and Accurate Implementation of GEneralized mixed model) [68, 69]. All tests were controlled by age at death, sex, years of education, five genetic principal components, and the genetic correlation matrix (modeled as a random effect). A meta-analysis was performed combining results from both ROS and MAP cohorts. No SV reached genome-wide significance ($P < 5 \times 10^{-8}$) for any of the phenotypes tested at the current sample size. Complete summary stats for each phenotype are provided in Additional files 2–5.

SVs in AD

SVs in AD GWAS loci

To investigate the impact of SVs in known AD GWAS loci, we examined 81 genome-wide significant loci collectively identified in six previous studies that did not examine SVs [77]. We first mapped the presence of SVs (discovered in ROS/MAP) in each locus. On average, 10 SVs were identified by locus across all 81 loci (Additional file 1: Fig. S2). As expected, complex genomic regions, such as the *HLA* locus, harbored a considerably higher number of SVs [75], followed by the *ABCA7* [47], the *TMEM121* [35], and *IDUA* [32] loci. 36 SVs were in LD with the lead variant in 10 of the 81 loci, with R^2 ranging from 0.217 to 0.964. Among these, 22 SVs were nominally associated ($P \leq 0.05$) with at least one of the 24 AD/ADRD phenotypes tested (Fig. 1 and Additional file 1: Table S3).

The SV with the strongest results ($P = 7.72 \times 10^{-4}$) was a 343-bp deletion (Fig. 2C), which deletes an *Alu* element at the 3'UTR of *TMEM106B* [78] and was in high LD with the lead variant at the *TMEM106B* locus (rs5011436; $R^2 = 0.96$) (Fig. 2A). In ROS/MAP participants, this SV was associated with multiple AD/ADRD phenotypes, including tangles density, cognitive resilience, TDP-43, and others, illustrating the pleiotropy among SVs (Fig. 2B). We further collected information on

SV-xQTL and found that the deletion was also associated with lower protein abundance of *TMEM106B* (Fig. 2D).

Another four associations were found with a relaxed $P < 0.01$. That included a 22,029-bp-long deletion (Fig. 3A) associated with diagnosis of MDD in ROS/MAP ($P = 0.0025$) and in LD ($R^2 = 0.44$) with the lead variant in the *IQCK* locus (chr16:19,308,163–20,308,163) (Fig. 3B). Two SVs at the *HLA* locus, an 86,768 bp deletion ($R^2 = 0.27$) and a 43,223-bp duplication ($R^2 = 0.39$) associated with cognitive resilience ($P = 0.002$) and MDD ($P = 0.003$), respectively. And one 1505 bp deletion at the *MYO15A* locus ($R^2 = 0.37$), associated with TDP-43 ($P = 0.007$). Although the AD association at the *IQCK* locus was identified only in one paper [2], nominally significant associations were also found in more powered AD GWAS [1] (Fig. 3C). In ROS/MAP, apart from depression status, cognitive resilience and cognitive decline also reached a nominal significant threshold (Fig. 3D). Depression is a well-known risk factor for AD, associated with an increased likelihood of developing dementia [79–82]. Although we did not find associations at the locus in GWAS studies for general depression [83–87], this association could represent a genetic link between late-life depression and AD. The SVs at the *HLA* locus overlap the genes *HLA-DRB1*, *HLA-DRB5*, and *HLA-DRB6*, characterizing the class II sub-region haplotypes. However, their association and LD patterns are overly complex to disentangle the causal variants from the haplotypes.

Replication and characterization of SVs previously linked with AD and other neurodegenerative diseases

To highlight another potential use of our catalog of SV-GWAS, we evaluated the replication of structural variants (SVs) identified in prior neurodegenerative disease GWAS. We collected results from five studies across different diseases, including Alzheimer's disease (AD) [17], Parkinson's disease (PD) [23], progressive supranuclear palsy (PSP) [21], Lewy body dementia (LBD) [20], and a combination of frontotemporal dementia and amyotrophic lateral sclerosis (FTD/ALS) [20]. The number of SVs tested in each study varied due to differing inclusion criteria for association testing, running from 19,248 SVs in AD to 3156 SVs tested in the PD GWAS (Fig. 4A). The PSP study reported 7 SVs reaching genome-wide significance ($P \leq 5 \times 10^{-8}$), followed by PD with 6, FTD/ALS with 2, LBD with 1. The AD GWAS performed by the ADSP consortium ($n = 12,908$) did not find SVs reaching genome-wide significance. Instead, they reported a list of SVs suggestively associated with AD (FDR < 0.20) [17].

To access the replication of associations, first, we matched SVs between each study and the ROS/MAP cohort using a reciprocal overlap criterion of 80%. On

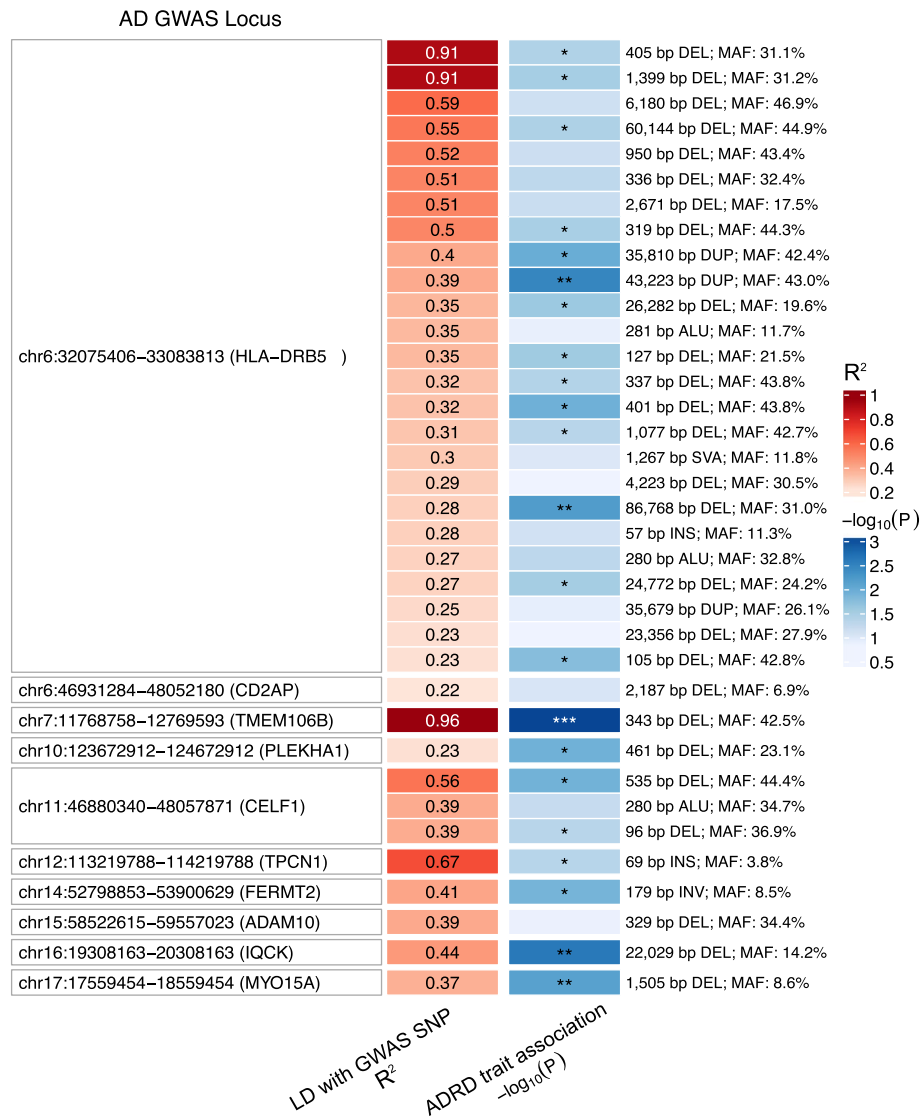


Fig. 1 Structural variants in LD with AD GWAS variants. Heatmap shows 36 SVs in LD ($R^2 > 0.2$) with 10 AD GWAS loci lead SNPs. The best association of each SV with ROS/MAP AD/ADRD phenotypes is highlighted in the blue colored column (shown as $-\log_{10}(\text{nominal } P)$; * $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$)

