## Title

Whole-genome sequencing analysis reveals new susceptibility loci and structural variants associated with progressive supranuclear palsy



## Journal

Molecular Neurodegeneration, 19(1)

## ISSN

1750-1326

## Authors

Wang, Hui
Chang, Timothy S
Dombroski, Beth A
et al.

## Publication Date

2024-08-01

## DOI

10.1186/s13024-024-00747-3

Peer reviewed

Molecular Neurodegeneration

RESEARCH ARTICLE

Open Access

# Whole-genome sequencing analysis reveals new susceptibility loci and structural variants associated with progressive supranuclear palsy

Hui Wang[1,2†], Timothy S. Chang[3†], Beth A. Dombroski[1,2], Po-Liang Cheng[1,2], Vishakha Patil[3], Leopoldo Valiente-Banuet[3], Kurt Farrell[4], Catriona Mclean[5], Laura Molina-Porcel[6,7], Alex Rajput[8], Peter Paul De Deyn[9,10], Nathalie Le Bastard[11], Marla Gearing[12], Laura Donker Kaat[13], John C. Van Swieten[13], Elise Dopper[13], Bernardino F. Ghetti[14], Kathy L. Newell[14], Claire Troakes[15], Justo G. de Yébenes[16], Alberto Rábano-Gutierrez[17], Tina Meller[18], Wolfgang H. Oertel[18], Gesine Respondek[19], Maria Stamelou[20,21], Thomas Arzberger[22,23], Sigrun Roeber[24], Ulrich Müller[24], Franziska Hopfner[43], Pau Pastor[25,26], Alexis Brice[27], Alexandra Durr[27], Isabelle Le Ber[27], Thomas G. Beach[28], Geidy E. Serrano[28], Lili-Naz Hazrati[29], Irene Litvan[30], Rosa Rademakers[31,32], Owen A. Ross[32], Douglas Galasko[30], Adam L. Boxer[33], Bruce L. Miller[33], Willian W. Seeley[33], Vivanna M. Van Deerlin[1], Edward B. Lee[1,34], Charles L. White III[35], Huw Morris[36], Rohan de Silva[37], John F. Crary[4], Alison M. Goate[38], Jeffrey S. Friedman[39], Yuk Yee Leung[1,2], Giovanni Coppola[3,40], Adam C. Naj[1,2,41], Li-San Wang[1,2], P. S. P. genetics study group, Clifton Dalgard[42], Dennis W. Dickson[32*], Günter U. Höglinger[43*], Gerard D. Schellenberg[1,2*], Daniel H. Geschwind[3,44,45*] and Wan-Ping Lee[1,2*]

## Abstract

**Background** Progressive supranuclear palsy (PSP) is a rare neurodegenerative disease characterized by the accumulation of aggregated tau proteins in astrocytes, neurons, and oligodendrocytes. Previous genome-wide association studies for PSP were based on genotype array, therefore, were inadequate for the analysis of rare variants as well as larger mutations, such as small insertions/deletions (indels) and structural variants (SVs).

**Method** In this study, we performed whole genome sequencing (WGS) and conducted association analysis for single nucleotide variants (SNVs), indels, and SVs, in a cohort of 1,718 cases and 2,944 controls of European ancestry. Of the 1,718 PSP individuals, 1,441 were autopsy-confirmed and 277 were clinically diagnosed.

†Hui Wang and Timothy S. Chang contributed equally to this work.

*Correspondence:
Dennis W. Dickson
Dickson.Dennis@mayo.edu
Günter U. Höglinger
guenter.hoeglinger@med.uni-muenchen.de
Gerard D. Schellenberg
gerardsc@pennmedicine.upenn.edu
Daniel H. Geschwind
dhg@mednet.ucla.edu
Wan-Ping Lee
wan-ping.lee@pennmedicine.upenn.edu
Full list of author information is available at the end of the article

**Results** Our analysis of common SNVs and indels confirmed known genetic loci at *MAPT*, *MOBP*, *STX6*, *SLCO1A2*, *DUSP10*, and *SP1*, and further uncovered novel signals in *APOE*, *FCHO1/MAP1S, KIF13A, TRIM24, TNXB, and ELOVL1*. Notably, in contrast to Alzheimer's disease (AD), we observed the *APOE* ε2 allele to be the risk allele in PSP. Analysis of rare SNVs and indels identified significant association in *ZNF592* and further gene network analysis identified a module of neuronal genes dysregulated in PSP. Moreover, seven common SVs associated with PSP were observed in the H1/H2 haplotype region (17q21.31) and other loci, including *IGH*, *PCMT1*, *CYP2A13*, and *SMCP*. In the H1/H2 haplotype region, there is a burden of rare deletions and duplications ($P = 6.73 \times 10^{-3}$) in PSP.

**Conclusions** Through WGS, we significantly enhanced our understanding of the genetic basis of PSP, providing new targets for exploring disease mechanisms and therapeutic interventions.

**Keywords** Progressive Supranuclear Palsy (PSP), Whole-Genome Sequencing (WGS), Genome-Wide Association Study (GWAS), Structural Variants (SVs), Apolipoprotein E (APOE)

## Background

Progressive supranuclear palsy (PSP) is a neurodegenerative disease that is pathologically defined by the accumulation of aggregated tau protein in multiple cortical and subcortical regions, especially involving the basal ganglia, dentate nucleus of the cerebellum midbrain [1]. An isoform of tau harboring 4 repeats of microtubule-binding domain (4R-tau) is particularly prominent in these tau aggregates [2]. Clinical manifestations of PSP include a range of phenotypes, including the initially described and most common, PSP-Richardson syndrome that presents with multiple features, including postural instability, vertical supranuclear palsy, and frontal dementia. However, there are several other phenotypes, such as PSP-Parkinsonism, PSP-Frontotemporal dementia, PSP-freezing of gait, PSP-speech and language disturbances, etc. [3]. Presentation of these phenotypes varies widely depending on the distribution and severity of the pathology [4–6].

Currently, the most recognized genetic risk locus for PSP is at the H1/H2 haplotype region covering *MAPT* gene at chromosome 17q21.31 [7], where individuals carrying the common H1 haplotype are more likely to develop PSP with an estimated odds ratio (OR) of 5.6 [8]. Previous studies usually ascribed the observed association in the H1/H2 haplotype to *MAPT* [7, 9, 10]. However, recent functional dissection of this region using multiple parallel reporter assays coupled to CRISPRi demonstrated multiple risk genes in the area in addition to *MAPT*, including *KANSL1* and *PLEKMHL1* [11]. Genome-wide association studies (GWASs) in PSP have identified common variants in *STX6*, *EIF2AK3*, *MOBP*, *SLCO1A2*, *DUSP10*, *RUNX2*, and *LRRK2* with moderate effect size [8, 12–14]. In addition, variants in *TRIM11* were identified as a genetic modifier of the PSP phenotype when comparing PSP with Richardson syndrome to PSP without Richardson syndrome [15].

To date, no comprehensive analysis of single nucleotide variants (SNVs), small insertions and deletions (indels), and structural variants (SVs) in PSP by whole genome sequencing has been conducted. To gain a more comprehensive understanding of the genetic underpinnings of PSP, we performed whole genome sequencing (WGS) and analyzed SNVs, indels, and SVs. As a result, we validated previously reported genes and unveiled new loci that provide novel insights into the genetic basis of PSP.

## Methods

### Study subjects

We performed WGS at 30 × coverage for 1,834 PSP cases and 128 controls from the PSP-NIH-CurePSP-Tau, PSP-CurePSP-Tau, PSP-UCLA, and AMPAD-MAYO cohorts included in Alzheimer's Disease Sequencing Project (ADSP, NG00067.v7) and used 3,008 controls from other cohorts in ADSP (Table S1) [16]. Control subjects were self-identified as non-Hispanic white. WGS data is available on The National Institute on Aging Genetics of Alzheimer's Disease Data Storage Site (NIAGADS) [17]. Among 1,834 individuals with PSP, 1,488 were autopsy-confirmed and 346 were clinical diagnosed. 34 of the clinically diagnosed PSP had subsequent autopsy, of which 29 had confirmed PSP and five did not have PSP on autopsy. These five subjects without PSP on autopsy were removed. We also removed related subjects (identify by descent > 0.25) and non-Europeans (subjects that were eight standard deviations away from the 1000 Genomes Project European samples [18, 19] using the first six principal components (PCs)), resulting in 1,718 individuals with PSP and 2,944 control subjects. Of the 1,718 PSP individuals, 1,441 were autopsy-confirmed and 277 were clinically diagnosed (Table 1). Among 1,718 PSP cases, 740 samples were included in previous GWASs (386 samples in Höglinger et al. stage 1 analysis [8], 107 samples in Höglinger et al. stage 2 analysis [8], and 247 samples in Chen et al. [13]) Among 2,944 controls, 113 controls from PSP-UCLA cohort (Table S2) were included in Chen et al. [13].

**Table 1** Characteristics of study participants

| | PSP (*n* = 1,718) | | Control (*n* = 2,944) |
|---|---|---|---|
| | **Autopsy Confirmed (*n* = 1,441)** | **Clinical Diagnosed (*n* = 277)** | |
| **Female** | 625 (43%) | 129 (46%) | 1,775 (60%) |
| **Age, y (SD)** | 68.38 (8.22) | 65.72 (7.68) | 81.19 (6.01) |
| **AF of *APOE* ε4**[a] | 13% | 11% | 17% |
| *ε4 carriers* | 350 (24%) | 57 (21%) | 905 (32%) |
| *Non-ε4 carriers* | 1,085 (75%) | 216 (78%) | 1,913 (65%) |
| *Data missing* | 6 (0.42%) | 4 (1%) | 126 (4%) |
| **AF of *APOE* ε2**[b] | 9% | 7% | 4% |
| *ε2 carriers* | 234 (16%) | 36 (13%) | 220 (8%) |
| *Non-ε2 carriers* | 1,193 (83%) | 238 (86%) | 2,522 (86%) |
| *Data missing* | 14 (1%) | 3 (1%) | 202 (7%) |
| **AF of H2**[c] | 6% | 5% | 23% |
| *H2 carriers* | 158 (11%) | 27 (10%) | 1,182 (40%) |
| *Non-H2 carriers* | 1,283 (89%) | 250 (90%) | 1,761 (60) |
| *Data missing* | 0 (0%) | 0 (0%) | 1 (0.03%) |

*SD* Standard deviation, *AF* Allele frequency

[a] *APOE* ε4 is represented by the genotypes of rs429358-C

[b] *APOE* ε2 is represented by the genotypes of rs7412-T

[c] H2 haplotype is determined by the genotypes of rs8070723-G

Considering that our sample set incorporated external controls from ADSP, initially collected for Alzheimer's Disease (AD) studies, there was a potential selection biases for *APOE* ε4 and ε2 in controls. To rigorously validate our findings linked to *APOE*, we broke down the allele frequencies of *APOE* ε4 and ε2 by cohorts (Table S2), reviewed the study design of each cohort, and created an additional sample set by excluding those cohorts with selection bias against *APOE* ε4 or ε2 (Supplementary Methods).

