Association of Structural Forms of 17q21.31 with the Risk of Progressive Supranuclear Palsy and MAPT Sub-haplotypes

Hui Wang1,2, PhD; Timothy S Chang3, MD, PhD; Beth A Dombroski1,2, PhD; Po-Liang Cheng1,2, PhD; Ya-Qin Si4, PhD; Albert Tucci4, PhD; Vishakha Patil5, MS; Leopoldo Valiente-Banuet5, MS; Kurt Farrell6, PhD; Catriona Mclean7, MD; Laura Molina-Porcel8,9, MD, PhD; Alex Rajput10, MD; Peter Paul De Deyn11,12, MD, PhD; Nathalie Le Bastard13, PhD; Marla Gearing14, PhD; Laura Donker Kaat15, MD, PhD; John C Van Swieten15, MD, PhD; Elise Dopper15, MD, PhD; Bernardino F Ghetti16, MD; Kathy L Newell16, MD; Claire Troakes17, PhD; Justo G de Yébenes18, MD, PhD; Alberto Rábano-Gutierrez19, MD, PhD; Tina Meller20, PhD; Wolfgang H Oertel20, MD, PhD; Gesine Respondek21, MD; Maria Stamelou22,23, MD; Thomas Arzberger24,25, MD; Sigrun Roeber26, MD; Ulrich Müller26, MD; Franziska Hopfner44, MD; Pau Pastor27,28, MD, PhD; Alexis Brice29, MD; Alexandra Durr29, MD, PhD; Isabelle Le Ber29, MD, PhD; Thomas G Beach30, MD, PhD; Geidy E Serrano30, PhD; Lili-Naz Hazrati31, MD, PhD; Irene Litvan32, MD; Rosa Rademakers33,34, PhD; Owen A Ross34, PhD; Douglas Galasko32, MD; Adam L Boxer35, MD, PhD; Bruce L Miller35, MD, PhD; Edward B Lee1,36, MD, PhD; Charles L White III37, MD; Huw R Morris38, MD, PhD; Rohan de Silva39, PhD; John F Crary6, MD, PhD; Alison M Goate40, PhD; Jeffrey S Friedman41, MD, PhD; Yuk Yee Leung1,2, PhD; Giovanni Coppola5,42, MD; Adam C Naj1,2,43, PhD; Li-San Wang1,2, PhD; PSP genetics study group; Dennis W Dickson34, MD; Günter U Höglinger44, MD; Jung-Ying Tzeng4,45, PhD; Daniel H Geschwind5,46,47, MD, PhD; Gerard D Schellenberg1,2, PhD; Wan-Ping Lee1,2, PhD

1Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
2Penn Neurodegeneration Genomics Center, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
3Movement Disorders Programs, Department of Neurology, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA
4Bioinformatics Research Center, North Carolina State University, NC, USA
5Movement Disorders Programs, Department of Neurology, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
6Department 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.

7Victorian Brain Bank, The Florey Institute of Neuroscience and Mental Health, Parkville, Victoria, Australia

8Alzheimer’s disease and other cognitive disorders unit. Neurology Service, Hospital Clínic, Fundació Recerca Clínica Barcelona (FRCB). Institut d’Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), University of Barcelona, Barcelona, Spain

9Neurological Tissue Bank of the Biobanc-Hospital Clínic-IDIBAPS, Barcelona, Spain

10Movement Disorders Program, Division of Neurology, University of Saskatchewan, Saskatoon, Saskatchewan, Canada

11Laboratory of Neurochemistry and Behavior, Experimental Neurobiology Unit, University of Antwerp, Wilrijk (Antwerp), Belgium

12Department of Neurology, University Medical Center Groningen, NL-9713 AV Groningen, Netherlands

13Fujirebio Europe NV, Technologiepark 6, 9052 Gent, Belgium

14Department of Pathology and Laboratory Medicine and Department of Neurology, Emory University School of Medicine, Atlanta, GA, USA

15Netherlands Brain Bank and Erasmus University, Netherlands

16Department of Pathology and Laboratory Medicine, Indiana University School of Medicine, Indianapolis, IN, USA

17London Neurodegenerative Diseases Brain Bank, King's College London, London, UK

18Autonomous University of Madrid, Madrid, Spain

19Fundación CIEN (Centro de Investigación de Enfermedades Neurológicas) - Centro Alzheimer Fundación Reina Sofia, Madrid, Spain