average, 59.16% of SVs were matched between ROS/MAP and the external studies, with overlaps ranging from 47.9% in AD to 69.4% in PD. After restricting to SVs that were tested for association in ROS/MAP, the final overlap included 5966 SVs for AD, 5704 for PSP, 3051 for LBD, 2805 for FTD/ALS, and 1815 for PD (Fig. 4B–F). The MAF correlations ranged from 0.59 in AD to 0.97 in LBD (Additional file 1: Fig S1). Among shared SVs that reached nominal significance ($P \leq 0.05$) for their respective traits and at least one of the 24 phenotypes tested in ROS/MAP, we found 223 ROS/MAP SVs matching with AD GWAS and mapping to 464 SV-trait associations (Fig. 4B), 107 SV-trait associations for 59 matching SVs

with PD GWAS (Fig. 4C), 603 SV-trait associations for 305 SVs matching PSP SVs (Fig. 4D); 175 SV-trait associations for 87 SVs matching LBD SVs (Fig. 4E), and 187 SV-trait associations for 99 SVs matching FTD/ALS SVs (Fig. 4F). Among the 16 SVs that reached genome-wide significance, 7 had matched SVs in ROS/MAP (6 from PSP and 1 from PD). Therefore, we also included SVs with a suggestive threshold of $P \leq 5 \times 10^{-4}$ for all studies, except for PSP, where we used $P \leq 5 \times 10^{-8}$, to compare with our results. This resulted in 59 candidate SVs (13 in AD, 15 in PD, 7 in PSP, 7 in LBD, and 17 in FTD/ALS), of which 24 had matched SVs tested in ROS/MAP (4 in AD, 4 in PD,

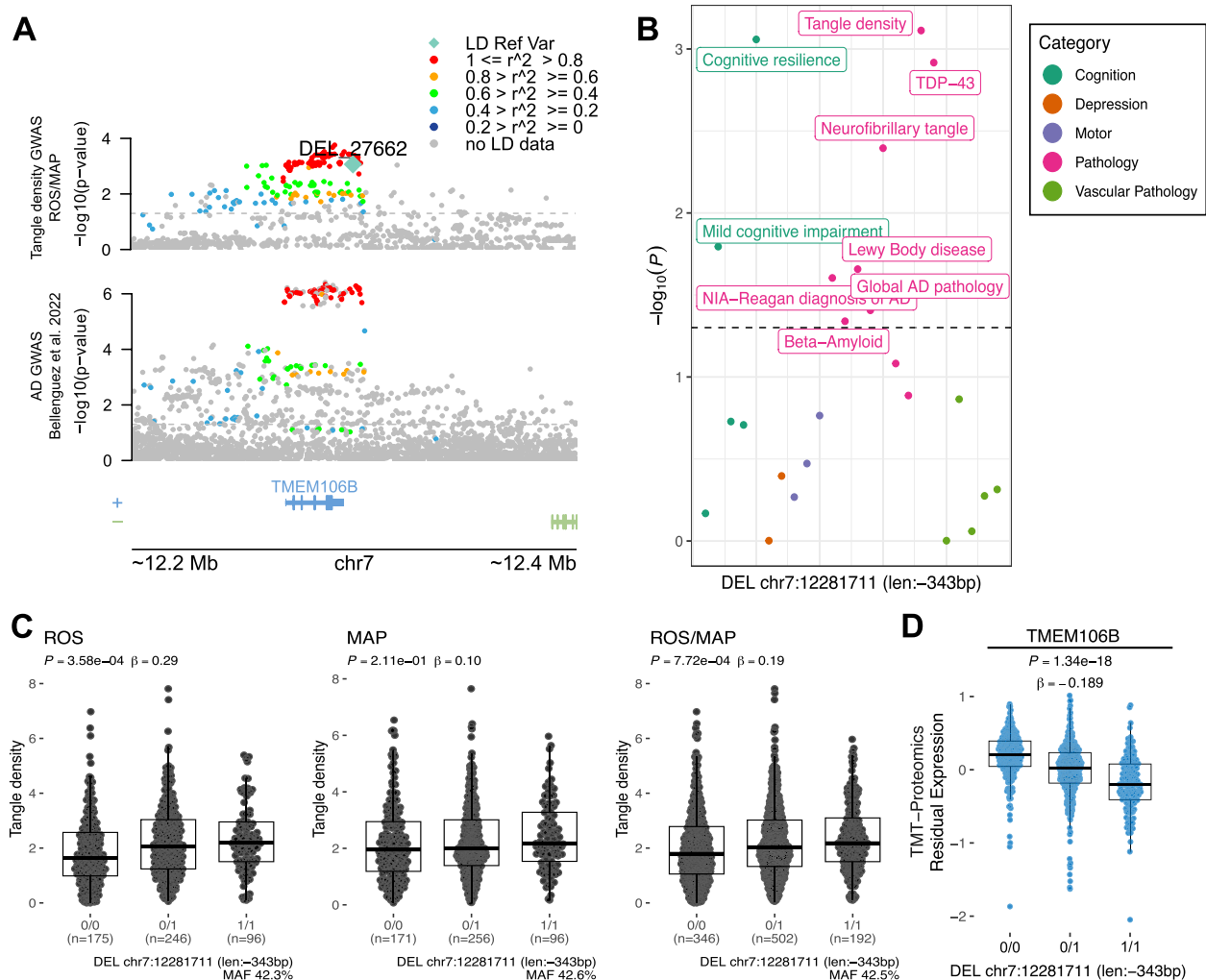


Fig. 2 Deletion associated with tangles density at *TMEM106B* locus. **A** Locus zoom plot for the *TMEM106B* locus (chr7:11,768,758–12,769,593) showing a 200-Kbp window around a 343-bp deletion with high LD with the lead SNP rs5011436. On the top, the y-axis shows the nominal P values (as $-\log_{10}$) for the association tests with tangles density in ROS/MAP participants. On the bottom, the y-axis shows the AD GWAS results from Bellenguez et al. [1]. The SV is plotted in a diamond shape, while SNPs are plotted in circles. Points are colored by the LD (R^2) to the SV. The dashed line represents nominal $P=0.05$. **B** shows the nominal P values (as $-\log_{10}$) for the association of the SV with all AD/ADRD traits tested. Dots are colored by phenotype category. The dashed line represents nominal $P=0.05$. **C** Boxplots showing the tangles density by the deletion alleles (ROS, MAP, ROS/MAP). **D** Boxplot shows the SV-pQTL between the deletion and protein expression levels of *TMEM106B* measured from DLPFC brain tissues of ROS/MAP participants

6 in PSP, 2 in LBD, and 8 in FTD/ALS). Among these, 16 SVs reached a nominal significance of $P \leq 0.05$ in at least one ROS/MAP phenotype (Table 1 and Additional file 1: Table S4). Notably, considering findings in ROS/MAP ($P \leq 5 \times 10^{-4}$) as discovery and SV GWAS studies as replication ($P \leq 0.05$), we found 15 SVs meeting the respective thresholds (Additional file 1: Table S5). Several notable findings emerged from this analysis. For instance, a 1483-bp deletion at the *CYP2A13* locus (chr19:4,110,280—GRCh38) significantly associated in the PSP GWAS ($P = 7.46 \times 10^{-9}$), showed associations with cognitive decline ($P = 1.9 \times 10^{-4}$) and another 4

phenotypes in ROS/MAP. Also, the top PSP associations in the *MAPT* locus (tagging the 1 Mb long inversion haplotype at 17q21.31), showed nominal association with motor function phenotypes ($P \leq 0.05$) in ROS/MAP. The *MAPT* inversion is also linked to multiple xQTLs, as we reported in our previous study [24]. Among other studies, we found a 994 bp duplication (chr12:25,590,144—GRCh38) in the *LMNTD1* locus, suggestively linked to AD and replicated with an association ($P = 3.3 \times 10^{-5}$) for neurofibrillary tangle density in ROS/MAP and another 5 phenotypes with $P \leq 0.05$. Similarly, a 3,958 bp deletion at the *DOCK5* locus linked to LBD ($P = 3.36 \times 10^{-4}$), was found to be associated