### SNVs/indels quality controls

Only biallelic variants were included in common (Minor Allele Frequency [MAF] > 0.01) SNVs/indels analysis. A biallelic site is a locus in the genome that contains two observed alleles, i.e., one reference allele and one alternative allele. Variants were removed if they were monomorphic, did not pass variant quality score recalibration (VQSR), had an average read depth ≥ 500, or if all calls have alignment depth (DP < 10) and genotype quality (GQ < 20). Individual calls with DP < 10 or GQ < 20 were set to missing. Indels were left aligned using the GRCh38 reference [20, 21]. Common variants with a missing rate < 0.1, 0.25 < allele balance for heterozygous calls (ABHet) < 0.75, and Hardy-Weiberg Equilibrium tests (HWE) in controls > $1 \times 10^{-5}$ were kept for analysis, leaving 7,945,112 SNVs/indels for analysis. Similar quality control procedures were applied to rare variants (Supplementary Methods). Then, we calculated the heritability of PSP using GCTA-LDMS [22] for common SNVs/indels (MAF > 0.01) and common plus rare SNVs/indels. A prevalence of 5 PSP cases per 100,000 individuals (0.00005) was used in the GCTA-LDMS analysis.

### Common SNVs/indels analysis

For association analysis, linear mixed model implemented in R Genesis [23] were used. Genetic relatedness matrix was obtained using KING [24]. PCs were obtained by PC-AiR [25] which accounts for sample relatedness. Sex and PC1-5 were adjusted in the linear mixed model. Age was not adjusted as more than half (1,159 of 1,718) of PSP cases had age missing. SNVs and indels with a $P < 1 \times 10^{-6}$ were reported along with the WGS quality metrics, such as QualByDepth (QD) and FisherStrand (FS), (Table S3).

For H1/H2 region, fine-mapping were analyzed using SuSie [26]. We ran the analysis several times assuming the number of maximum causal variants were from 2 to 10. The only variant (rs242561) robust to the choice of maximum causal variants was reported. For major histocompatibility complex (MHC) region on chromosome 6, we imputed HLA alleles for HLA-A, HLA-B, HLA-C, HLA-DQB1 and HLA-DRB1 using CookHLA [27]. HLA alleles in linkage disequilibrium (LD) ($R^2 > 0.1$) with the most significant SNV (rs367364) in this region were reported (Table S9). Then, we used linear regression

models and performed association analysis for each HLA allele, adjusting for sex and PC1-5 (Table S10). To avoid potential confounding effects (particularly for *APOE* alleles), we also performed association analysis (Table S4, Table S5) for SNVs/indels with a $P < 1 \times 10^{-6}$ when excluding subjects from the three cohorts with selection bias against *APOE* alleles (ADSP-FUS1-APOEextremes, ADSP-FUS1-StEPAD1, and CacheCounty) along with cohorts with less than 10 subjects (NACC-Genentech, FASe-Families-WGS, and KnightADRC-WGS) (Table S2). We also performed additional experimental validation using TaqMan assay/Sanger sequencing to confirm the genotype of *APOE* observed from WGS (Supplementary Methods, Table S6).

#### Rare SNVs/indels analysis

For aggregated tests of rare variants, we considered rare protein truncating variants (PTVs) and PTVs/damaging missense variants. Variant were annotated with ANNOVAR (version 2020–06-07) [28] and Variant Effect Predictor (VEP, version 104.3) [29]. PTVs were in protein coding genes (Ensembl version 104) [30] and had VEP consequence as stop gained, splice acceptor, splice donor or frameshift. Damaging missense variants were in protein coding genes (Ensembl version 104) and had a VEP consequence as missense, CADD score ≥ 15, and PolyPhen-2 HDIV of probably damaging. Rare variants were selected based on a MAF < 0.01% from gnomAD and a MAF < 1% in our dataset. The number of alternative allele variants in PTVs and PTVs/damaging missense variants was similar across sequencing centers and when evaluated for loss of function intolerant genes (observed/expected score upper confidence interval < 0.35 [31]) (Fig. S14).

We performed SKAT-O and gene burden testing (SKATBinary, method = 'burden') for PTVs and PTV/damaging missense variants (Supplementary Methods). We also considered only PTVs or PTVs/missense variants in loss of function intolerant genes (observed/expected score upper confidence interval < 0.35 [31]) when performing the tests. *P*-values were FDR corrected for the number of genes with a total minor allele count (MAC) ≥ 10. As SKAT-O does not calculate an odds ratio, we calculated the odds ratio of significant genes using logistic regression with the same covariates as SKAT-O and burden testing, and the same variant weights.

We evaluated the C1 module, a gene set, which was previously shown to be composed of neuronal genes and enriched for common variants in PSP [32]. We performed a permutation test ($N = 1000$) of random gene set modules from brain expressed genes that contained the same number of genes as C1. From the human protein atlas (www.proteinatlas.org) [33], brain expressed genes were defined as the union of unique proteins from the cerebral cortex, basal ganglia and midbrain ($N = 15{,}638$). We calculated SKAT-O *P*-values from these random gene modules to determine the null distribution. We calculated the unadjusted odds ratio of significant genes or gene sets by summing the number of alternate alleles in the gene set among the total number alleles in cases and controls. Normalized quantification (TPM) gene expression across tissues was obtained from Genotype-Tissue Expression (GTEx) [34]. The expression of ZNF592 and C1 module (summarized as an eigengene [35]) were plotted.

#### SV detection and filtering

For each sample, SVs were called by Manta (v1.6.0) [36] and Smoove (v0.2.5) [37] with default parameters. Calls from Manta and Smoove were merged by Svimmer [38] to generate a union of two call sets for a sample. Then, all individual sample VCF files were merged together by Svimmer as input to Graphtyper2 (v2.7.3) [38] for joint genotyping. SV calls after joint-genotyping are comparable across the samples, therefore, can be used directly in genome-wide association analysis [38]. A subset of SV calls was defined as high-quality calls [38]. Details of SV calling pipeline were in our previous study [39].For each individual SV reported, Samplot [40] or IGV [41] were used to keep only high-confident CNVs and inversions that are supported by read depth or split reads; for insertions, we kept high-confident insertions that are high-quality and not in the masked regions (Supplementary Methods).

#### SV analysis

For SV association, more strict sample filtering was applied: outlier samples with too many (larger than median + 4*MAD) CNV/insertion calls or too little (smaller than median—4*MAD) high-quality CNV/insertion calls were removed. There were 4,432 samples (1,703 cases and 2,729 controls) remaining for PSP SV association analysis. Due to more false positives being picked up, the genomic inflation would be high ($\lambda = 1.89$, Fig. S9) if all SVs were included in the analysis. Therefore, we restricted our analysis to high-quality SVs only, making the genomic inflation drop to 1.27 (Fig. S9). The 14,792 high-quality common SVs (MAF > 0.1) with call rate > 0.5 were included in the analysis. Mixed model implemented in R Genesis were used for association. Sex, PCR information, SV PCs 1–5, and SNV PCs 1–5 were adjusted in the mixed model. After association, we manually inspect deletions, duplications, and inversions by Samplot or IGV to keep only those with support from read depth, split read or insert size. For insertions, those not on masked regions were reported.

For SVs inside the H1/H2 region, all SVs those that are not high-quality are included. Then, we removed SVs with missing rate > 0.5 and manual inspect deletions, duplications, and inversions by Samplot or IGV to keep only those with support from read depth, split read or insert size. For insertions, those high-quality ones not on masked regions were kept for analysis. LD between SVs was calculated using PLINK (V1.90 beta) [42].Rare SV burden on H1/H2 region was evaluated by SKAT-O [43] adjusting for gender and PCs 1–5. As SKAT-O does not calculate an odds ratio, we calculated the odds ratio using logistic regression with the same covariates.

## Results

### Common SNVs and indels associated with PSP

We conducted whole genome sequencing at 30 × coverage in 4,662 European-ancestry samples (1,718 individuals with PSP of which 1,441 were autopsy confirmed and 277 were clinically diagnosed and 2,944 control subjects, Table 1). We successfully replicated the association of known loci at *MAPT*, *MOBP* and *STX6* [8, 12, 13] and identified a novel signal in *APOE* with a genome-wide significance of $P < 5 \times 10^{-8}$ (Fig. 1, Fig. S1, Table 2, Table S3). Furthermore, eight loci were of potential interest ($5 \times 10^{-8} < P < 1 \times 10^{-6}$), including two loci reported genome-wide significant (*SLCO1A2* and *DUSP10*) [12, 13] and one locus (*SP1*) reported with a $P$ of $4.1 \times 10^{-7}$ [13], as well as five new loci in *FCHO1*/*MAP1S*, *KIF13A*, *TRIM24*, *ELOVL1* and *TNXB*.

### MAPT, MOBP and STX6

In the *MAPT* region, a multitude of SNVs and indels in high linkage disequilibrium (LD) with the H1/H2 haplotype remains the most significant association with PSP (Fig. S2A). From our analysis, the prominent signal within the *MAPT* region is rs62057121 ($P = 7.45 \times 10^{-78}$, β = -1.32, MAF = 0.15). Fine mapping suggests that rs242561 ($P = 4.49 \times 10^{-74}$, β = -1.23, MAF = 0.16) is likely to be a causal SNV underlying the statistical significance. The SNP rs242561 is located in an enhancer region, containing an antioxidant response element that binds with NRF2/sMAF protein complex. The T allele of rs242561 showed a stronger binding affinity for NRF2/sMAF in ChIP-seq analysis, therefore inducing significantly higher transactivation of the *MAPT* gene [44]. rs242561 and rs62057151 were both in high LD ($r^2 > 0.9$) with H1/H2 (defined by the 238 bp deletion in *MAPT* intron 9) and represented the same association signal as the H1/H2. In previous studies [8, 45], the H1c tagging SNV (rs242557) inside the H1/H2 region was found to be significant when conditioning on H1/H2. We confirmed that rs242557 (G/A) was genome-wide significant after adjusting for H1/H2 ($P = 3.68 \times 10^{-15}$, β = 0.39, MAF = 0.42). In H2 background, only rs242557-G allele is observed, while

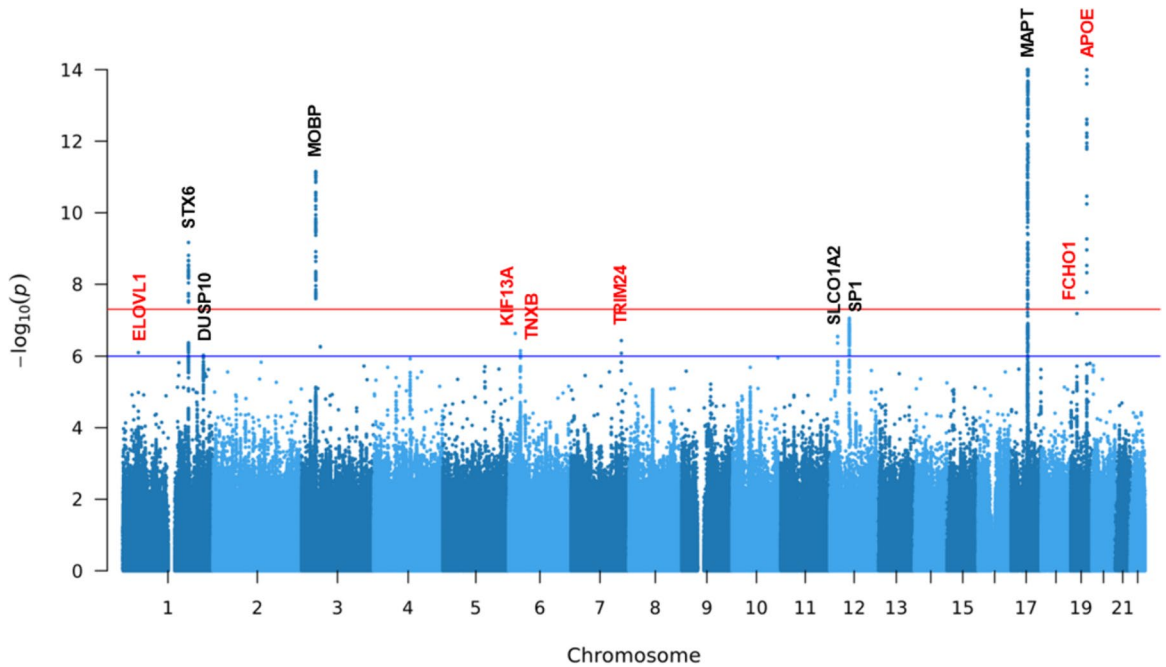


**Fig. 1** Manhattan plot of SNVs/indels for PSP. Loci with a $P < 1 \times 10^{-6}$ are annotated (novel loci in red and known loci in black). Variants with a $P$ value below $1 \times 10^{-14}$ are not shown. The red horizontal line represents genome-wide significance level ($5 \times 10^{-8}$). The blue horizontal line represents loci of potential interest ($1 \times 10^{-6}$)