20Department of Neurology, Philipps-Universität, Marburg, Germany

21German Center for Neurodegenerative Diseases (DZNE), Munich, Germany

22Parkinson’s disease and Movement Disorders Department, HYGEIA Hospital, Athens, Greece

23European University of Cyprus, Nicosia, Cyprus

24Department of Psychiatry and Psychotherapy, University Hospital Munich, Ludwig-Maximilians-University Munich, Germany

25Center for Neuropathology and Prion Research, Ludwig-Maximilians-University Munich, Germany

26German Brain Bank, Neurobiobank Munich, Germany

27Unit of Neurodegenerative diseases, Department of Neurology, University Hospital Germans Trias i Pujol, Badalona, Barcelona, Spain

28Neurosciences, The Germans Trias i Pujol Research Institute (IGTP) Badalona, Badalona, Spain

29Sorbonne Université, Paris Brain Institute – Institut du Cerveau – ICM, Inserm U1127, CNRS UMR 7225, APHP - Hôpital Pitié-Salpêtrière, Paris, France
Search Terms:
Progressive Supranuclear Palsy (PSP), H1 and H2 haplotypes, MAPT, Copy number variation (CNV), 17q21.31

Word Count: 3902 (Key Points, Abstract, and Main Text)

Corresponding Author:
Wan-Ping Lee
D103 Medical Research Laboratories, University of Pennsylvania, 3700 Hamilton Walk, Philadelphia, PA 19104, USA
Phone: +1 412-880-8674
Email: wan-ping.lee@pennmedicine.upenn.edu
108 Statistical Analysis conducted by Hui Wang, Ph.D., University of Pennsylvania
Key Points

**Question**: Do large copy number variations (i.e., α, β, and γ) inside 17q21.31 contribute to the risk of progressive supranuclear palsy (PSP) independently from the H1 and H2 haplotypes? Do structural forms of 17q21.31, characterized by combinations of α, β, and γ, present divergent risk to the development of PSP? Are structural forms of 17q21.31 associated with MAPT sub-haplotypes, such as H1c?

**Findings**: In this case-control study of 1,684 individuals with PSP and 2,392 control subjects, the copy number of γ duplication was independently associated with the risk of the disease. H1 haplotypes with more γ duplications (H1β1γ2, H1β1γ3, and H1β1γ4) displayed a higher odds ratio for PSP when compared to H1β1γ1. Notably, H1β1γ3 was observed to be in linkage disequilibrium with H1c, a widely recognized MAPT sub-haplotype associated with PSP.

**Meaning**: The association between the H1 and H2 haplotypes and PSP involves multiple contributing factors, including the copy number of γ duplication.
Abstract

Importance: The chromosome 17q21.31 region, containing a 900 Kb inversion that defines H1 and H2 haplotypes, represents the strongest genetic risk locus in progressive supranuclear palsy (PSP). In addition to H1 and H2, various structural forms of 17q21.31, characterized by the copy number of α, β, and γ duplications, have been identified. However, the specific effect of each structural form on the risk of PSP has never been evaluated in a large cohort study.

Objective: To assess the association of different structural forms of 17q21.31, defined by the copy numbers of α, β, and γ duplications, with the risk of PSP and MAPT sub-haplotypes.

Design, setting, and participants: Utilizing whole genome sequencing data of 1,684 (1,386 autopsy confirmed) individuals with PSP and 2,392 control subjects, a case-control study was conducted to investigate the association of copy numbers of α, β, and γ duplications and structural forms of 17q21.31 with the risk of PSP. All study subjects were selected from the Alzheimer’s Disease Sequencing Project (ADSP) Umbrella NG00067.v7. Data were analyzed between March 2022 and November 2023.

Main outcomes and measures: The main outcomes were the risk (odds ratios [ORs]) for PSP with 95% CIs. Risks for PSP were evaluated by logistic regression models.

Results: The copy numbers of α and β were associated with the risk of PSP only due to their correlation
with H1 and H2, while the copy number of γ was independently associated with the increased risk of PSP. Each additional duplication of γ was associated with 1.10 (95% CI, 1.04-1.17; \( P = 0.0018 \)) fold of increased risk of PSP when conditioning H1 and H2. For the H1 haplotype, addition γ duplications displayed a higher odds ratio for PSP: the odds ratio increases from 1.21 (95% CI 1.10-1.33, \( P = 5.47 \times 10^{-5} \)) for H1β1γ1 to 1.29 (95% CI 1.16-1.43, \( P = 1.35 \times 10^{-6} \)) for H1β1γ2, 1.45 (95% CI 1.27-1.65, \( P = 3.94 \times 10^{-8} \)) for H1β1γ3, and 1.57 (95% CI 1.10-2.26, \( P = 1.35 \times 10^{-2} \)) for H1β1γ4. Moreover, H1β1γ3 is in linkage disequilibrium with H1c (\( R^2 = 0.31 \)), a widely recognized \( MAPT \) sub-haplotype associated with increased risk of PSP. The proportion of \( MAPT \) sub-haplotypes associated with increased risk of PSP (i.e., H1c, H1d, H1g, H1o, and H1h) increased from 34% in H1β1γ1 to 77% in H1β1γ4.