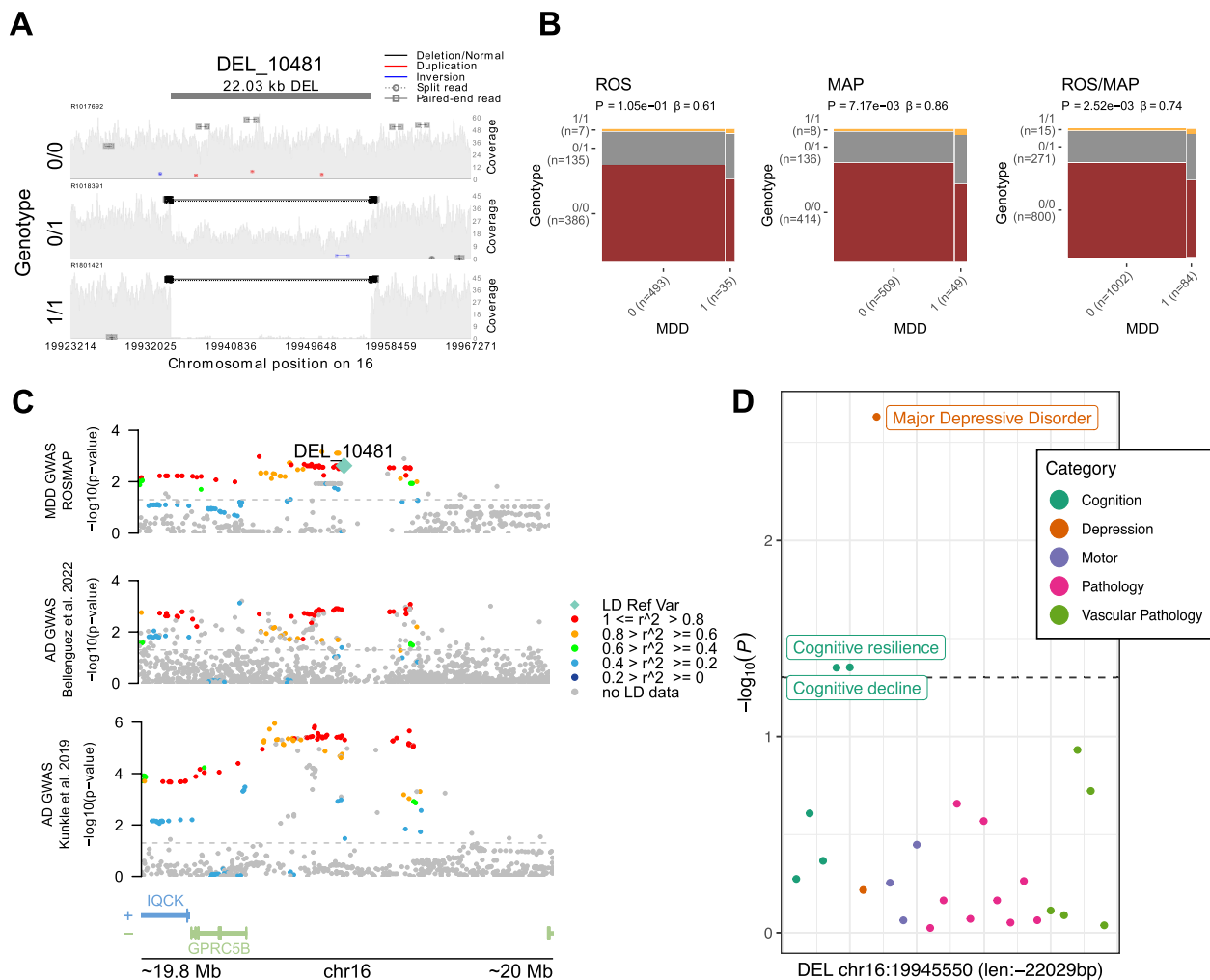


Fig. 3 Twenty-two Kbp deletion associated with major depressive disorder at *IQCK* locus. **A** Mapping of sequencing reads at the locus from three representative individuals corresponding to the possible genotypes: deletion not present (0/0), heterozygous deletion (0/1), and homozygous deletion (1/1). **B** Mosaic plots showing the proportion of individuals with MDD and respective deletion alleles for ROS, MAP, and ROS/MAP. **C**, locus zoom plot for the AD GWAS *IQCK* locus (chr16:19,308,163–20,308,163) showing a 200-Kbp window around the deletion in LD with the lead SNP rs7185636. The scatter plot on the top shows the nominal P values (as $-\log_{10}$) for the association tests with MDD status in ROS/MAP participants. The scatter plots in the middle and bottom show AD GWAS results from Bellenguez et al. (2022) [1] and Kunkle et al. (2019) [2], respectively. The 22-Kbp deletion is plotted in a diamond shape, while SNPs are plotted in circles. Points are colored by the LD (R^2) to the SV, as measured in ROS/MAP. The dashed line represents nominal $P=0.05$. **D** shows the nominal P values (as $-\log_{10}$) for the association of the 22 Kbp deletion with all AD/ADRD traits tested. Dots are colored by phenotype category. The dashed line represents nominal $P=0.05$

with motor function ($P=0.008$) and another 2 phenotypes. With these results, we demonstrate that leveraging our SV-GWAS catalogs not only enables the replication of SV associations with neurodegenerative diseases but also allows deeper insights into their potential roles across diverse phenotypes, while highlighting shared genetic architecture and candidate loci for further investigation.

Discussion

GWAS have enabled the identification of dozens of common SNVs and small indels (insertion-deletions) contributing to AD/ADRD traits [1, 87–93]. The contribution of

SVs, by contrast, lags. Previous studies were restricted mainly to either CNVs, as in autism [94]; rare SVs, as in schizophrenia [95–97]; or repeat expansions, as in amyotrophic lateral sclerosis (ALS), FTD, and Huntington's disease. Here, by leveraging the deep ROS/MAP phenotypic data, we performed for the first time a comprehensive analysis of the role of SVs in AD/ADRD traits by linking common SV alleles to multiple clinical and neuropathological traits.

We investigated the impact of SVs in known AD GWAS loci. By mapping the presence of SVs in each of the 81 loci, we identified 26 SVs in moderate to high LD