**Table 2** Top associations from genome-wide association study

| SNV | Chr | Position | Ref | Alt | AF (Alt) | β (Alt) | P | Gene | eQTL/sQTL |
|---|---|---|---|---|---|---|---|---|---|
| **Genome-Wide Significance ($P < 5 \times 10^{-8}$)** | | | | | | | | | |
| rs62057121 | 17 | 45823394 | G | A | 0.15 | -1.32 | $7.45 \times 10^{-78}$ | MAPT | LRRC37A4P[c*] |
| rs4420638 | 19 | 44919689 | A | G | 0.20 | -0.57 | $2.91 \times 10^{-19}$ | APOE | TOMM40[b] |
| rs7412 | 19 | 44908822 | C | T | 0.06 | 0.87 | $9.57 \times 10^{-16}$ | APOE | |
| rs11708828 | 3 | 39458158 | C | T | 0.46 | -0.35 | $7.04 \times 10^{-12}$ | MOBP | PRSA[c] |
| rs10753232 | 1 | 180980990 | C | T | 0.44 | 0.31 | $6.79 \times 10^{-10}$ | STX6 | STX6[a*] |
| **Loci of Potential Interest ($P < 1 \times 10^{-6}$)** | | | | | | | | | |
| rs56251816 | 19 | 17750888 | A | G | 0.22 | 0.35 | $6.57 \times 10^{-08}$ | FCHO1/MAP1S | |
| rs12817984 | 12 | 53410523 | T | G | 0.16 | -0.37 | $8.91 \times 10^{-08}$ | SP1 | SP1[a*] |
| rs4712314 | 6 | 17833813 | G | T | 0.51 | 0.27 | $2.37 \times 10^{-07}$ | KIF13A | |
| rs74651308 | 12 | 21323155 | G | A | 0.07 | 0.51 | $2.86 \times 10^{-07}$ | SLCO1A2 | |
| rs111593852 | 7 | 138449166 | C | T | 0.02 | 0.87 | $3.75 \times 10^{-07}$ | TRIM24 | |
| rs367364 | 6 | 32052169 | C | T | 0.13 | -0.37 | $7.07 \times 10^{-07}$ | TNXB | CYP21A1P[c*] |
| rs839764 | 1 | 43367703 | T | A | 0.41 | 0.27 | $7.94 \times 10^{-07}$ | ELOVL1 | TIE1[a*] |
| rs12026659 | 1 | 221976623 | G | A | 0.21 | 0.31 | $9.48 \times 10^{-07}$ | DUSP10 | |

*Chr* Chromosome, *Ref* Reference allele, *Alt* Alternative allele, *AF* Allele frequency

[*] Represents the SNV regulates multiple genes, and the gene with the smallest *P*-value was shown here (eQTL/sQTL for the brain region was obtained through GTEx)

[a] SNVs with significant eQTL hits

[b] SNVs with significant sQTL hits

[c] SNVs with both eQTL and sQTL hits

both rs242557-G and rs242557-A exist in H1. Therefore, rs242557 is in LD with H1/H2 ($r^2 = 0.14$, D' = 1, $P < 0.0001$). The $r^2$ is relatively low because rs242557-A (AF = 0.42) has a lower allele frequency and could not tag or substitute H1 (AF = 0.81). To pinpoint the causal genes underlying the association in H1/H2 requires additional functional study. For example, Cooper et al. [11] analyzed transcriptional regulatory activity of SNVs and suggested *PLEKHM1* and *KANSL1* were probable causal genes in H1/H2 besides *MAPT*. In *MOBP* (rs11708828, $P = 7.04 \times 10^{-12}$, β = -0.35, MAF = 0.46, Fig. S2B) and *STX6* (rs10753232, $P = 6.79 \times 10^{-10}$, β = 0.31, MAF = 0.44, Fig. S2C), the associated variants were of high allele frequency and exhibited moderate effect size.

### APOE and risk of PSP

One newly identified significant locus from our analysis is the well-known AD risk gene, *APOE*. We observed a significant association between the *APOE* ε2 haplotype and an elevated risk of PSP ($P = 9.57 \times 10^{-16}$, β = 0.87, MAF = 0.06, Table 2, Fig. S3B). The *APOE* ε2 haplotype is encoded by rs429358-T and rs7412-T, which is considered a protective allele in AD. The increased risk of *APOE* ε2 in PSP has been previously reported in a Japanese cohort, albeit with a relatively small sample size [46]. Furthermore, Zhao et al. [47] confirmed that *APOE* ε2 is linked to increased tau pathology in the brains of individuals with PSP and reported a higher frequency of homozygosity of *APOE* ε2 in PSP with an odds ratio of 4.41. Consistent with these findings, our dataset exhibited a higher frequency of homozygosity of rs7412-T in PSP, yielding an odds ratio of 3.91.

For *APOE* ε4 allele, contrary to its association with AD, we observed that rs429358-C exhibits a protective effect against PSP ($P = 5.71 \times 10^{-18}$, β = -0.60, MAF = 0.16, Table 2). The lead SNV demonstrating this protective association from our analysis is rs4420638 ($P = 2.91 \times 10^{-19}$, β = -0.57, MAF = 0.20, Fig. S3A), which is in LD ($r^2 = 0.74$) with rs429358. In a previous PSP GWAS conducted by Hoglinger et al. [8], another *APOE* ε4 tagging SNV (rs2075650, $r^2 = 0.52$ with rs429358) was also found to be diminished (MAF_case = 0.11 and MAF_control = 0.15) in PSP, although not reaching significance ($P = 1.28 \times 10^{-5}$). Notably, in our analysis, rs2075650 reached genome-wide significance ($P = 3.39 \times 10^{-13}$, β = -0.51, MAF = 0.15). *APOE* ε4 or ε2 displayed an independent effect for PSP risk without a significant epistatic interaction with H1/H2 haplotype ($P > 0.05$) (Fig. S4).

Given that our dataset included external controls from ADSP collected for Alzheimer's disease studies, there were potential selection biases for *APOE* ε4 and ε2 in controls. To address this concern, we analyzed the allele frequencies of *APOE* ε4 and ε2 by cohorts (Table S2) and indicated cohorts with potential selection bias. The association analysis excluding these cohorts shows the ε2 SNV (rs7412, $P = 1.23 \times 10^{-12}$, β = 0.70, MAF = 0.06)

remained genome-wide significant and ε4 SNV (rs429358, $P=0.02$, β = -0.16, MAF = 0.14) was nominally significant (Table S4, Table S5). Despite removing ADSP controls with a potential selection bias for *APOE* ε4 and ε2, the allele frequency of *APOE2* is still higher in external databases (AF = 0.0752—0.1060; Table 3) compared to controls in our study (AF = 0.0454; Table S4). This indicates there could still be additional factors affecting the collection of controls in ADSP.

#### Other loci of potential interest

Eight loci were of potential interest ($5\times10^{-8} < P < 1\times10^{-6}$) in our analysis of which three, *SLCO1A2*, *DUSP10*, and *SP1*, were previously reported [12, 13]. In *SLCO1A2*, the lead SNV rs74651308 ($P=2.86\times10^{-7}$, β = 0.51, MAF = 0.07, Fig. S5A) is intronic and in LD ($r^2=0.98$) with missense SNV rs11568563 ($P=1.45\times10^{-6}$, β = 0.47, MAF = 0.07), which was reported in a previous study [12]. About 250 kb upstream of *DUSP10* lies the previously reported SNV rs6687758 [12] ($P=3.36\times10^{-6}$, β = 0.29, MAF = 0.21), which is in LD ($r^2=0.98$) with the lead SNV rs12026659 in our analysis ($P=9.48\times10^{-7}$, β = 0.31, MAF = 0.21, Fig. S5B). In *SP1*, the reported indel rs147124286 [13] ($P=4.39\times10^{-7}$, β = -0.35, MAF = 0.16) is in LD ($r^2=0.995$) with the lead SNV rs12817984 ($P=8.91\times10^{-8}$, β = -0.37, MAF = 0.16, Fig. S5C). Notably, disruption of a transcriptional network centered on *SP1* by causal variants has been implicated previously in PSP [11].

Five newly discovered loci are in *FCHO1/MAP1S, KIF13A, TRIM24, TNXB,* and *ELOVL1*. Within *FCHO1/MAP1S*, the most significant signal (rs56251816, $P=6.57\times10^{-8}$, β = 0.35, MAF = 0.22, Fig. S6A) is in the intron of *FCHO1*. rs56251816 is a significant expression quantitative trait locus (eQTL) for both *FCHO1* and *MAP1S* (13 kb upstream of *FCHO1*) in the Genotype-Tissue Expression (GTEx) project [52]. *MAP1S* encodes a microtubule associated protein that is involved in microtubule bundle formation, aggregation of mitochondria and autophagy [53], and therefore, is more relevant than *FCHO1* regarding PSP. *KIF13A*, which encodes a microtubule-based motor protein was also of potential interest (rs4712314, $P=2.37\times10^{-7}$, β = 0.27, AF = 0.51, Fig. S6B). The significance in genes involved in microtubule-based processes, such as *MAPT*, *MAP1S* and *KIF13A*, implicates the neuronal cytoskeleton as a convergent aspect of PSP etiology.

Besides, *TRIM24* (rs111593852, $P=3.75\times10^{-7}$, β = 0.87, MAF = 0.02, Fig. S7A), *TNXB* (rs367364, $P=7.07\times10^{-7}$, β = -0.37, MAF = 0.13, Fig. S7B) and *ELOVL1* (rs839764, $P=7.94\times10^{-7}$, β = 0.27, MAF = 0.41, Fig. S7C) were also of potential interest. *TRIM24* is involved in transcriptional initiation and shows differential expression in individuals with Parkinson's disease [54, 55]. *TNXB* is located in the major histocompatibility complex (MHC) region on chromosome 6 and makes a matricellular protein called tenascin-X [56].. *ELOVL1* encodes an enzyme that elongates fatty acids and can cause a neurological disorder with ichthyotic keratoderma, spasticity, hypomyelination and dysmorphic features [57]. Furthermore, we found a few SNV/indels with $P<1\times10^{-6}$ but without other supporting variants in LD (Fig.S1, Table S3). These signals could be due to sequencing errors and need further experimental validation.