**Conclusions and relevance:** This study revealed that the copy number of γ was associated with the risk of PSP independently from H1 and H2. The H1 haplotype with more γ duplications showed a higher odds ratio for PSP and were associated with \( MAPT \) sub-haplotypes with increased risk of PSP. These findings expand our understanding of how the complex structure at 17q21.31 affect the risk of PSP.
Introduction

Progressive supranuclear palsy (PSP) is a neurodegenerative disease with two characteristic clinical features, i.e. postural instability and ocular motor abnormalities\(^1\). Other symptoms and signs, such as cognitive disfunction and problems with swallowing, vary and can get worse over time depending on the distribution of pathology and severity of diseases\(^2,3\). The main pathology of PSP is the accumulation of tau in the brain, leading to the presence of neurofibrillary tangles and threads along with tufted astrocytes and oligodendroglial coiled bodies\(^4,5\). An isoform of tau harboring 4 repeats (4R) of microtubule-binding domain is particularly prominent in these tau aggregates\(^6\).

The most recognized genetic risk locus for PSP is situated on 17q21.31. The region can be divided into two major haplotypes, H1 and H2, characterized by a 900 Kb inversion found in 20\% of Europeans\(^7\). Individuals carrying the H1 haplotype, compared to the H2 haplotype, are more likely to develop PSP, with an estimated odds ratio (OR) around 5 in Europeans\(^8,9\). The association was identified through various variants in linkage disequilibrium (LD) with H1 and H2, such as a dinucleotide repeat (TG\(_n\)) in \(MAPT\) intron \(^9\(^{10}\), a 238-bp deletion in \(MAPT\) intron \(^9\(^7\), and a multitude of single nucleotide variants (SNVs) in this region\(^8,9\). However, the causal variants underlying the association remain unclear due to numerous SNVs, short insertion/deletions (indels), and structural variations (SVs) introduced by complex genomic rearrangements in 17q.21.31.

Based on LD structure in the \(MAPT\) region, H1 and H2 haplotypes can be further categorized into more than 20 common sub-haplotypes\(^11,13\). Among them, H1c, H1d, H1g, and H1o were significantly associated with the increased risk of PSP\(^11,13\). Particularly, H1c and H1d are nominally associated with
measures of severity of tau pathology in PSP cases\textsuperscript{11}. However, these sub-haplotypes were inferred from LD structure limited to \textit{MAPT} gene (~150 Kb) and do not represent the intricacies of structural forms on 17q21.31 (~1.5 Mb). To address this limitation, we constructed the ten structural forms of 17q21.31, characterized by the copy numbers of $\alpha$, $\beta$, and $\gamma$ (150Kb, 300Kb, and 218Kb; eFigure 1 in \textit{Supplement})\textsuperscript{14,15}, and evaluated their association with the risk of PSP in this study. We found that the H1 haplotype with more copies $\gamma$ duplication showed an increased risk of PSP. The result is in accordance with the higher portion of H1c and other \textit{MAPT} sub-haplotypes associated with increased risk of PSP in the H1 haplotypes with more copies of $\gamma$. Through this study, we gained a better understanding how the haplotypic structure of 17q.21.31 correlates with the risk of PSP.

\textbf{Methods}

\textit{Study subjects and quality control}

All study subjects and WGS data are available on The National Institute on Aging Genetics of Alzheimer's Disease Data Storage Site (NIAGADS)\textsuperscript{16} under Alzheimer’s Disease Sequencing Project (ADSP) Umbrella NG00067.v7\textsuperscript{17}. We inferred ancestry of subjects by GRAF-pop\textsuperscript{18} and selected 4,618 subjects (1,797 cases and 2,821 controls) of European ancestry for analysis. WGS were performed at 30x coverage (eTable 1).