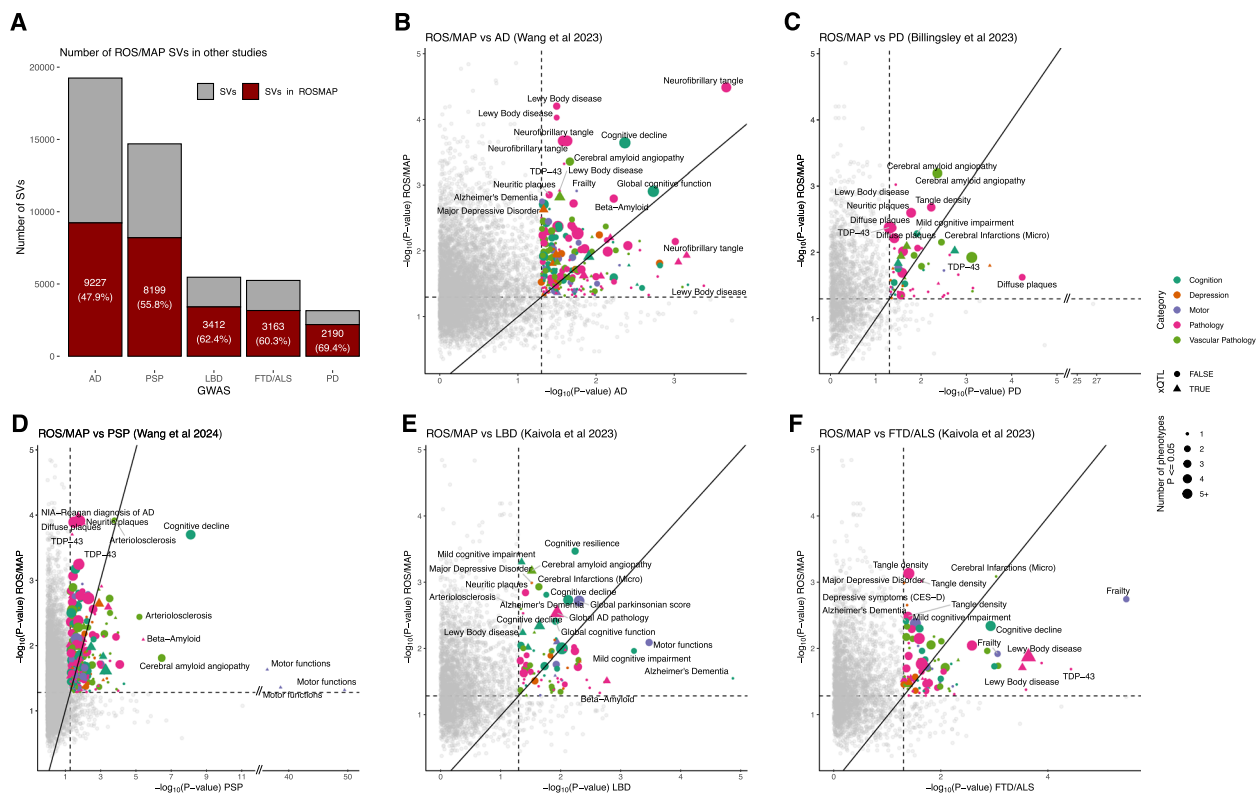


Fig. 4 Replication of SV associations with SV-GWAS for neuro diseases. **A** Total number of SVs tested in each GWAS study and the number of matching SVs also mapped in ROS/MAP. **B–F** Scatter plots showing matching SVs association P values (as $-\log_{10}$) between ROS/MAP SV-phenotypes (y-axis) and respective GWAS case-control studies (x-axis). **B** AD [17], **C** PD [23], **D** PSP [21], **E** LBD [20], **F** FTD/ALS [20]. In each scatter plot, colors represent the ROS/MAP phenotype category. For each SV, only the ROS/MAP phenotype with the lowest P value is shown, the size of the dots represents the number of phenotypes linked to the SV at $P \leq 0.05$. SVs also mapped as xQTL are shown as triangles. Dashed lines represent nominal $P = 0.05$

with GWAS lead SNPs and directly associated (nominal $P \leq 0.05$) with AD/ADRD phenotypes in ROS/MAP. The most significant SV was a 343-bp deletion at the 3'UTR of *TMEM106B*, associated with multiple phenotypes, including cognitive resilience, tangles density, and TDP-43, and in high LD with the lead variant at the locus. This SV has already been reported as being the likely causal variant at the locus, with reported risks for frontotemporal lobar dementia with TDP-43 inclusions (FTLD-TDP) [98] and neurodegeneration [99]. The proposed mechanism involved is via an aging-related negative feedback loop mediated by the presence of the *AluYb8* that dysregulates TDP-43 due to increasing demethylation of *TMEM106B* [99]. *TMEM106B* variants have also been reported to affect cell fraction in brain tissues from ROS/MAP, specifically for a subpopulation of excitatory neurons [100, 101].

Our analysis demonstrates the utility of the catalog of SV-GWAS for replicating and extending findings from

neurodegenerative disease GWAS. Across five studies (AD, PD, PSP, LBD, and FTD/ALS), we matched an average of 59.16% of SVs with ROS/MAP and identified 16 SVs showing nominal significance with at least one phenotype. Notable findings include the *CYP2A13* deletion, a PSP GWAS hit, which was associated with cognitive decline, and other phenotypes in ROS/MAP. Also, the *MAPT* haplotype inversion, the major PSP risk factor, replicated with motor function phenotypes. In AD, a duplication at the *LMNTD1* locus was associated with neurofibrillary tangle density, while a *DOCK5* deletion linked to LBD was associated with motor function and other traits. These findings highlight the value of deeply phenotyped cohorts like ROS/MAP in validating and expanding SV associations across neurodegenerative diseases. By focusing on SVs shared between studies and leveraging the extensive phenotypic data in ROS/MAP, we provide further evidence for the contribution of SVs to the genetic architecture of neurodegenerative diseases.