#### Rare SNVs/indels and network analysis

The heritability of PSP for common SNVs and indels (MAF > 0.01) was estimated to be 20%, while common plus rare SNVs/indels was estimated to be 23% from our analysis using GCTA-LDMS [22]. Therefore, we performed aggregated tests for rare SNVs and indels, and identified *ZNF592* (SKAT-O FDR = 0.043, burden test

**Table 3** Allele Frequency of APOE ε4 SNV (rs429358) and ε2 SNV (rs7412)

| Studies | rs429358 | | rs7412 | |
|---|---|---|---|---|
| | AF (Case) | AF (Control) | AF (Case) | AF (Control) |
| PSP WGS (This study) | 0.1279 | 0.1742 | 0.0844 | 0.0414 |
| PSP GWAS [48] | 0.1159 | 0.1366 | 0.0826 | 0.0794 |
| 1000 Genomes Project [18] | | 0.1512 | | 0.0771 |
| ExAC<br>European (non-Finnish) [49] | | 0.2078 | | 0.1060 |
| gnomAD V4<br>European (non-Finnish) [50] | | 0.1506 | | 0.0783 |
| TOPMed Freeze 8<br>NFE (Non-Finnish European) | | 0.1501 | | 0.0752 |
| ADSP R3 Non-Hispanic White [51] | 0.3139<br>(AD as cases) | 0.1803 | 0.0244<br>(AD as cases) | 0.0406 |

FDR = 0.041) with an of OR = 1.08 (95% CI: 1.008–1.16) (Fig. 2, Table 4, Table S7) for protein truncating or damaging missense variants. There was no genomic inflation with a λ = 1.07 (Fig. 2). Risk in *ZNF592* was imparted by 16 unique variants, with one splice donor and 15 damaging missense variants (Table S7). *ZNF592* has not been previously associated with PSP but showed moderate RNA expression in the cerebellum compared to other tissues from GTEx (Fig. S8). There were no significant genes identified when evaluating PTVs only or when restricting to loss of function intolerant genes.

Considering that genes do not operate alone, but rather within signaling pathways and networks, we and others have shown that better understanding of disease mechanisms can be achieved through gene network analysis [58–60]. Therefore, we scrutinized rare variants within a network framework, focusing on co-expression network analysis performed in PSP post mortem brain that had previously identified a brain co-expression module, C1, which was conserved at the protein interaction level and enriched for common variants in PSP [32]. We found this C1 neuronal module was significantly enriched with PSP rare variants ($P = 0.006$, OR [95% CI] = 1.31 [1.01–1.70], Table 4; Table S8). Genes from the C1 module were more likely to be loss of function intolerant compared to the background of all brain expressed genes (Fig. S8). To ensure that this was association not spurious, we performed permutation testing using random gene modules of brain expressed genes with the same number of genes as C1. The C1 module remains significant (Permutation $P = 0.078$). Exploring GTEx, we found that C1 genes are highly expressed in brain tissues including the cerebellum, frontal cortex, and basal ganglia (Fig. S8), consistent with regions affected in this disorder.

**SVs associated with PSP**

Seven high-confident SVs achieved genome-wide significance with PSP (Table 5, Fig. S9), including three deletions tagging the H2 haplotype. The most significant signal is a 238 bp deletion in *MAPT* intron 9 (Fig. S10A, chr17:46009357–46009595, $P = 3.14 \times 10^{-50}$, AF = 0.16) that has been reported on the H2 haplotype [61, 62] and is in LD ($r^2 = 0.99$) with the lead SNV, rs62057121 (chr17:45823394, $P = 7.45 \times 10^{-78}$, β = -1.32, MAF = 0.15), in the *MAPT* region. Adding to this, two other deletions, one spanning 314 bp (Fig. S10B, chr17:46146541–46146,855, AF = 0.19) and the other covering 323 bp (Fig. S10C, chr17:46099028–46099351, AF = 0.22), both are Alu elements and in LD ($r^2 > 0.8$) with the top signal (the 238 bp deletion). This observation indicates that transposable elements may play an important role in the evolution of H1/H2 haplotype structure.

Beyond the identified SVs in the H1/H2 region, we uncovered a significant deletion (chr14:105864208–105916743, $P = 4.74 \times 10^{-14}$, AF = 0.01) within the immunoglobulin heavy locus (*IGH*), which is a complex SV region (Fig. S11) related to antigen recognition. Moreover, a 619 bp deletion (chr6:149762615–149763234, $P = 8.60 \times 10^{-12}$, AF = 0.55; Fig. S10D) in *PCMT1* displayed increased risk of PSP with an odds ratio of 4.19. The odds ratio increased to 8.38 when comparing 1,244 individuals with homozygous deletions in *PCMT1* with the rest of sample set. *PCMT1* encodes a type II class of protein carboxyl methyltransferase enzyme that is highly expressed in the brain [63] and is able to ameliorate $A\beta_{25\text{-}35}$ induced neuronal apoptosis [64, 65]. Additionally, we found a deletion between *CYP2F1* and *CYP2A13* (chr19:41102802–41104285, AF = 0.17) and an insertion in *SMCP* (chr1:152880979–152880979, AF = 0.74)

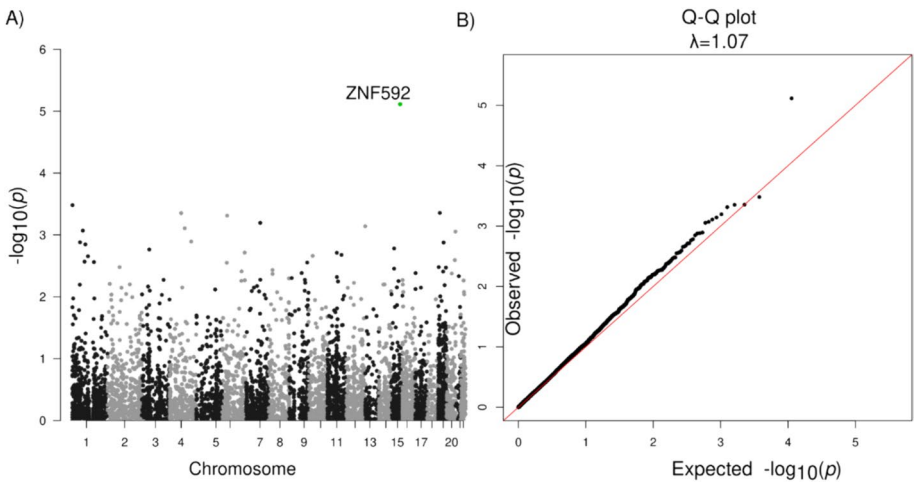


**Fig. 2** Association analysis of rare SNVs/indels. **A** Manhattan plot for genes with protein truncating variants or damaging missense variants. **B** Q-Q plot of gene *P*-values with protein truncating variants or damaging missense variants

**Table 4** Association analysis of ZNF592 and the C1 module

| Gene | Variants | Total MAC | Case MAC | Control MAC | Fraction Case | Fraction Control | OR (95% CI) | SKAT-O | | Burden | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | FDR | P | FDR | P |
| ZNF592 | 16 | 19 | 8 | 11 | 0.0023 | 0.0018 | 1.08 (1.01–1.16) | 0.044 | $7.60\times10^{-6}$ | 0.041 | $7.30\times10^{-06}$ |
| **Module** | **Variants** | **Total MAC** | **Case MAC** | **Control MAC** | **Fraction Case** | **Fraction Control** | **OR (95% CI)** | **Permutation test** | **P** | **Permutation test** | **P** |
| C1 | 180 | 234 | 101 | 133 | 0.029 | 0.022 | 1.31 (1.01–1.70) | 0.19 | 0.048 | 0.078 | 0.006 |

**Table 5** Significant structural variants from association analysis ($P < 5 \times 10^{-8}$)

| Name | N | AF | beta | P | AF (case) | AF (control) | Odds Ratio | Fisher's P | Gene |
|---|---|---|---|---|---|---|---|---|---|
| chr17:46009357–46009595:DEL[a] | 4357 | 0.16 | -1.22 | $3.14 \times 10^{-50}$ | 0.054 | 0.23 | 0.19 | $5.80 \times 10^{-118}$ | MAPT |
| chr17:46146541–46146855:DEL[a] | 3697 | 0.19 | -1.12 | $2.13 \times 10^{-39}$ | 0.079 | 0.25 | 0.26 | $1.58 \times 10^{-83}$ | KANSL1 |
| chr17:46099028–46099351:DEL[a] | 3699 | 0.22 | -1.07 | $3.88 \times 10^{-37}$ | 0.11 | 0.28 | 0.33 | $2.05 \times 10^{-66}$ | KANSL1 |
| chr14:105864208–105916743:DEL | 4378 | 0.010 | -1.53 | $4.74 \times 10^{-14}$ | 0.0053 | 0.014 | 0.39 | $1.33 \times 10^{-04}$ | IGH |
| chr6:149762615–149763234:DEL | 3811 | 0.55 | 0.50 | $8.60 \times 10^{-12}$ | 0.75 | 0.42 | 4.19 | $6.00 \times 10^{-182}$ | PCMT1 |
| chr19:41102802–41104285:DEL | 2921 | 0.17 | 0.64 | $7.46 \times 10^{-09}$ | 0.21 | 0.14 | 1.59 | $5.95 \times 10^{-11}$ | CYP2A13 |
| chr1:152880979–152880979:INS | 2872 | 0.74 | 0.67 | $2.37 \times 10^{-08}$ | 0.79 | 0.71 | 1.62 | $1.46 \times 10^{-13}$ | SMCP |

[a] Represents SVs with DNA samples available and PCR validated

which were also significant (Table 5). The 1.5 kb deletion (chr19:41102802–41104285) almost completely overlaps the SINE-VNTR-Alus (SVA) transposon region annotated by RepeatMasker [66].

### SVs in H1/H2 haplotype region

The H1/H2 region stands out as the pivotal genetic risk factor for PSP [8, 67]. The H2 haplotype exhibits a reduced odds ratio of 0.19, as we observed the allele frequency of the 238 bp H2-tagging deletion is 23% in PSP and only 5% in control ($P < 2.2 \times 10^{-16}$). Moreover, our analysis pointed out five common (MAF > 0.01) and 12 rare deletions and duplications in the region (Table 6), ranging from 88 bp to 47 kb. Additionally, one common and four rare high-confidence insertions were reported in the region.