Among 4,618 samples, we filtered 183 samples with abnormally low reads mapped (aligned read depth < 1.7x) to $\alpha$, $\beta$ or $\gamma$ region (eFigure 2 in \textit{Supplement}) and ten samples with high genotyping missing rate (> 0.05). Next, 244 related samples inferred by KING\textsuperscript{19} (duplicates, monozygotic twins,
parent-offsprings, full-siblings, and 2nd degree relatives) were removed. We used the 238-bp deletion between exons 9 and 10 of \( MAPT \) to determine the H1 and H2 haplotypes of each sample. The genotype calls of the 238-bp deletion were obtained from our previous SV work\(^\text{20} \). 75 subjects were removed due to missing or failed genotype of the 238-bp deletion. Given the specification of H1/H2 genotype, determined by the 238 bp deletion, and the copy numbers of \( \alpha \), \( \beta \) and \( \gamma \), we can ascertain the ten structural forms (eFigure 1 in Supplement) in each individual. We removed 30 individuals (eFigure 3 and eFigure 4 in Supplement) since their structural forms could not be decided based on the copy numbers of \( \alpha \), \( \beta \) and \( \gamma \). This discordance might be due to subjects carrying undiscovered structural forms or genotyping errors on the copy numbers of \( \alpha \), \( \beta \) and \( \gamma \). As a result, 4,076 subjects (Table 1; \( N_{PSP} = 1,684 \), \( N_{control} = 2,392 \)) remained for statistical analyses in this study. Of the 1,684 individuals with PSP, 1,386 were autopsy confirmed.
Table 1. Characteristics of PSP cases and controls

<table>
<thead>
<tr>
<th></th>
<th>Overall (N = 4,076)</th>
<th>PSP (N = 1,684)</th>
<th>Control (N = 2,392)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age</strong> (SD)</td>
<td>78.49 (8.50)</td>
<td>68.03 (8.17)</td>
<td>81.04 (6.37)</td>
</tr>
<tr>
<td><strong>Sex (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>2168 (53.19%)</td>
<td>739 (43.88%)</td>
<td>1429 (59.74%)</td>
</tr>
<tr>
<td>Male</td>
<td>1908 (46.81%)</td>
<td>945 (56.12%)</td>
<td>963 (40.26%)</td>
</tr>
<tr>
<td><strong>H1/H2 status</strong> (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H1H1</td>
<td>2958 (72.57%)</td>
<td>1511 (89.73%)</td>
<td>1447 (60.49%)</td>
</tr>
<tr>
<td>H1H2</td>
<td>975 (23.92%)</td>
<td>168 (9.98%)</td>
<td>807 (33.74%)</td>
</tr>
<tr>
<td>H2H2</td>
<td>143 (3.51%)</td>
<td>5 (0.30%)</td>
<td>138 (5.77%)</td>
</tr>
<tr>
<td><strong>Structural forms of 17q21.31 (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H1β1γ1</td>
<td>2446 (30.00%)</td>
<td>1097 (32.57%)</td>
<td>1349 (28.20%)</td>
</tr>
<tr>
<td>H1β1γ2</td>
<td>1552 (19.04%)</td>
<td>739 (21.94%)</td>
<td>813 (16.99%)</td>
</tr>
<tr>
<td>H1β1γ3</td>
<td>987 (12.11%)</td>
<td>496 (14.73%)</td>
<td>491 (10.26%)</td>
</tr>
<tr>
<td>H1β1γ4</td>
<td>126 (1.55%)</td>
<td>65 (1.93%)</td>
<td>61 (1.28%)</td>
</tr>
<tr>
<td>H1β2γ1</td>
<td>1716 (21.05%)</td>
<td>774 (22.98%)</td>
<td>942 (19.69%)</td>
</tr>
<tr>
<td>H1β3γ1</td>
<td>64 (0.79%)</td>
<td>19 (0.56%)</td>
<td>45 (0.94%)</td>
</tr>
<tr>
<td>H2α1γ1</td>
<td>7 (0.09%)</td>
<td>1 (0.03%)</td>
<td>6 (0.13%)</td>
</tr>
<tr>
<td>H2α1γ2</td>
<td>99 (1.21%)</td>
<td>14 (0.42%)</td>
<td>85 (1.78%)</td>
</tr>
<tr>
<td>H2α2γ1</td>
<td>33 (0.40%)</td>
<td>2 (0.06%)</td>
<td>31 (0.65%)</td>
</tr>
<tr>
<td>H2α2γ2</td>
<td>1122 (13.76%)</td>
<td>161 (4.78%)</td>
<td>961 (20.09%)</td>
</tr>
</tbody>
</table>