Table 1 Replication of SV-GWAS studies

SV-GWAS study			Replication in ROS/MAP			
SV (GRCh38)	P	Effect (SE) ^a	ROS/MAP Matching SV	Best SV-trait P and beta (SE)	Phenotypes ($P \leq 0.05$)	Locus
<i>Alzheimer's disease (Wang et al. 2023) [17]</i>						
DUP chr12:25590144 (len: -994 bp MAF: 4.2%)	2.16×10^{-4}	-0.05 (0.01)	DUP_950 chr12:25743078 (len: 833 bp MAF: 6.6%)	Neurofibrillary tangle $P = 3.28 \times 10^{-5}$ Beta (SE) = -0.16 (0.04)	Neurofibrillary tangle/tangle density/beta-amyloid/global AD pathology/NIA-Reagan diagnosis of AD/diffuse plaques	<i>LMNTD1</i>
DEL chr5:178679826 (len: -5012 bp MAF: 5.9%)	4.19×10^{-4}	-0.03 (0.01)	DEL_25213 chr5:178106851 (len: -5017 bp MAF: 2.8%)	Lewy body disease $P = 3.37 \times 10^{-2}$ Beta (SE) = 0.60 (0.68)	Lewy body disease	<i>ZNF354A</i>
<i>Parkinson's disease (Billingsley et al. 2023) [23]</i>						
INS chr3:189440563 (len: 143 bp MAF: 41.7%)	5.94×10^{-5}	0.79 (0.06)	INS_3482 chr3:189158352 (len: 145 bp MAF: 27.0%)	Diffuse plaques $P = 2.36 \times 10^{-2}$ Beta (SE) = -0.05 (0.02)	Diffuse plaques/NIA-Reagan diagnosis of AD	<i>TPRG1</i>
INS chr4:827662 (len: 134 bp MAF: 48.9%)	3.08×10^{-4}	0.82 (0.05)	INS_3579 chr4:821448 (len: 134 bp MAF: 47.6%)	Major Depressive Disorder $P = 1.57 \times 10^{-2}$ Beta (SE) = 0.40 (0.23)	Major depressive disorder	<i>CPLX1</i>
<i>Progressive supranuclear palsy (Wang et al. 2024) [21]</i>						
DEL chr17:46009357 (len: 238 bp MAF: 16.2%)	3.14×10^{-50}	-1.22 (0.08)	DEL_11943 chr17:44086723 (len: -242 bp MAF: 18.0%)	Motor functions $P = 4.61 \times 10^{-2}$ Beta (SE) = -0.03 (0.01)	Motor functions	<i>MAPT</i>
DEL chr17:46146541 (len: 314 bp MAF: 18.8%)	2.13×10^{-39}	-1.12 (0.09)	DEL_11947 chr17:44223907 (len: -347 bp MAF: 19.5%)	Motor functions $P = 4.19 \times 10^{-2}$ Beta (SE) = -0.03 (0.01)	Motor functions	<i>KANSL1</i>
DEL chr17:46099028 (len: 323 bp MAF: 21.9%)	3.88×10^{-37}	-1.07 (0.08)	DEL_11945 chr17:44176394 (len: -344 bp MAF: 19.3%)	Motor functions $P = 2.24 \times 10^{-2}$ Beta (SE) = -0.03 (0.01)	Motor functions	<i>KANSL1</i>
DEL chr19:41102802 (len: 1483 bp MAF: 17.1%)	7.46×10^{-9}	0.64 (0.11)	DEL_14138 chr19:41608707 (len: -1486 bp MAF: 12.3%)	Cognitive decline $P = 1.94 \times 10^{-4}$ Beta (SE) = -0.02 (0.01)	Cognitive decline/cognitive resilience/global cognitive function/Alzheimer's dementia/mild cognitive impairment	<i>CYP2A13</i>
<i>Lewy body dementia (Kaivola et al. 2023) [20]</i>						
DEL chr18:68414892 (len: 52 bp MAF: 4.2%)	1.35×10^{-5}	0.64 (0.10)	DEL_13123 chr18:66082127 (len: -54 bp MAF: 6.7%)	Alzheimer's dementia $P = 2.71 \times 10^{-2}$ Beta (SE) = 0.40 (0.18)	Alzheimer's dementia	<i>TMX3</i>
DEL chr8:25209169 (len: 3967 bp MAF: 18.3%)	3.36×10^{-4}	0.83 (0.05)	DEL_29898 chr8:25066678 (len: -3985 bp MAF: 18.3%)	Motor functions $P = 7.91 \times 10^{-3}$ Beta (SE) = -0.03 (0.01)	Motor functions/mild cognitive impairment/global parkinsonian score	<i>DOCK5</i>
<i>Frontotemporal dementia/amyotrophic lateral sclerosis (Kaivola et al. 2023) [20]</i>						
CPX chr17:45603799 (len: 673211 bp MAF: 18.3%)	3.48×10^{-6}	0.77 (0.06)	INV_1434 chr17:43671041 (len: 658522 bp MAF: 5.0%)	Frailty $P = 1.76 \times 10^{-3}$ Beta (SE) = 0.14 (0.18)	Frailty/mild cognitive impairment	<i>ARL17B</i>
DEL chr11:82145368 (len: 6935 bp MAF: 9.5%)	3.75×10^{-5}	1.34 (0.07)	DEL_4930 chr11:81856411 (len: -6938 bp MAF: 10.8%)	TDP-43 $P = 1.97 \times 10^{-2}$ Beta (SE) = -0.38 (0.16)	TDP-43	<i>FAM181B</i>
DEL chr9:110262382 (len: 5230 bp MAF: 14.6%)	1.32×10^{-4}	1.26 (0.06)	DEL_32128 chr9:113,024,577 (len: -5342 bp MAF: 20.8%)	NIA-Reagan diagnosis of AD $P = 1.60 \times 10^{-2}$ Beta (SE) = -0.22 (0.31)	NIA-Reagan diagnosis of AD	<i>TXN</i>
DEL chr2:183220800 (len: 5418 bp MAF: 5.1%)	2.33×10^{-4}	1.41 (0.09)	DEL_16436 chr2:184085521 (len: -5428 bp MAF: 4.7%)	Lewy body disease $P = 1.29 \times 10^{-2}$ Beta (SE) = 0.55 (0.59)	Lewy body disease/cerebral infarctions (gross)/beta-amyloid/depressive symptoms (CES-D)/frailty/motor functions	<i>NUP35</i>
DEL chr4:69257657 (len: 109005 bp MAF: 11.0%)	2.59×10^{-4}	1.28 (0.07)	DEL_21919 chr4:70123596 (len: -108693 bp MAF: 2.8%)	Lewy body disease $P = 4.03 \times 10^{-2}$ Beta (SE) = 0.73 (0.40)	Lewy body disease	<i>UGT2B28</i>
DEL chr19:36349389 (len: 6781 bp MAF: 13.8%)	3.03×10^{-4}	1.25 (0.06)	DEL_14072 chr19:36840751 (len: -6334 bp MAF: 19.9%)	Neurofibrillary tangle $P = 1.87 \times 10^{-2}$ Beta (SE) = 0.05 (0.02)	Neurofibrillary tangle/tangle density	<i>ZFP14</i>

^a Effects shown in terms of log(odds ratio) in all studies, with exception of AD and PSP, where values of beta were reported

While our results represent a step forward in understanding the effects of common genetic variation in AD/ADRD traits, important limitations must be noted: (1) the power for association discovery is constrained by the current sample size; (2) the replication of associations in independent samples is limited to available AD-related phenotypes and might not capture the same nuances from ROS/MAP; (3) SV calling is restricted to deletions, insertions, inversions, and duplication and is still prone to falsely discovered variants and low sensitivity (especially for insertions); (4) tandem repeats are not likely to be mapped in our data, since these require another specific set of tools for detection; (5) the suggestive associations do not represent suggestive causal effects on the traits, especially when LD is present, which would require a more precise fine-mapping analysis; (6) analyses were restricted to germline common autosomal structural variation; (7) since the individuals in this study have a European genetic background, these associations might not transfer to ancestrally diverse population-based data. These limitations could be overcome with additional sample size and deeper sequencing data (e.g., long-reads). Most importantly, these results should be interpreted as suggestive associations, and a more comprehensive replication in an independent sample data set is crucial for the validity of the findings.

Conclusions

Here, we performed genome-wide association analyses of SVs with 24 clinical and neuropathological phenotypes measured from two ongoing cohort studies of aging and dementia that benefit from having a comprehensive set of standardized phenotypes measured from the same individuals. We mapped the occurrence of SVs in 81 known AD GWAS loci and identified SVs in LD with respective lead variants in 10 of these loci. Among these, we highlighted a deletion of an *Alu* element in the *TMEM106B* locus, which was associated with tangles density, cognitive resilience, TDP-43, and also linked to protein abundance level changes in human brains (SV-pQTL). We also highlighted a 22-Kb deletion at the *IQCK* locus, associated with major depression disorder (a known risk factor for dementia). Further, we were able to replicate findings from other SV-GWAS studies, showing the capability of our results as a catalog for replication and characterization of the role of SVs in neurodegenerative diseases. Therefore, we believe this work will be of interest to the research community, not only for the findings described but also as a resource for future studies.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13073-025-01444-6>.

Additional file 1: Supplementary Material. Table S1. Demographic and clinical characteristics of study cohorts. Table S2. Variables and categories of each phenotype tested. Table S3. Table S4. Replication of SV-GWAS studies. Table S5. Replication of ROS/MAP SV-trait associations. Fig S1. Correlation of Minor Allele Frequencies between the ROS/MAP dataset and neurodegenerative disease datasets for structural variants. Fig S2. Overview of SVs in 81 AD GWAS loci

Additional file 2. AssociationResults_README.md. Description of columns for Additional files 2,3 and 4. Text file

Additional file 3. MAP_association_results.tsv.gz. Full summary statistics for SV-trait association in MAP cohort. Tabulated compressed file

Additional file 4. ROS_association_results.tsv.gz. Full summary statistics for SV-trait association in ROS cohort. Tabulated compressed file

Additional file 5. ROSMAP_meta_analysis_results.tsv.gz. An abstract or a condensed presentation of the substance of a body of material. Tabulated compressed file

Acknowledgements

We thank the participants of ROS/MAP cohorts for their essential contributions and gift to these projects, as well the investigators and staff at the Rush Alzheimer's Disease Center.

Authors' contributions

RAV performed the main analysis, drafted the manuscript, and designed the project. KDP, YL, and BN assisted on the analysis. APW, TSW, and NTF provided and processed proteomics data. ASB, YW, JMF, CG, ST, and DAB edited the manuscript and assisted in project design. LLB, JAS, PLD, CG, and DAB funded the project. All authors read and approved the final manuscript.

Funding

This work has been supported the following National Institute on Aging (NIA) grants: P30AG10161 (DAB), P30AG72975 (JAS), R01AG15819 (DAB), R01AG17917 (DAB), U01AG46152 (PLD, DAB), U01AG61356 (PLD, DAB), R01AG22018 (LLB), U01AG079847 (CG). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Data availability

No new sequencing data was generated as part of this current study. complete summary statistics for all phenotypes tested and all relevant code used in this study are publicly available on github (https://github.com/rushalz/adrd_sv_gwas) [102]. ros/map resources, including individual-level genotyping and phenotypic data, can be requested at <https://www.radc.rush.edu>.