**Table 6** High-confident structural variants in the H1/H2 haplotype region

| Name | Size | N | AF | AF (PSP) | AF (Control) | Gene | Annotation |
|---|---|---|---|---|---|---|---|
| chr17:46099028–46099351:DEL[a]* | 323 | 3,699 | 0.24 | 0.11 | 0.28 | KANSL1 | intron |
| chr17:46146541–46146855:DEL[a]* | 314 | 3,697 | 0.21 | 0.08 | 0.25 | KANSL1 | intron |
| chr17:46237619–46238142:DEL[a] | 523 | 3,686 | 0.19 | 0.09 | 0.22 | MAPK8IP1P1 | intergenic |
| chr17:46009357–46009595:DEL[a]* | 238 | 4,357 | 0.19 | 0.05 | 0.23 | MAPT | intron |
| chr17:46277789–46282210:DEL | 4,421 | 4,233 | 0.12 | 0.03 | 0.15 | ARL17B | intron |
| chr17:46113802–46113802:INS | 311 | 2,464 | 0.31 | 0.32 | 0.32 | KANSL1 | intron |
| Name | Size | N | N (Carriers) | N (PSP) | N (Control) | Gene | Annotation |
| chr17:46811121–46811289:DEL[a] | 168 | 2,614 | 36 | 15 | 21 | WNT3 | intron |
| chr17:45847702–45851880:DEL[a] | 4,178 | 4,427 | 31 | 17 | 14 | MAPT-AS1 | splicing |
| chr17:46837153–46839088:DEL[a] | 1,935 | 4,415 | 12 | 8 | 4 | WNT9B | intron |
| chr17:45918825–45920861:DEL[a] | 2,036 | 4,422 | 1 | 0 | 1 | MAPT | intron |
| chr17:45916681–45920693:DEL | 4,012 | 4,430 | 3 | 0 | 3 | MAPT | intron |
| chr17:45570198–45572012:DEL | 1,814 | 4,243 | 3 | 2 | 1 | AC091132.4 | intron |
| chr17:45334194–45381549:DEL[a] | 47,355 | 4,430 | 1 | 0 | 1 | AC003070.2 | transcript ablation |
| chr17:45311955–45312258:DEL | 303 | 4,365 | 2 | 0 | 2 | MAP3K14 | intron |
| chr17:45894637–45914976:DUP[a] | 20,339 | 4,260 | 1 | 1 | 0 | MAPT-AS1 | transcript amplification |
| chr17:45993882–45993970:DEL[a] | 88 | 4,283 | 1 | 1 | 0 | MAPT | splicing |
| chr17:45665996–45666370:DEL[a] | 374 | 4,412 | 1 | 1 | 0 | LINC02210-CRHR1 | TFBS ablation |
| chr17:45879141–45881180:DEL | 2,039 | 4,431 | 1 | 1 | 0 | MAPT-AS1 | intron |
| chr17:45741582–45741582:INS | 315 | 4,420 | 10 | 4 | 6 | LINC02210-CRHR1 | intergenic |
| chr17:45929579–45929579:INS | 453 | 3,025 | 5 | 1 | 4 | MAPT | intron |
| chr17:46754483–46754483:INS | 330 | 3,692 | 12 | 2 | 10 | NSF | intron |

*AF* Allele frequency, *N* Number of individuals with non-missing genotypes

[*] High-quality SVs that were included in association analysis

[a] Represents SVs with DNA samples available and PCR validated

Of the five common deletions and duplications (Fig. S12), three show genome-wide significant association with the disease (Table 5); four are located in regions with transposable elements (SVA, L1, or Alu) and in LD ($r^2$ from 0.63 to 0.92) with the 238 bp H2-tagging deletion. This further highlights the important role of transposable elements in shaping the landscape of H1/H2 region.

Among the 12 rare deletions and duplications (Fig. S13), five are located in potentially functional regions, such as splice sites, exons, and transcription factor binding sites (Table 6). Particularly, one deletion (chr17:45993882–45993970) in exon 9 of *MAPT* was identified in a PSP patient, adding to previous reports of exonic deletions in the *MAPT* in frontotemporal dementia, such as deletion of exon 10 [68] and exons 6–9 [69] in *MAPT*. Using the SKAT-O test ($N=4{,}432$), the 12 rare CNVs displayed a significantly higher burden in PSP than controls ($P=0.01$, OR = 1.64).

## Discussion

Through comprehensive analysis of whole genome sequence, we identified SNVs, indels and SVs contributing to the risk of PSP. For common SNVs, previously reported regions, including *MAPT*, *MOBP*, STX6, *SLCO1A2*, *DUSP10*, and *SP1* [8, 12, 13] were replicated in our analysis and novel loci in *APOE*, *FCHO1*/*MAP1S*, *KIF13A*, *TRIM24*, *ELOVL1*, and *TNXB* were discovered. *EIF2AK3* which was significantly associated with PSP in a previous GWAS [8] did not reach significance in our study. In the current study, the SNV with the lowest *P* around *EIF2AK3* was rs13003510 ($P=8.30\times10^{-5}$, $\beta=0.22$, MAF = 0.3).

The *APOE* loci was of particular interest as it is a common risk factor for AD, explaining more than a 1/3 of population attributable risk [70, 71]. In contrast to AD, the ε4 tagging allele rs429358 was protective in PSP and the ε2 tagging allele rs7412 was deleterious. After removing ADSP controls with a potential selection bias for *APOE* ε4 and ε2, ε2 remained genome-wide significant and ε4 showed nominal significance. This observation is particularly intriguing since both AD and PSP have intracellular aggregated tau as a prominent neuropathologic feature. Notably, both ε2 allele and ε4 allele have been associated with tau pathology burden in the brain of mice models [47, 72], which raises the question of distinct tau species in 4R-PSP versus 3R-4R-AD. It is also notable that the ε2 allele is also associated with increased risk for age-related macular degeneration (AMD), and the ε4 allele was associated with decreased risk [73, 74]. These results demonstrate that the same variant may have opposite effects in different degenerative diseases. This is especially important, given the advent of gene editing as a therapeutic modality, and programs focused on changing *APOE* ε4 to ε2. Although this therapy would likely decrease risk for AD, our results indicate that it could increase risk for PSP, in addition to AMD. From this standpoint, caution is warranted in germ-line genome editing until the broad spectrum of phenotypes associated with human genetic variation is understood.

Moreover, MHC region, which encodes genes that present antigens and are involved in synaptic plasticity, axonal regeneration and neuroinflammation [75, 76], belonged to one of loci of interest in our association analysis. We found that the most significant SNV in this region (rs367364) was in LD with a few HLA alleles (i.e., HLA-A*01:01, HLA-B*08:01, HLA-C*07:01, HLA-DQB1*02:01 and HLA-DRB1*03:01; Table S9), though only HLA-C*07:01 showed nominal significance in association with PSP ($P=9.19\times10^{-3}$; Table S10). This suggests that the effect of rs367364 on PSP could not be explained by individual HLA alleles. Since HLA-A*01:01-B*08:01-C*07:01-DQB1*02:01- DRB1*03:01 is the most common HLA-A-B-C-DQB1-DRB1 haplotype in Europeans (AF = 0.074) [77], comprehensive analysis of the HLA haplotypes and their contribution to the risk of PSP is needed in the future.

Burden association tests are an highly valuable for addressing sample size limitations in analyzing rare variants [78]. Indeed, burden testing allowed us to identify *ZNF592*, a classical C2H2 zinc finger protein (ZNF) [79, 80], as a candidate risk gene. ZNF proteins have been causative or strongly associated with large numbers of neurodevelopmental disease [81, 82] and neurodegenerative disease including Parkinson's disease [83] and Alzheimer's disease [84, 85]. *ZNF592* was initially thought to be responsible for autosomal recessive spinocerebellar ataxia 5 from a consanguineous family with neurodevelopmental delay including cerebellar ataxia and intellectual disability due to a homozygous G1046R substitution [86]. However, further analysis of this family identified *WDR73* to be the most likely causative gene, consistent with Galloway-Mowat syndrome, although *ZNF592* may have contributed to the phenotype [87].

We also extended classical gene-based burden analysis to consider rare risk burden in the context of a gene set defined by co-expression networks [32, 88]. We leveraged combined previous proteomic and transcriptomic analysis of post-mortem brain from patients afflicted with PSP, and showed that rare variants enrich in the C1 neuronal module, which was the same module enriched with common variants [32]. This, along with our recent work identifying a neuronally-enriched transcription factor network centered around SP1 disrupted by PSP common genetic risk, suggests that although PSP neuropathologically is defined by tufted

astrocytes and oligodendroglial coiled bodies [6, 89, 90], initial causal drivers of PSP appear to be primarily neuronal.

In analysis of SVs, we found deletions in *PCMT1* and *IGH* were significantly associated with PSP. The *IGH* deletions are in a complex region on chromosome 14 that encodes immunoglobins recognizing foreign antigens. The size of the *IGH* deletion varies across individuals (Fig. S9). In addition, the *IGH* deletions can be accompanied by other deletions, duplications, and inversions (Fig. S9). These combined make the experimental validation of the deletion challenging. The *PCMT1* deletion is common (AF = 0.55) with an odds ratio of 8.38 for PSP in homozygous individuals.

There were limitations to this study. First, not all PSP were pathologically confirmed (of the 1,718 PSP individuals, 1,441 were autopsy-confirmed and 277 were clinically-diagnosed). The specificity of the National Institute of Neurological Disorders and Stroke and Society for PSP (NINDS-SPSP) from 1997 [91] was shown to be 95% to 100% for probable PSP and around 80% to 93% for possible PSP [92–94]. The 2017 Movement Disorder Society PSP (MDS-PSP) clinical criteria were developed to improve the sensitivity for PSP patients with variant syndrome that were not reflective of PSP-Richardson Syndrome [3]. The MDS criteria also have shown a small decrease in specificity but improved sensitivity in clinicopathological studies [95, 96]. Additionally, the majority of control samples in this study were from ADSP and were initially collected as controls for AD studies. Although samples with a potential selection bias for *APOE* ε4 and ε2 were removed, the allele frequency of *APOE* ε2 in controls was still lower compared to external databases (Tables 3 and S4), indicating that there could be additional factors affecting the collection of controls in ADSP. For example, if individuals had an AD family history, they might be more willing to volunteer to serve as controls in ADSP therefore contributing to the lower allele frequency of *APOE2*. To clarify this, future replication studies using independent datasets are needed to validate the effects of *APOE* ε4 and ε2 in PSP.

This work represents an important first step; future work is necessary to further delineate the rare genetic risk in PSP harbored in coding and noncoding regions. These results may come to fruition as additional genomic analytical methods are developed, sample size increased, and orthogonal genomic data are integrated. While PSP is rare, it is the most common primary tauopathy, and studying this disease is critical to understanding common pathological mechanisms across tauopathies. Further work to include individuals with diverse ancestry background will also improve our understanding of genetic architecture of the disease.

## Conclusion

In conclusion, this study significantly advances our understanding of the genetic basis of PSP through WGS from this study. Previous GWAS signals were validated, and *APOE2* was found to the risk allele for PSP from the analysis of common SNVs and indels. Additionally, the analysis of rare SNVs/indels and SVs has revealed additional genetic targets, including *ZNF592*, *IGH*, *PCMT1*, *CYP2A13*, and *SMCP*, opening new avenues for investigating disease mechanisms and potential therapeutic interventions.

## Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13024-024-00747-3.

Supplementary Material 1.

Supplementary Material 2.

**Acknowledgements**

This project is supported by CurePSP, courtesy of a donation from the Morton and Marcine Friedman Foundation. We are indebted to the Biobanc-Hospital Clinic-FRCB-IDIBAPS and Center for Neurodegenerative Disease Research at Penn for samples and data procurement. The PSP genetics study group is a multisite collaboration including: German Center for Neurodegenerative Diseases (DZNE), Munich; Department of Neurology, LMU Hospital, Ludwig-Maximilians-Universität (LMU), Munich, Germany (Franziska Hopfner, Günter Höglinger); German Center for Neurodegenerative Diseases (DZNE), Munich; Center for Neuropathology and Prion Research, LMU Hospital, Ludwig-Maximilians-Universität (LMU), Munich, Germany (Sigrun Roeber, Jochen Herms); Justus-Liebig-Universität Gießen, Germany (Ulrich Müller); MRC Centre for Neurodegeneration Research, King's College London, London, UK (Claire Troakes); Movement Disorders Unit, Neurology Department and Neurological Tissue Bank and Neurology Department, Hospital Clínic de Barcelona, University of Barcelona, Barcelona, Catalonia, Spain (Ellen Gelpi; Yaroslau Compta); Department of Neurology and Netherlands Brain Bank, Erasmus Medical Centre, Rotterdam, The Netherlands (John C. van Swieten); Division of Neurology, Royal University Hospital, University of Saskatchewan, Canada (Alex Rajput); Australian Brain Bank Network in collaboration with the Victorian Brain Bank Network, Australia (Fairlie Hinton), Department of Neurology, Hospital Ramón y Cajal, Madrid, Spain (Justo García de Yebenes). The acknowledgement of PSP cohorts is listed below, whereas the acknowledgement of ADSP cohorts for control samples can be found in the supplementary materials. The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. The data used for the analyses described in this manuscript were obtained from: https://gtexportal.org/home/datasets the GTEx Portal on 1/27/2022. We also thank to Drs. Murray Grossman and Hans Kretzschmar for their valuable contribution to this work.