*a1,130 PSP cases and 111 controls have missing age.
*bH1/H2 status was determined by the genotype of a 238-bp H2 tagging deletion.
*cStructural forms of 17q21.31 were inferred by expectation-maximization (EM) algorithm.
Calling the copy numbers α, β and γ duplications from WGS

The genomic coordinates on HG38 of α (chr17:46135415-46289349), β (chr17:46087894-46356512), and γ (chr17:46289349-46707123) were obtained from two previous studies14,15 (eFigure 1 in Supplement). Segmental duplications can introduce mapping challenges and thus inaccurate calling of the number of copies21–23. To address this, we removed segmental duplicated regions inside the α, β and γ (eFigure 5 in Supplement) when calculating aligned read depth. Subsequently, the copy numbers of α, β, and γ were obtained based on the aligned read depth on chr17:46135415-46203287, chr17:46106189-46135415, and chr17:46356512-46489410/chr17:46565081-46707123, respectively. Copies of α, β and γ were genotyped by assessing aligned read depth within each 1 Kb bin on the specified regions using CNVpytor (Version 1.3.1)24. Then, we employed K-means25 to assign an integer copy number for α, β and γ for the 4,076 individuals.

Validation of CNV calling

To validate the copy numbers of α, β and γ called from WGS, 66 samples were genotyped using TaqMan CNV assay. Overall, the copy number of α, β, or γ inferred by aligned read depth from WGS were highly consistent (α, R = 0.87; γ, R = 0.96) with that from TaqMan assay (eFigure 6 in Supplement). The experimental procedure was as follows: Copy number at the α region (Context Sequence: TTAGTCAATTTCCTTAGCCAACC; chr17:46171656-46171680, GRCh38) was determined using a predesigned TaqMan CNV assay (Applied Biosystems Part No. Hs01788222_cn).
Copy number at the $\gamma$ region (Context Sequence: AGAAAAAAAGCATTGACTCCCAACCC; chr17:46419867-46419891 and chr17:46637445-46637469, GRCh38) was determined using a custom designed TaqMan CNV assay (Applied Biosystems; Forward Primer: TGGCACAATGACCA TCGAGATT, Reverse Primer: CTGCCATCTTGTGGGTGTCA, Reporter Sequence: AAGGGTTGGAGTCAATGCTTTT). For the copy number assay reaction, 20 ng of genomic DNA was combined with TaqMan Master Mix (Applied Biosystems Part No. 4371357), 20X TaqMan CNV assay, and 20x RNase P TaqMan Copy Number Reference Assay (Applied Biosystems Part No. 4403328) in a total volume of 20ul. The DNA amplification and quantification were carried out in a QuantStudio 12K Flex instrument (Applied Biosystems) in a 96 well format with the following program: 50°C 2min, 95°C 10min followed by 40 cycles of 95°C 15sec, 60°C 1min. The results were analyzed using the TaqMan CopyCaller software (Version 2.0) (Applied Biosystems).

**Determine haplotypic contributions to diploid copy number**

For approximately sixty percent of the samples, only one combination of the structural forms (eFigure 1 in Supplement) was possible based on the H1 and H2 genotypes, determined by the 238-bp deletion, and the copy numbers of $\alpha$, $\beta$ and $\gamma$. For the rest of the samples, multiple haplotypic combinations were possible. Therefore, we initially assigned an equal likelihood to all possible haplotypic combinations that were consistent with the detected copy number of $\alpha$, $\beta$, and $\gamma$. Then, the following expectation-maximization (EM) loop described in the previous study$^{14}$ were repeated: from the probabilistic inferences of structural form of 17q21.31 in each sample, we estimated an allele
frequency for each structural form of 17q21.31; we then re-estimated the relative likelihood of each
combination of haplotypes in each sample, given the population-level allele frequency. In this way, we
intended to eliminate haplotypic combinations that are theoretically possible but extremely unlikely
given the apparent frequencies of haplotypes as estimated from the rest of the population. The revised
probabilistic estimates then allowed a new population-level estimate of haplotype frequencies. We
repeated this EM loop process until estimates of haplotype frequency converged. The allele frequency of
each structural form of 17q21.31 after convergence were showed in eFigure 1 in Supplement.