Declarations

Ethics approval and consent to participate

The ROS and MAP studies were approved by an Institutional Review Board (IRB) of Rush University Medical Center, Chicago, IL. All participants agreed to an annual clinical evaluation and signed an informed consent and an Anatomic Gift Act agreeing to post-mortem brain donation. All procedures performed in studies involving human participants were in accordance with the ethical standards of the Institutional Review Board of Rush University Medical Center and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. Each participant signed an informed consent form to participate in the study.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Rush Alzheimer's Disease Center, Rush University Medical Center, 1750 W Harrison St, Chicago, IL 60612, USA. ²Department of Psychiatry, University of California, Davis, Davis, CA, USA. ³VA Northern California Health Care System, Davis, CA, USA. ⁴Department of Neurology, University of California, Davis, Davis, CA, USA. ⁵Department of Neurology and Department of Biochemistry, Goizueta Alzheimer's Disease Research Center, Emory University School of Medicine, Atlanta, GA, USA. ⁶Center for Translational and Computational Neuroimmunology, Department of Neurology, Columbia University Irving Medical Center, New York, NY, USA. ⁷Department of Psychiatry, SUNY Upstate Medical University, Syracuse, NY, USA.

Received: 3 September 2024 Accepted: 24 February 2025

Published online: 04 March 2025

References

- Bellenguez C, Küçükali F, Jansen IE, Kleiweidam L, Moreno-Grau S, Amin N, et al. New insights into the genetic etiology of Alzheimer's disease and related dementias. *Nat Genet.* 2022;54(4):412–36.
- Kunkle BW, Grenier-Boley B, Sims R, Bis JC, Damotte V, Naj AC, et al. Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing. *Nat Genet.* 2019;51(3):414–30.
- Wightman DP, Jansen IE, Savage JE, Shadrin AA, Bahrami S, Holland D, et al. A genome-wide association study with 1,126,563 individuals identifies new risk loci for Alzheimer's disease. *Nat Genet.* 2021;53(9):1276–82.
- Jansen IE, Savage JE, Watanabe K, Bryois J, Williams DM, Steinberg S, et al. Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nat Genet.* 2019;51(3):404–13.
- Schwartzentruber J, Cooper S, Liu JZ, Barrio-Hernandez I, Bello E, Kumasaka N, et al. Genome-wide meta-analysis, fine-mapping and integrative prioritization implicate new Alzheimer's disease risk genes. *Nat Genet.* 2021;53(3):392–402.
- de Rojas I, Moreno-Grau S, Tesi N, Grenier-Boley B, Andrade V, Jansen IE, et al. Common variants in Alzheimer's disease and risk stratification by polygenic risk scores. *Nat Commun.* 2021;12(1):3417.
- Collins RL, Brand H, Karczewski KJ, Zhao X, Alfoldi J, Francioli LC, et al. A structural variation reference for medical and population genetics. *Nature.* 2020;581(7809):444–51.
- Sudmant PH, Rausch T, Gardner EJ, Handsaker RE, Abyzov A, Huddleston J, et al. An integrated map of structural variation in 2,504 human genomes. *Nature.* 2015;526(7571):75–81.
- Chiang C, Scott AJ, Davis JR, Tsang EK, Li X, Kim Y, et al. The impact of structural variation on human gene expression. *Nat Genet.* 2017;49(5):692–9.
- Scott AJ, Chiang C, Hall IM. Structural variants are a major source of gene expression differences in humans and often affect multiple nearby genes. *Genome Res.* 2021;31(12):2249–57.
- Blom ES, Viswanathan J, Kilander L, Helisalmi S, Soininen H, Lannfelt L, et al. Low prevalence of APP duplications in Swedish and Finnish patients with early-onset Alzheimer's disease. *Eur J Hum Genet.* 2008;16(2):171–5.
- Hooli BV, Mohapatra G, Mattheisen M, Parrado AR, Roehr JT, Shen Y, et al. Role of common and rare APP DNA sequence variants in Alzheimer disease. *Neurology.* 2012;78(16):1250–7.
- Kasuga K, Shimohata T, Nishimura A, Shiga A, Mizuguchi T, Tokunaga J, et al. Identification of independent APP locus duplication in Japanese patients with early-onset Alzheimer disease. *J Neurol Neurosurg Psychiatry.* 2009;80(9):1050–2.
- Rovelet-Lecrux A, Hannequin D, Raux G, Le Meur N, Laquerrière A, Vital A, et al. APP locus duplication causes autosomal dominant early-onset Alzheimer disease with cerebral amyloid angiopathy. *Nat Genet.* 2006;38(1):24–6.
- Sleegers K, Brouwers N, Gijssels I, Theuns J, Goossens D, Wauters J, et al. APP duplication is sufficient to cause early onset Alzheimer's dementia with cerebral amyloid angiopathy. *Brain.* 2006;129(Pt 11):2977–83.
- Wang H, Wang L-S, Schellenberg G, Lee W-P. The role of structural variations in Alzheimer's disease and other neurodegenerative diseases. *Front Aging Neurosci.* 2022;14:1073905.
- Wang H, Dombroski BA, Cheng PL, Tucci A, Si YQ, Farrell JJ, et al. Structural variation detection and association analysis of whole-genome-sequence data from 16,905 Alzheimer's diseases sequencing project subjects. *medRxiv.* 2023.09.13.23295505.
- Baker M, Litvan I, Houlden H, Adamson J, Dickson D, Perez-Tur J, et al. Association of an extended haplotype in the tau gene with progressive supranuclear palsy. *Hum Mol Genet.* 1999;8(4):711–5.
- DeJesus-Hernandez M, Mackenzie IR, Boeve BF, Boxer AL, Baker M, Rutherford NJ, et al. Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron.* 2011;72(2):245–56.
- Kaivola K, Chia R, Ding J, Rasheed M, Fujita M, Menon V, et al. Genome-wide structural variant analysis identifies risk loci for non-Alzheimer's dementias. *Cell Genom.* 2023;3(6):100316.
- Wang H, Chang TS, Dombroski BA, Cheng P-L, Patil V, Valiente-Banueta L, et al. Whole-genome sequencing analysis reveals new susceptibility loci and structural variants associated with progressive supranuclear palsy. *Mol Neurodegener.* 2024;19(1):61.
- Zabetian CP, Hutter CM, Factor SA, Nutt JG, Higgins DS, Griffith A, et al. Association analysis of MAPT H1 haplotype and subhaplotypes in Parkinson's disease. *Ann Neurol.* 2007;62(2):137–44.
- Billingsley KJ, Ding J, Jerez PA, Illarionova A, Levine K, Grenn FP, et al. Genome-wide analysis of structural variants in Parkinson disease. *Ann Neurol.* 2023;93(5):1012–22.
- Vialle RA, de Paiva LK, Bennett DA, Cray JF, Raj T. Integrating whole-genome sequencing with multi-omic data reveals the impact of structural variants on gene regulation in the human brain. *Nat Neurosci.* 2022;25(4):504–14.
- Bennett DA, Buchman AS, Boyle PA, Barnes LL, Wilson RS, Schneider JA. Religious orders study and rush memory and aging project. *J Alzheimers Dis.* 2018;64(5):S161–89.
- Bennett DA, Schneider JA, Arvanitakis Z, Wilson RS. Overview and findings from the religious orders study. *Curr Alzheimer Res.* 2012;9(6):628–45.
- Bennett DA, Schneider JA, Buchman AS, Barnes LL, Boyle PA, Wilson RS. Overview and findings from the rush memory and aging project. *Curr Alzheimer Res.* 2012;9(6):646–63.
- Bennett DA, Schneider JA, Arvanitakis Z, Kelly JF, Aggarwal NT, Shah RC, et al. Neuropathology of older persons without cognitive impairment from two community-based studies. *Neurology.* 2006;66(12):1837–44.
- Bennett DA, Schneider JA, Buchman AS, Mendes de Leon C, Bienias JL, Wilson RS. The rush memory and aging project: study design and baseline characteristics of the study cohort. *Neuroepidemiology.* 2005;25(4):163–75.
- De Jager PL, Ma Y, McCabe C, Xu J, Vardarajan BN, Felsky D, et al. A multi-omic atlas of the human frontal cortex for aging and Alzheimer's disease research. *Sci Data.* 2018;7(5):180142.
- Bennett DA, Schneider JA, Aggarwal NT, Arvanitakis Z, Shah RC, Kelly JF, et al. Decision rules guiding the clinical diagnosis of Alzheimer's disease in two community-based cohort studies compared to standard practice in a clinic-based cohort study. *Neuroepidemiology.* 2006;27(3):169–76.
- McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's disease. *Neurology.* 1984;34(7):939–44.
- Bennett DA, Wilson RS, Schneider JA, Evans DA, Beckett LA, Aggarwal NT, et al. Natural history of mild cognitive impairment in older persons. *Neurology.* 2002;59(2):198–205.
- Schneider JA, Arvanitakis Z, Bang W, Bennett DA. Mixed brain pathologies account for most dementia cases in community-dwelling older persons. *Neurology.* 2007;69(24):2197–204.