AMP-AD (sa000011) data: Mayo RNAseq Study- Study data were provided by the following sources: The Mayo Clinic Alzheimer's Disease Genetic Studies, led by Dr. Nilufer Ertekin-Taner and Dr. Steven G. Younkin, Mayo Clinic, Jacksonville, FL using samples from the Mayo Clinic Study of Aging, the Mayo Clinic Alzheimer's Disease Research Center, and the Mayo Clinic Brain Bank. Data collection was supported through funding by NIA grants P50 AG016574, R01 AG032990, U01 AG046139, R01 AG018023, U01 AG006576, U01 AG006786, R01 AG025711, R01 AG017216, R01 AG003949, NINDS grant R01 NS080820, CurePSP Foundation, and support from Mayo Foundation. Study data includes samples collected through the Sun Health Research Institute Brain and Body Donation Program of Sun City, Arizona. 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 Arizona Alzheimer's Disease Core 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.
PSP-NIH-CurePSP-Tau (sa000015) data: This project was funded by the NIH grant UG3NS104095 and supported by grants U54NS100693 and U54AG052427. Queen Square Brain Bank is supported by the Reta Lila Weston Institute for Neurological Studies and the Medical Research Council UK. The Mayo Clinic Florida had support from a Morris K. Udall Parkinson's Disease Research Center of Excellence (NINDS P50 #NS072187), CurePSP and the Tau Consortium. The samples from the University of Pennsylvania are supported by NIA grant P01AG017586.
PSP-CurePSP-Tau (sa000016) data: This project was funded by the Tau Consortium, Rainwater Charitable Foundation, and CurePSP. It was also supported by NINDS grant U54NS100693 and NIA grants U54NS100693 and U54AG052427. Queen Square Brain Bank is supported by the Reta Lila Weston Institute for Neurological Studies and the Medical Research Council UK. The Mayo Clinic Florida had support from a Morris K. Udall Parkinson's Disease Research Center of Excellence (NINDS P50 #NS072187), CurePSP and the Tau Consortium. The samples from the University of Pennsylvania are supported by NIA grant P01AG017586. Tissues were received from the Victorian Brain Bank, supported by The Florey Institute of Neuroscience and Mental Health, The Alfred and the Victorian Forensic Institute of Medicine and funded in part by Parkinson's Victoria and MND Victoria. We are grateful to the Sun Health Research Institute Brain and Body Donation Program of Sun City, Arizona for the provision of human biological materials (or specific description, e.g. brain tissue, cerebrospinal fluid). 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 Arizona Alzheimer's Disease Core 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. Biomaterial was provided by the Study Group DESCRIBE of theClinical Research of the German Center for Neurodegenerative Diseases (DZNE).
PSP_UCLA (sa000017) data: Thank to the AL-108-231 investigators, Adam L Boxer, Anthony E Lang, Murray Grossman, David S Knopman, Bruce L Miller, Lon S Schneider, Rachelle S Doody, Andrew Lees, Lawrence I Golbe, David R Williams, Jean-Cristophe Corvol, Albert Ludolph, David Burn, Stefan Lorenzl, Irene Litvan, Erik D Roberson, Günter U Höglinger, Mary Koestler, Cliff ord R Jack Jr, Viviana Van Deerlin, Christopher Randolph, Iryna V Lobach, Hilary W Heuer, Illana Gozes, Lesley Parker, Steve Whitaker, Joe Hirman, Alistair J Stewart, Michael Gold, and Bruce H Morimoto.

### Authors' contributions

Study design: TSC, DD, GUH, GDS, DHG, and WPL. Sample collection, brain biospecimens, and neuropathological examinations: TSC, CM, LM, AR, PPDD, NLB, MG, LDK, JCVS, ED, BFG, KLN, CT, JGdY, ARG, TM, WHO, GR, UM, FH, TA, SR, PP, AB, AD, ILB, TGC, GES, LNH, IL, RR, OR, DG, ALB, BLM, WWS, VMVD, EBL, CLW, HM, JH, RdS, JFC, AMG, GC, and DHG. Genotype or phenotype acquisition: HW, TSC, VP, LVB, KF, AN, LSW, GDS, DHG, and WPL. Variant detection and variant quality check: HW, TSC, VP, LVB, KF, YYL, and WPL. Statistical analyses and interpretation of results: HW, TSC, KF, AN, GDS, DHG, and WPL. Experimental validation: BAD and PLC. Draft of the manuscript: HW, TSC, GDS, DHG, and WPL. All authors read, critically revised, and approved the manuscript.

### Funding

This work was supported by NIH 5UG3NS104095, the Rainwater Charitable Foundation, and CurePSP. HW and PLC are supported by RF1-AG074328, P30-AG072979, U54-AG052427 and U24-AG041689. TSC is supported by NIH K08AG065519 and the Larry L Hillblom Foundation 2021-A-005-SUP. KF was supported by CurePSP 685–2023-06-Pathway and K01 AG070326. MG is supported by P30 AG066511. BFG and KLN are supported by P30 AG072976 and R01 AG080001. TGB and GES are supported by P30AG072980. IR is supported by 2R01AG038791-06A, U01NS100610, R25NS098999, U19 AG063911-1 and 1R21NS114764-01A1. OR is support by U54 NS100693. DG is supported by P30AG062429. ALB is supported by U19AG063911, R01AG073482, R01AG038791, and R01AG071756. BLM is supported by P01 AG019724, R01 AG057234 and P0544014. VMV is supported by P01-AG-066597, P01-AG-017586. HRM is supported by CurePSP, PSPA, MRC, and Michael J Fox Foundation. RDS is supported by CurePSP, PSPA, and Reta Lila Weston Trust. JFC is supported by R01 AG054008, R01 NS095252, R01 AG060961, R01 NS086736, R01 AG062348, P30 AG066514, the Rainwater Charitable Foundation / Tau Consortium, Karen Strauss Cook Research, and Scholar Award, Stuart Katz & Dr. Jane Martin. AMG is supported by the Tau Consortium and U54-NS123746. YYL is supported by U54-AG052427; U24-AG041689. LSW is supported by U01AG032984, U54AG052427, and U24AG041689. GUH was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy within the framework of the Munich Cluster for Systems Neurology (EXC 2145 SyNergy – ID 390857198); Deutsche Forschungsgemeinschaft (DFG, HO2402/18–1 MSAomics); German Federal Ministry of Education and Research (BMBF, 01KU1403A EpiPD; 01EK1605A HitTau; 01DH18025 TauTherapy). DHG is supported by 3UH3NS104095, Tau Consortium. WPL is supported by RF1-AG074328; P30-AG072979; U54-AG052427; U24-AG041689. Cases from Banner Sun Health Research Institute were supported by the NIH (U24 NS072026, P30 AG19610 and P30AG072980), 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 Mayo Clinic Brain Bank is supported through funding by NIA grants P50 AG016574, CurePSP Foundation, and support from Mayo Foundation.

### Availability of data and materials

All raw data (CRAM files and FASTQ files), SNVs/SVs calls (VCF files) and phenotype files are available through NIAGADS Data Sharing Service (https://dss.niagads.org/) under accession number NG00067.
Code for analyses have been deposited at: https://github.com/whtop/PSP-Whole-Genome-Sequencing-Analysis.

## Declarations

### Consent for publication

Not applicable.

### Competing interests

Laura Molina-Porcel received income from Biogen as a consultant in 2022. Gesine Respondek is now employed by Roche (Hoffmann-La Roche, Basel, Switzerland) since 2021. Her affiliation whilst completing her contribution to this manuscript was German Center for Neurodegenerative Diseases (DZNE), Munich, Germany. Thomas G Beach is a consultant for Aprinoia Therapeutics and a Scientific Advisor and stock option holder for Vivid Genomics. Huw Morris is employed by UCL. In the last 12 months he reports paid consultancy from Roche, Aprinoia, AI Therapeutics and Amylyx; lecture fees/honoraria—BMJ, Kyowa Kirin, Movement Disorders Society. Huw Morris is a co-applicant on a patent application related to C9ORF72—Method for diagnosing a neurodegenerative disease (PCT/GB2012/052140). Giovanni Coppola is currently an employee of Regeneron Pharmaceuticals. Alison Goate serves on the SAB for Genentech and Muna Therapeutics.

### Author details

[1]Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. [2]Penn Neurodegeneration Genomics Center, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. [3]Movement Disorders Programs, Department of Neurology, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA. [4]Department of Pathology, Department of Artificial Intelligence & Human Health, Nash Family, Department of Neuroscience, Ronald M. Loeb Center for Alzheimer's Disease, Friedman Brain, Institute, Neuropathology Brain Bank & Research CoRE, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [5]Victorian Brain Bank, The Florey Institute of Neuroscience and Mental Health, Parkville, VIC, Australia. [6]Alzheimer's Disease and Other Cognitive Disorders Unit. Neurology Service, Hospital Clínic, Fundació Recerca Clínic Barcelona (FRCB). Institut d'Investigacions Biomediques August Pi I Sunyer (IDIBAPS), University of Barcelona, Barcelona, Spain. [7]Neurological Tissue Bank of the Biobanc-Hospital Clínic-IDIBAPS, Barcelona, Spain. [8]Movement Disorders Program, Division of Neurology, University of Saskatchewan, Saskatoon, SK, Canada. [9]Laboratory

of Neurochemistry and Behavior, Experimental Neurobiology Unit, University of Antwerp, Wilrijk (Antwerp), Belgium. [10]Department of Neurology, University Medical Center Groningen, NL-9713 AV Groningen, Netherlands. [11]Fujirebio Europe NV, Technologiepark 6, 9052 Ghent, Belgium. [12]Department of Pathology and Laboratory Medicine and Department of Neurology, Emory University School of Medicine, Atlanta, GA, USA. [13]Netherlands Brain Bank and Erasmus University, Rotterdam, Netherlands. [14]Department of Pathology and Laboratory Medicine, Indiana University School of Medicine, Indianapolis, IN, USA. [15]London Neurodegenerative Diseases Brain Bank, King's College London, London, UK. [16]Autonomous University of Madrid, Madrid, Spain. [17]Fundación CIEN (Centro de Investigación de Enfermedades Neurológicas) - Centro Alzheimer Fundación Reina Sofía, Madrid, Spain. [18]Department of Neurology, Philipps-Universität, Marburg, Germany. [19]German Center for Neurodegenerative Diseases (DZNE), Munich, Germany. [20]Parkinson's Disease and Movement Disorders Department, HYGEIA Hospital, Athens, Greece. [21]European University of Cyprus, Nicosia, Cyprus. [22]Department of Psychiatry and Psychotherapy, University Hospital Munich, Ludwig-Maximilians-University Munich, Munich, Germany. [23]Center for Neuropathology and Prion Research, Ludwig-Maximilians-University Munich, Munich, Germany. [24]German Brain Bank, Neurobiobank Munich, Munich, Germany. [25]Unit of Neurodegenerative Diseases, Department of Neurology, University Hospital Germans Trias I Pujol, Badalona, Barcelona, Spain. [26]Neurosciences, The Germans Trias I Pujol Research Institute (IGTP) Badalona, Badalona, Spain. [27]Sorbonne Université, Paris Brain Institute – Institut du Cerveau – ICM, Inserm U1127, CNRS UMR 7225, APHP - Hôpital Pitié-Salpêtrière, Paris, France. [28]Banner Sun Health Research Institute, Sun City, AZ, USA. [29]University McGill, Montreal, QC, Canada. [30]Department of Neuroscience, University of California, San Diego, CA, USA. [31]VIB Center for Molecular Neurology, University of Antwerp, Antwerp, Belgium. [32]Department of Neuroscience, Mayo Clinic Jacksonville, Jacksonville, FL, USA. [33]Memory and Aging Center, University of California, San Francisco, CA, USA. [34]Center for Neurodegenerative Disease Research, University of Pennsylvania School of Medicine, Philadelphia, PA, USA. [35]University of Texas Southwestern Medical Center, Dallas, TX, USA. [36]Departmento of Clinical and Movement Neuroscience, University College of London, London, UK. [37]Reta Lila Weston Institute, UCL Queen Square Institute of Neurology, London, UK. [38]Department of Genetics and Genomic Sciences, New York, NY, USA; Icahn School of Medicine at Mount Sinai, New York, NY, USA. [39]Friedman Bioventure, Inc., Del Mar, CA, USA. [40]Department of Psychiatry, Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles, CA, USA. [41]Department of Biostatistics, Epidemiology, and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. [42]Department of Anatomy Physiology and Genetics, the American Genome Center, Uniformed Services University of the Health Sciences, Bethesda, MD, USA. [43]Department of Neurology, LMU University Hospital, Ludwig-Maximilians-Universität (LMU) München; German Center for Neurodegenerative Diseases (DZNE), Munich, Germany; and Munich Cluster for Systems Neurology (SyNergy), Munich, Germany. [44]Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA. [45]Institute of Precision Health, University of California, Los Angeles, Los Angeles, CA, USA.

Received: 29 January 2024 Accepted: 22 July 2024
Published online: 16 August 2024

## References

1. Hauw JJ, Daniel SE, Dickson D, Horoupian DS, Jellinger K, Lantos PL, et al. Preliminary NINDS neuropathologic criteria for Steele-Richardson-Olszewski syndrome (progressive supranuclear palsy). Neurology. 1994;44(11):2015–2015.
2. Stamelou M, Respondek G, Giagkou N, Whitwell JL, Kovacs GG, Höglinger GU. Evolving concepts in progressive supranuclear palsy and other 4-repeat tauopathies. Nat Rev Neurol. 2021;17(10):601–20.
3. Hoglinger GU, Respondek G, Stamelou M, Kurz C, Josephs KA, Lang AE, et al. Clinical Diagnosis of Progressive Supranuclear Palsy: The Movement Disorder Society Criteria. Mov Disord Off J Mov Disord Soc. 2017;32(6):853–64.
4. Lukic MJ, Respondek G, Kurz C, Compta Y, Gelpi E, Ferguson LW, et al. Long-Duration Progressive Supranuclear Palsy: Clinical Course and Pathological Underpinnings. Ann Neurol. 2022;92(4):637–49.
5. Ali F, Martin PR, Botha H, Ahlskog JE, Bower JH, Masumoto JY, et al. Sensitivity and specificity of diagnostic criteria for progressive supranuclear palsy. Mov Disord. 2019;34(8):1144–53.
6. Kovacs GG, Lukic MJ, Irwin DJ, Arzberger T, Respondek G, Lee EB, et al. Distribution patterns of tau pathology in progressive supranuclear palsy. Acta Neuropathol (Berl). 2020;140(2):99–119.
7. Wen Y, Zhou Y, Jiao B, Shen L. Genetics of progressive supranuclear palsy: a review. J Park Dis. 2021;11(1):93–105.
8. Höglinger GU, Melhem NM, Dickson DW, Sleiman PM, Wang LS, Klei L, et al. Identification of common variants influencing risk of the tauopathy progressive supranuclear palsy. Nat Genet. 2011;43(7):699–705.
9. Borroni B, Agosti C, Magnani E, Di Luca M, Padovani A. Genetic bases of Progressive Supranuclear Palsy: the MAPT tau disease. Curr Med Chem. 2011;18(17):2655–60.
10. Rademakers R, Cruts M, Van Broeckhoven C. The role of tau (MAPT) in frontotemporal dementia and related tauopathies. Hum Mutat. 2004;24(4):277–95.
11. Cooper YA, Teyssier N, Dräger NM, Guo Q, Davis JE, Sattler SM, et al. Functional regulatory variants implicate distinct transcriptional networks in dementia. Science. 2022;377(6608):eabi8654.
12. Sanchez-Contreras MY, Kouri N, Cook CN, Heckman MG, Finch NA, Caselli RJ, et al. Replication of progressive supranuclear palsy genome-wide association study identifies SLCO1A2 and DUSP10 as new susceptibility loci. Mol Neurodegener. 2018;13(1):1–10.
13. Chen JA, Chen Z, Won H, Huang AY, Lowe JK, Wojta K, et al. Joint genome-wide association study of progressive supranuclear palsy identifies novel susceptibility loci and genetic correlation to neurodegenerative diseases. Mol Neurodegener. 2018;13(1):1–11.
14. Jabbari E, Koga S, Valentino RR, Reynolds RH, Ferrari R, Tan MM, et al. Genetic determinants of survival in progressive supranuclear palsy: a genome-wide association study. Lancet Neurol. 2021;20(2):107–16.
15. Jabbari E, Woodside J, Tan MM, Shoai M, Pittman A, Ferrari R, et al. Variation at the TRIM11 locus modifies progressive supranuclear palsy phenotype. Ann Neurol. 2018;84(4):485–96.
16. Beecham GW, Bis JC, Martin ER, Choi SH, DeStefano AL, Van Duijn CM, et al. The Alzheimer's Disease Sequencing Project: study design and sample selection. Neurol Genet. 2017;3(5).
17. Kuzma A, Valladares O, Cweibel R, Greenfest-Allen E, Childress DM, Malamon J, et al. NIAGADS: The NIA Genetics of Alzheimer's Disease Data Storage Site. Alzheimers Dement. 2016;12(11):1200–3.
18. Consortium 1000 Genomes Project. A global reference for human genetic variation. Vol. 526, Nature. Nature Publishing Group; 2015. p. 68.
19. Lowy-Gallego E, Fairley S, Zheng-Bradley X, Ruffier M, Clarke L, Flicek P. Variant calling on the GRCh38 assembly with the data from phase three of the 1000 Genomes Project. Wellcome Open Res. 2019;30(4):50.
20. Genome Reference Consortium. GRCh38 reference 000001405.15 [Internet]. [cited 2022 Jun 22]. Available from: https://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000/001/405/GCA_000001405.15_GRCh38/seqs_for_alignment_pipelines.ucsc_ids/GCA_000001405.15_GRCh38_no_alt_analysis_set.fna.gz.
21. Schneider VA, Graves-Lindsay T, Howe K, Bouk N, Chen HC, Kitts PA, et al. Evaluation of GRCh38 and de novo haploid genome assemblies demonstrates the enduring quality of the reference assembly. Genome Res. 2017;27(5):849–64.
22. Yang J, Bakshi A, Zhu Z, Hemani G, Vinkhuyzen AA, Lee SH, et al. Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. Nat Genet. 2015;47(10):1114–20.
23. Gogarten SM, Sofer T, Chen H, Yu C, Brody JA, Thornton TA, et al. Genetic association testing using the GENESIS R/Bioconductor package. Bioinformatics. 2019;35(24):5346–8.
24. 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.
25. Conomos MP, Miller MB, Thornton TA. Robust inference of population structure for ancestry prediction and correction of stratification in the presence of relatedness. Genet Epidemiol. 2015;39(4):276–93.
26. Zou Y, Carbonetto P, Wang G, Stephens M. Fine-mapping from summary data with the "Sum of Single Effects" model. PLoS Genet. 2022;18(7):e1010299.