35. Wilson RS, Boyle PA, Yang J, James BD, Bennett DA. Early life instruction in foreign language and music and incidence of mild cognitive impairment. *Neuropsychology*. 2015;29(2):292–302.
36. De Jager PL, Shulman JM, Chibnik LB, Keenan BT, Raj T, Wilson RS, et al. A genome-wide scan for common variants affecting the rate of age-related cognitive decline. *Neurobiol Aging*. 2012;33(5):1017.e1–15.
37. Oveisgharan S, Yang J, Yu L, Burba D, Bang W, Tasaki S, et al. Estrogen receptor genes, cognitive decline, and Alzheimer disease. *Neurology*. 2023;100(14):e1474–87.
38. Boyle PA, Wang T, Yu L, Wilson RS, Dawe R, Arfanakis K, et al. To what degree is late life cognitive decline driven by age-related neuropathologies? *Brain*. 2021;144(7):2166–75.
39. Buchman AS, Shulman JM, Nag S, Leurgans SE, Arnold SE, Morris MC, et al. Nigral pathology and parkinsonian signs in elders without Parkinson disease. *Ann Neurol*. 2012;71(2):258–66.
40. Buchman AS, Boyle PA, Wilson RS, Tang Y, Bennett DA. Frailty is associated with incident Alzheimer's disease and cognitive decline in the elderly. *Psychosom Med*. 2007;69(5):483–9.
41. Buchman AS, Wilson RS, Bienias JL, Bennett DA. Change in frailty and risk of death in older persons. *Exp Aging Res*. 2009;35(1):61–82.
42. Buchman AS, Boyle PA, Wilson RS, Fleischman DA, Leurgans S, Bennett DA. Association between late-life social activity and motor decline in older adults. *Arch Intern Med*. 2009;169(12):1139–46.
43. Bennett DA, Wilson RS, Schneider JA, Bienias JL, Arnold SE. Cerebral infarctions and the relationship of depression symptoms to level of cognitive functioning in older persons. *Am J Geriatr Psychiatry*. 2004;12(2):211–9.
44. Kohout FJ, Berkman LF, Evans DA, Cornoni-Huntley J. Two shorter forms of the CES-D (center for epidemiological studies depression) depression symptoms index. *J Aging Health*. 1993;5(2):179–93.
45. Wilson RS, Barnes LL, Mendes de Leon CF, Aggarwal NT, Schneider JS, Bach J, et al. Depressive symptoms, cognitive decline, and risk of AD in older persons. *Neurology*. 2002;59(3):364–70.
46. Wilson RS, Capuano AW, Boyle PA, Hoganson GM, Hizez LP, Shah RC, et al. Clinical-pathologic study of depressive symptoms and cognitive decline in old age. *Neurology*. 2014;83(8):702–9.
47. Boyle PA, Yu L, Leurgans SE, Wilson RS, Brookmeyer R, Schneider JA, et al. Attributable risk of Alzheimer's dementia attributed to age-related neuropathologies. *Ann Neurol*. 2019;85(1):114–24.
48. Bennett DA, Wilson RS, Schneider JA, Evans DA, Aggarwal NT, Arnold SE, et al. Apolipoprotein E epsilon4 allele, AD pathology, and the clinical expression of Alzheimer's disease. *Neurology*. 2003;60(2):246–52.
49. Bennett DA, Schneider JA, Tang Y, Arnold SE, Wilson RS. The effect of social networks on the relation between Alzheimer's disease pathology and level of cognitive function in old people: a longitudinal cohort study. *Lancet Neurol*. 2006;5(5):406–12.
50. Wilson RS, Arnold SE, Schneider JA, Tang Y, Bennett DA. The relationship between cerebral Alzheimer's disease pathology and odour identification in old age. *J Neurol Neurosurg Psychiatry*. 2007;78(1):30–5.
51. Nag S, Yu L, Wilson RS, Chen E-Y, Bennett DA, Schneider JA. TDP-43 pathology and memory impairment in elders without pathologic diagnoses of AD or FTL. *Neurology*. 2017;88(7):653–60.
52. Schneider JA, Arvanitakis Z, Yu L, Boyle PA, Leurgans SE, Bennett DA. Cognitive impairment, decline and fluctuations in older community-dwelling subjects with Lewy bodies. *Brain*. 2012;135(Pt 10):3005–14.
53. Buchman AS, Leurgans SE, Nag S, Bennett DA, Schneider JA. Cerebrovascular disease pathology and parkinsonian signs in old age. *Stroke*. 2011;42(11):3183–9.
54. Arvanitakis Z, Capuano AW, Leurgans SE, Buchman AS, Bennett DA, Schneider JA. The relationship of cerebral vessel pathology to brain microinfarcts. *Brain Pathol*. 2017;27(1):77–85.
55. Boyle PA, Yu L, Nag S, Leurgans S, Wilson RS, Bennett DA, et al. Cerebral amyloid angiopathy and cognitive outcomes in community-based older persons. *Neurology*. 2015;85(22):1930–6.
56. Arvanitakis Z, Leurgans SE, Barnes LL, Bennett DA, Schneider JA. Microinfarct pathology, dementia, and cognitive systems. *Stroke*. 2011;42(3):722–7.
57. Schneider JA, Bienias JL, Wilson RS, Berry-Kravis E, Evans DA, Bennett DA. The apolipoprotein E epsilon4 allele increases the odds of chronic cerebral infarction [corrected] detected at autopsy in older persons. *Stroke*. 2005;36(5):954–9.
58. Rausch T, Zichner T, Schlattl A, Stütz AM, Benes V, Korbel JO. DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics*. 2012;28(18):i333–9.
59. Layer RM, Chiang C, Quinlan AR, Hall IM. LUMPY: a probabilistic framework for structural variant discovery. *Genome Biol*. 2014;15(6):R84.
60. Chen X, Schulz-Trieglaff O, Shaw R, Barnes B, Schlesinger F, Källberg M, et al. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics*. 2016;32(8):1220–2.
61. Chen K, Wallis JW, McLellan MD, Larson DE, Kalicki JM, Pohl CS, et al. BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. *Nat Methods*. 2009;6(9):677–81.
62. Abyzov A, Urban AE, Snyder M, Gerstein M. CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. *Genome Res*. 2011;21(6):974–84.
63. Abyzov A, Li S, Kim DR, Mohiyuddin M, Stütz AM, Parrish NF, et al. Analysis of deletion breakpoints from 1,092 humans reveals details of mutation mechanisms. *Nat Commun*. 2015;6:7256.
64. Gardner EJ, Lam VK, Harris DN, Chuang NT, Scott EC, Pittard WS, et al. The mobile element locator tool (MELT): population-scale mobile element discovery and biology. *Genome Res*. 2017;27(11):1916–29.
65. Jeffares DC, Jolly C, Hoti M, Speed D, Shaw L, Rallis C, et al. Transient structural variations have strong effects on quantitative traits and reproductive isolation in fission yeast. *Nat Commun*. 2017;8:14061.
66. Pedersen B, Layer R, Quinlan AR. smooove: structural variant calling and genotyping with existing tools. GitHub repository. 2021. <https://github.com/brentp/smoove>.
67. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM. Robust relationship inference in genome-wide association studies. *Bioinformatics*. 2010;26(22):2867–73.
68. Zheng X, Davis JW. SAIGEgds—an efficient statistical tool for large-scale PheWAS with mixed models. *Bioinformatics*. 2021;37(5):728–30.
69. Zhou W, Nielsen JB, Fritsche LG, Dey R, Gabrielsen ME, Wolford BN, et al. Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nat Genet*. 2018;50(9):1335–41.
70. Han B, Eskin E. Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies. *Am J Hum Genet*. 2011;88(5):586–98.
71. Seifar F, Fox EJ, Shantaraman A, Liu Y, Dammer EB, Modeste E, et al. Large-scale deep proteomic analysis in Alzheimer's disease brain regions across race and ethnicity. *bioRxiv*. 2024.04.22.590547.
72. Higginbotham L, Carter EK, Dammer EB, Haque RU, Johnson ECB, Duong DM, et al. Unbiased classification of the human brain proteome resolves distinct clinical and pathophysiological subtypes of cognitive impairment. *bioRxiv*. 2022.07.22.501017.
73. Wingo AP, Fan W, Duong DM, Gerasimov ES, Dammer EB, Liu Y, et al. Shared proteomic effects of cerebral atherosclerosis and Alzheimer's disease on the human brain. *Nat Neurosci*. 2020;23(6):696–700.
74. Yu L, Tasaki S, Schneider JA, Arfanakis K, Duong DM, Wingo AP, et al. Cortical proteins associated with cognitive resilience in community-dwelling older persons. *JAMA Psychiat*. 2020;77(11):1172–80.
75. Lawrence M, Gentleman R, Carey V. rtracklayer: an R package for interfacing with genome browsers. *Bioinformatics*. 2009;25(14):1841–2.
76. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26(6):841–2.
77. Andrews SJ, Renton AE, Fulton-Howard B, Podlesny-Drabiniok A, Marcora E, Goate AM. The complex genetic architecture of Alzheimer's disease: novel insights and future directions. *EBioMedicine*. 2023;90(104511):104511.
78. Rodney A, Karanjeet K, Benzow K, Koob MD. A common Alu insertion in the 3'UTR of TMEM106B is associated with risk of dementia. *Alzheimers Dement*. 2024;20(7):5071–7.
79. Bellou V, Belbasis L, Tzoulaki I, Middleton LT, Ioannidis JPA, Evangelou E. Systematic evaluation of the associations between environmental risk factors and dementia: an umbrella review of systematic reviews and meta-analyses. *Alzheimers Dement*. 2017;13(4):406–18.
80. Dafsari FS, Jessen F. Depression—an underrecognized target for prevention of dementia in Alzheimer's disease. *Transl Psychiatry*. 2020;10(1):160.