27. Cook S, Choi W, Lim H, Luo Y, Kim K, Jia X, et al. Accurate imputation of human leukocyte antigens with CookHLA. Nat Commun. 2021;12(1):1264.
28. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38(16):e164.
29. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, et al. The Ensembl Variant Effect Predictor. Genome Biol. 2016;17(1):122.
30. Cunningham F, Allen JE, Allen J, Alvarez-Jarreta J, Amode MR, Armean IM, et al. Ensembl 2022. Nucleic Acids Res. 2022;50(D1):D988–95.
31. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature. 2020;581(7809):434–43.
32. Swarup V, Chang TS, Duong DM, Dammer EB, Dai J, Lah JJ, et al. Identification of Conserved Proteomic Networks in Neurodegenerative Dementia. Cell Rep. 2020;31(12):107807.
33. Sjöstedt E, Zhong W, Fagerberg L, Karlsson M, Mitsios N, Adori C, et al. An atlas of the protein-coding genes in the human, pig, and mouse brain. Science. 2020;367(6482):eaay5947.
34. Melé M, Ferreira PG, Reverter F, DeLuca DS, Monlong J, Sammeth M, et al. Human genomics. The human transcriptome across tissues and individuals. Science. 2015;348(6235):660–5.
35. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;29(9):559.
36. 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.
37. Layer RM, Chiang C, Quinlan AR, Hall IM. LUMPY: a probabilistic framework for structural variant discovery. Genome Biol. 2014;15(6):1–19.
38. Eggertsson HP, Kristmundsdottir S, Beyter D, Jonsson H, Skuladottir A, Hardarson MT, et al. GraphTyper2 enables population-scale genotyping of structural variation using pangenome graphs. Nat Commun. 2019;10(1):1–8.
39. Wang H, Dombroski BA, Cheng PL, Tucci A, Si Y qin, 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;
40. Belyeu JR, Chowdhury M, Brown J, Pedersen BS, Cormier MJ, Quinlan AR, et al. Samplot: a platform for structural variant visual validation and automated filtering. Genome Biol. 2021;22(1):1–13.
41. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform. 2013;14(2):178–92.
42. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81(3):559–75.
43. Lee S, Emond MJ, Bamshad MJ, Barnes KC, Rieder MJ, Nickerson DA, et al. Optimal Unified Approach for Rare-Variant Association Testing with Application to Small-Sample Case-Control Whole-Exome Sequencing Studies. Am J Hum Genet. 2012;91(2):224–37.
44. Wang X, Campbell MR, Lacher SE, Cho HY, Wan M, Crowl CL, et al. A polymorphic antioxidant response element links NRF2/sMAF binding to enhanced MAPT expression and reduced risk of Parkinsonian disorders. Cell Rep. 2016;15(4):830–42.
45. Anaya F, Lees A, Silva R. Tau gene promoter rs242557 and allele-specific protein binding. Transl Neurosci. 2011 Jan 1 [cited 2023 Nov 9];2(2). Available from: https://doi.org/10.2478/s13380-011-0021-6/html
46. Sawa A, Amano N, Yamada N, Kajio H, Yagishita S, Takahashi T, et al. Apolipoprotein E in progressive supranuclear palsy in Japan. Mol Psychiatry. 1997;2(4):341–2.
47. Zhao N, Liu CC, Van Ingelgom AJ, Linares C, Kurti A, Knight JA, et al. APOE ε2 is associated with increased tau pathology in primary tauopathy. Nat Commun. 2018;9(1):4388.
48. Farrell K, Humphrey J, Chang T, Zhao Y, Leung YY, Kuksa PP, et al. Genetic, transcriptomic, histological, and biochemical analysis of progressive supranuclear palsy implicates glial activation and novel risk genes. bioRxiv. 2023;2023–11.
49. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature. 2016;536(7616):285–91.
50. Chen S, Francioli LC, Goodrich JK, Collins RL, Kanai M, Wang Q, et al. A genome-wide mutational constraint map quantified from variation in 76,156 human genomes. bioRxiv. 2022;2022–03.
51. Lee WP, Choi SH, Shea MG, Cheng PL, Dombroski BA, Pitsillides AN, et al. Association of Common and Rare Variants with Alzheimer's Disease in over 13,000 Diverse Individuals with Whole-Genome Sequencing from the Alzheimer's Disease Sequencing Project. medRxiv. 2023;2023–09.
52. Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, et al. The genotype-tissue expression (GTEx) project. Nat Genet. 2013;45(6):580.
53. Xie R, Nguyen S, McKeehan K, Wang F, McKeehan WL, Liu L. Microtubule-associated protein 1S (MAP1S) bridges autophagic components with microtubules and mitochondria to affect autophagosomal biogenesis and degradation. J Biol Chem. 2011;286(12):10367–77.
54. Shi L, Huang C, Luo Q, Xia Y, Liu H, Li L, et al. Pilot study: molecular risk factors for diagnosing sporadic Parkinson's disease based on gene expression in blood in MPTP-induced rhesus monkeys. Oncotarget. 2017;8(62):105606.
55. Pan M, Li X, Xu G, Tian X, Li Y, Fang W. Tripartite Motif Protein Family in Central Nervous System Diseases. Cell Mol Neurobiol. 2023;1–23.
56. Valcourt U, Alcaraz LB, Exposito JY, Lethias C, Bartholin L. Tenascin-X: beyond the architectural function. Cell Adhes Migr. 2015;9(1–2):154–65.
57. Kutkowska-Kaźmierczak A, Rydzanicz M, Chlebowski A, K\losowska-Kosicka K, Mika A, Gruchota J, et al. Dominant ELOVL1 mutation causes neurological disorder with ichthyotic keratoderma, spasticity, hypomyelination and dysmorphic features. J Med Genet. 2018;55(6):408–14.
58. Swarup V, Hinz FI, Rexach JE, Noguchi K ichi, Toyoshiba H, Oda A, et al. Identification of evolutionarily conserved gene networks mediating neurodegenerative dementia. Nat Med. 2019;25(1):152–64.
59. Swarup V, Chang TS, Duong DM, Dammer EB, Dai J, Lah JJ, et al. Identification of conserved proteomic networks in neurodegenerative dementia. Cell Rep. 2020 [cited 2023 Nov 9];31(12). Available from: https://www.cell.com/cell-reports/pdf/S2211-1247(20)30788-9.pdf.
60. Parikshak NN, Gandal MJ, Geschwind DH. Systems biology and gene networks in neurodevelopmental and neurodegenerative disorders. Nat Rev Genet. 2015;16(8):441–58.
61. 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.
62. Wang H, Wang LS, Schellenberg G, Lee WP. The role of structural variations in Alzheimer's disease and other neurodegenerative diseases. Front Aging Neurosci. 2023.
63. Mizobuchi M, Murao K, Takeda R, Kakimoto Y. Tissue-specific expression of isoaspartyl protein carboxyl methyltransferase gene in rat brain and testis. J Neurochem. 1994;62(1):322–8.
64. Wu X, Jia G, Yang H, Sun C, Liu Y, Diao Z. Neural stem cell-conditioned medium upregulated the PCMT1 expression and inhibited the phosphorylation of MST1 in SH-SY5Y cells induced by Aβ 25–35. Biocell. 2022;46(2):471.
65. Shi L, Al-Baadani A, Zhou K, Shao A, Xu S, Chen S, et al. PCMT1 ameliorates neuronal apoptosis by inhibiting the activation of MST1 after subarachnoid hemorrhage in rats. Transl Stroke Res. 2017;8:474–83.
66. Smit, AFA, Hubley, R & Green, P. RepeatMasker Open-4.0. 2013–2015 <http://www.repeatmasker.org>.
67. Chen JA, Chen Z, Won H, Huang AY, Lowe JK, Wojta K, et al. Joint genome-wide association study of progressive supranuclear palsy identifies novel susceptibility loci and genetic correlation to neurodegenerative diseases. Mol Neurodegener. 2018;13(1):41.
68. Rizzu P, Van Swieten JC, Joosse M, Hasegawa M, Stevens M, Tibben A, et al. High prevalence of mutations in the microtubule-associated protein tau in a population study of frontotemporal dementia in the Netherlands. Am J Hum Genet. 1999;64(2):414–21.
69. Rovelet-Lecrux A, Lecourtois M, Thomas-Anterion C, Le Ber I, Brice A, Frebourg T, et al. Partial deletion of the MAPT gene: A novel mechanism of FTDP-17. Hum Mutat. 2009;30(4):E591-602.
70. Farrer LA, Cupples LA, Haines JL, Hyman B, Kukull WA, Mayeux R, et al. Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease: a meta-analysis. JAMA. 1997;278(16):1349–56.
71. Selkoe DJ, Podlisny MB. Deciphering the genetic basis of Alzheimer's disease. Annu Rev Genomics Hum Genet. 2002;3(1):67–99.

72. Shi Y, Yamada K, Liddelow SA, Smith ST, Zhao L, Luo W, et al. ApoE4 markedly exacerbates tau-mediated neurodegeneration in a mouse model of tauopathy. Nature. 2017;549(7673):523–7.
73. Rasmussen KL, Tybjærg-Hansen A, Nordestgaard BG, Frikke-Schmidt R. Associations of Alzheimer disease–protective APOE variants with age-related macular degeneration. JAMA Ophthalmol. 2023;141(1):13–21.
74. Klaver CC, Kliffen M, van Duijn CM, Hofman A, Cruts M, Grobbee DE, et al. Genetic association of apolipoprotein E with age-related macular degeneration. Am J Hum Genet. 1998;63(1):200–6.
75. Cebrián C, Loike JD, Sulzer D. Neuronal MHC-I expression and its implications in synaptic function, axonal regeneration and Parkinson's and other brain diseases. Front Neuroanat. 2014;8:114.
76. Dressman D, Elyaman W. T Cells: A Growing Universe of Roles in Neurodegenerative Diseases. Neuroscientist. 2022;28(4):335–48.
77. Maiers M, Gragert L, Klitz W. High-resolution HLA alleles and haplotypes in the United States population. Hum Immunol. 2007;68(9):779–88.
78. Lee S, Abecasis GR, Boehnke M, Lin X. Rare-variant association analysis: study designs and statistical tests. Am J Hum Genet. 2014;95(1):5–23.
79. Cassandri M, Smirnov A, Novelli F, Pitolli C, Agostini M, Malewicz M, et al. Zinc-finger proteins in health and disease. Cell Death Discov. 2017;3(1):1–12.
80. Fedotova AA, Bonchuk AN, Mogila VA, Georgiev PG. C2H2 Zinc Finger Proteins: The Largest but Poorly Explored Family of Higher Eukaryotic Transcription Factors. Acta Naturae. 2017;9(2):47–58.
81. Bu S, Lv Y, Liu Y, Qiao S, Wang H. Zinc Finger Proteins in Neuro-Related Diseases Progression. Front Neurosci. 2021;18(15):760567.
82. Al-Naama N, Mackeh R, Kino T. C2H2-Type Zinc Finger Proteins in Brain Development, Neurodevelopmental, and Other Neuropsychiatric Disorders: Systematic Literature-Based Analysis. Front Neurol. 2020;11:32.
83. Shin JH, Ko HS, Kang H, Lee Y, Lee YI, Pletinkova O, et al. PARIS (ZNF746) Repression of PGC-1α Contributes to Neurodegeneration in Parkinson's Disease. Cell. 2011;144(5):689–702.
84. Li R, Strohmeyer R, Liang Z, Lue LF, Rogers J. CCAAT/enhancer binding protein delta (C/EBPdelta) expression and elevation in Alzheimer's disease. Neurobiol Aging. 2004;25(8):991–9.
85. Ko CY, Chang LH, Lee YC, Sterneck E, Cheng CP, Chen SH, et al. CCAAT/enhancer binding protein delta (CEBPD) elevating PTX3 expression inhibits macrophage-mediated phagocytosis of dying neuron cells. Neurobiol Aging. 2012;33(2):422.e11-25.
86. Nicolas E, Poitelon Y, Chouery E, Salem N, Levy N, Mégarbané A, et al. CAMOS, a nonprogressive, autosomal recessive, congenital cerebellar ataxia, is caused by a mutant zinc-finger protein, ZNF592. Eur J Hum Genet EJHG. 2010;18(10):1107–13.
87. Vodopiutz J, Seidl R, Prayer D, Khan MI, Mayr JA, Streubel B, et al. WDR73 Mutations Cause Infantile Neurodegeneration and Variable Glomerular Kidney Disease. Hum Mutat. 2015;36(11):1021–8.
88. Parikshak NN, Luo R, Zhang A, Won H, Lowe JK, Chandran V, et al. Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. Cell. 2013;155(5):1008–21.
89. Takahashi M, Weidenheim KM, Dickson DW, Ksiezak-Reding H. Morphological and biochemical correlations of abnormal tau filaments in progressive supranuclear palsy. J Neuropathol Exp Neurol. 2002;61(1):33–45.
90. Roemer SF, Grinberg LT, Crary JF, Seeley WW, McKee AC, Kovacs GG, et al. Rainwater Charitable Foundation criteria for the neuropathologic diagnosis of progressive supranuclear palsy. Acta Neuropathol (Berl). 2022;144(4):603–14.
91. Litvan I, Agid Y, Calne D, Campbell G, Dubois B, Duvoisin RC, et al. Clinical research criteria for the diagnosis of progressive supranuclear palsy (Steele-Richardson-Olszewski syndrome): report of the NINDS-SPSP international workshop. Neurology. 1996;47(1):1–9.
92. Osaki Y, Ben-Shlomo Y, Lees AJ, Daniel SE, Colosimo C, Wenning G, et al. Accuracy of clinical diagnosis of progressive supranuclear palsy. Mov Disord Off J Mov Disord Soc. 2004;19(2):181–9.
93. Lopez OL, Litvan I, Catt KE, Stowe R, Klunk W, Kaufer DI, et al. Accuracy of four clinical diagnostic criteria for the diagnosis of neurodegenerative dementias. Neurology. 1999;53(6):1292–9.
94. Litvan I, Campbell G, Mangone CA, Verny M, McKee A, Chaudhuri KR, et al. Which clinical features differentiate progressive supranuclear palsy (Steele-Richardson-Olszewski syndrome) from related disorders? A clinicopathological study. Brain J Neurol. 1997;120(Pt 1):65–74.
95. Hokelekli FO, Duffy JR, Clark HM, Utianski RL, Botha H, Ali F, et al. Autopsy Validation of Progressive Supranuclear Palsy-Predominant Speech/Language Disorder Criteria. Mov Disord Off J Mov Disord Soc. 2022;37(1):213–8.
96. Gazzina S, Respondek G, Compta Y, Allinson KSJ, Spillantini MG, Molina-Porcel L, et al. Neuropathological validation of the MDS-PSP criteria with PSP and other frontotemporal lobar degeneration. bioRxiv; 2019 [cited 2023 Aug 7]. p. 520510. Available from: https://doi.org/10.1101/520510v1


## Publisher's Note