81. Harerimana NV, Liu Y, Gerasimov ES, Duong D, Beach TG, Reiman EM, et al. Genetic evidence supporting a causal role of depression in Alzheimer's disease. *Biol Psychiatry*. 2022;92(1):25–33.
82. Holmquist S, Nordström A, Nordström P. The association of depression with subsequent dementia diagnosis: a Swedish nationwide cohort study from 1964 to 2016. *PLoS Med*. 2020;17(1):e1003016.
83. Coleman JRI, Peyrot WJ, Purves KL, Davis KAS, Rayner C, Choi SW, et al. Genome-wide gene-environment analyses of major depressive disorder and reported lifetime traumatic experiences in UK Biobank. *Mol Psychiatry*. 2020;25(7):1430–46.
84. Adams MJ, Thorp JG, Jermy BS, Kwong ASF, Köiv K, Grotzinger AD, et al. Genome-wide meta-analysis of ascertainment and symptom structures of major depression in case-enriched and community cohorts. *Psychol Med*. 2024;54(12):3459–68.
85. Meng X, Navoly G, Giannakopoulou O, Levey DF, Koller D, Pathak GA, et al. Multi-ancestry genome-wide association study of major depression aids locus discovery, fine mapping, gene prioritization and causal inference. *Nat Genet*. 2024;56(2):222–33.
86. Wray NR, Ripke S, Mattheisen M, Trzaskowski M, Byrne EM, Abdellaoui A, et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat Genet*. 2018;50(5):668–81.
87. Howard DM, Adams MJ, Clarke T-K, Hafferty JD, Gibson J, Shirali M, et al. Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nat Neurosci*. 2019;22(3):343–52.
88. Chia R, Sabir MS, Bandres-Ciga S, Saez-Atienzar S, Reynolds RH, Gustavsson E, et al. Genome sequencing analysis identifies new loci associated with Lewy body dementia and provides insights into its genetic architecture. *Nat Genet*. 2021;53(3):294–303.
89. Farrell K, Kim S, Han N, Iida MA, Gonzalez EM, Otero-Garcia M, et al. Genome-wide association study and functional validation implicates JADE1 in tauopathy. *Acta Neuropathol*. 2022;143(1):33–53.
90. Nalls MA, Blauwendraat C, Vallerga CL, Heilbron K, Bandres-Ciga S, Chang D, 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*. 2019;18(12):1091–102.
91. Sherva R, Zhang R, Sahelijo N, Jun G, Anglin T, Chanfreau C, et al. African ancestry GWAS of dementia in a large military cohort identifies significant risk loci. *Mol Psychiatry*. 2023;28(3):1293–302.
92. Zhang S, Cooper-Knock J, Weimer AK, Shi M, Moll T, Marshall JNG, et al. Genome-wide identification of the genetic basis of amyotrophic lateral sclerosis. *Neuron*. 2022;110(6):992–1008.e11.
93. Yan Q, Nho K, Del-Aguila JL, Wang X, Risacher SL, Fan KH, et al. Genome-wide association study of brain amyloid deposition as measured by Pittsburgh Compound-B (PiB)-PET imaging. *Mol Psychiatry*. 2021;26(1):309–21.
94. Marshall CR, Noor A, Vincent JB, Lionel AC, Feuk L, Skaug J, et al. Structural variation of chromosomes in autism spectrum disorder. *Am J Hum Genet*. 2008;82(2):477–88.
95. Halvorsen M, Huh R, Oskolkov N, Wen J, Netotea S, Giusti-Rodriguez P, et al. Increased burden of ultra-rare structural variants localizing to boundaries of topologically associated domains in schizophrenia. *Nat Commun*. 2020;11(1):1842.
96. Sebat J, Levy DL, McCarthy SE. Rare structural variants in schizophrenia: one disorder, multiple mutations; one mutation, multiple disorders. *Trends Genet*. 2009;25(12):528–35.
97. Walsh T, McClellan JM, McCarthy SE, Addington AM, Pierce SB, Cooper GM, et al. Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science*. 2008;320(5875):539–43.
98. Chemparathy A, Le Guen Y, Zeng Y, Gorzynski J, Jensen T, Yang C, et al. A 3'UTR insertion is a candidate causal variant at the *TMEM106B* locus associated with increased risk for FTLD-TDP. *bioRxiv*. 2023.07.06.23292312.
99. Salazar A, Tesi N, Knoop L, Pijnenburg Y, van der Lee S, Wijesekera S, et al. An AluYb8 retrotransposon characterises a risk haplotype of *TMEM106B* associated in neurodegeneration. *bioRxiv*. 2023.07.16.23292721.
100. Fujita M, Gao Z, Zeng L, McCabe C, White CC, Ng B, et al. Cell subtype-specific effects of genetic variation in the Alzheimer's disease brain. *Nat Genet*. 2024;56(4):605–14.
101. Green GS, Fujita M, Yang HS, Taga M, McCabe C, Cain A, et al. Cellular dynamics across aged human brains uncover a multicellular cascade leading to Alzheimer's disease. *bioRxiv*. 2023. <https://doi.org/10.1101/2023.03.07.531493>.
102. Vialle RA. *ADRD_SV_GWAS*. GitHub Repository. 2024. https://github.com/RushAlz/ADRD_SV_GWAS.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.