Phasing for MAPT sub-haplotypes

The six SNVs (rs1467967, rs242557, rs3785883, rs2471738, rs8070723, and rs7521)\textsuperscript{11–13} on MAPT
were employed to defined the 26 MAPT sub-haplotypes (eTable 2 in Supplement). We phased the six
SNVs with other SNVs and indels in chr17:43000000-48000000 to determine the MAPT sub-haplotypes.
The SNV genotypes for the study subjects were called in our previous work\textsuperscript{26}. Variants were removed if
they were monomorphic, did not pass variant quality score recalibration, had an average read depth \( \geq \)
500, or if all calls have DP<10 & GQ<20. Individual calls with a DP<10 or GQ<20 were set to missing.
Then, common variants (MAF > 0.01) with 0.25 < ABHet < 0.75 were phased using SHAPEIT4\textsuperscript{27}
(Version 4.2.2).

Linkage disequilibrium between structural forms of 17q21.31 and MAPT sub-haplotypes

To phase the structural forms of 17q21.31 together with MAPT sub-haplotypes, we encoded the
copy numbers of $\alpha$, $\beta$ and $\gamma$ as multi-allelic CNVs by a series of surrogate bi-allelic markers with 0/1 alleles\(^{14}\) (eTable 3 in Supplement). Then, SHAPEIT\(^{27}\) (Version 4.2.2) were used for phasing the copy numbers of $\alpha$, $\beta$ and $\gamma$ together with SNVs/indels. SNVs and indels inside $\alpha$, $\beta$, and $\gamma$ regions (chr17:46087000-46708000) were not included when phasing. After phasing, we calculated the linkage disequilibrium (LD) between structural forms of 17q21.31 and MAPT sub-haplotypes.

Statistical Analysis

Statistical analyses were performed for the 4,076 individuals ($N_{\text{PSP}} = 1,684$, $N_{\text{control}} = 2,392$). For the association of the copy numbers of $\alpha$, $\beta$, and $\gamma$ with PSP, logistic regression adjusting for sex and PC1-5 for population substructure was employed. Age was not adjusted in the regression model with following considerations: first, more than half of the PSP cases have age missing (Table 1); second, controls (average age, 81) was much older compared to PSP (average age, 63) and well over the mean age-of-onset for PSP which is 63 years\(^{28}\). Individuals with the H2H2 genotype are imbalanced (5 cases, 138 controls) and with few cases, therefore, statistical analysis for this subgroup was not included.

To evaluate the association of the structural forms of 17q21.31 with PSP, each structural form with allele frequency $> 1\%$ is compared with the rest of structural forms using logistic regression model adjusting for sex and PC1-5. To evaluate the association of MAPT sub-haplotypes with PSP, each MAPT sub-haplotypes with allele frequency $> 1\%$ was compared with the rest of sub-haplotypes using logistic regression adjusting for sex and PC1-5.
Results

The copy numbers of α, β, and γ and PSP risk

The H1 and H2 haplotypes are the most prominent genetic risk factor for PSP. Individuals with the H2 haplotype showed significantly lower risk of disease (OR, 0.19; 95%CI, 0.16-0.22; \( P = 3.00 \times 10^{-79} \)). Structural forms of 17q21.31 can be characterized by three large duplications α, β, and γ. However, their specific contributions to the risk of PSP have not been carefully assessed. In this study, we first evaluated whether the copy numbers of α, β, or γ are associated to the risk of PSP independently from H1 and H2 (eFigure 7 in Supplement).

We found that each additional copy of γ was associated with 1.08 (95%CI, 1.02-1.15; \( P = 0.014 \)) fold of increased risk of PSP in H1H1 individuals and 1.29 (%95CI, 1.06-1.56; \( P = 0.0096 \)) fold of increased risk of PSP in H1H2 individuals (Table 2). The association of the copy number of γ with PSP in H2H2 individuals was not evaluated since there were only five H2H2 PSP samples. On average, γ was associated with 1.10 (95% CI, 1.04-1.17; \( P = 0.0018 \)) fold of increased risk of PSP after adjusting for sex, PC1-5, and allele count (0, 1, or 2) of the H2 haplotype (Table 2). Without adjusting for H2, the effect of γ would be neglected (OR, 0.98; 95%CI, 0.93-1.04; \( P = 0.60 \), Table 2) because the H2 haplotype has increased copies (usually two copies) of γ while displayed significantly lower risk of PSP.

Table 2. Association between the copy numbers of α, β, γ and risk of PSP

<table>
<thead>
<tr>
<th>Model: Sex, PC1-5</th>
<th>Model: Sex, PC1-5</th>
</tr>
</thead>
<tbody>
<tr>
<td>(H1H1 carriers, N=2,958)</td>
<td>(H1H2 carriers, N=975)</td>
</tr>
<tr>
<td>(PSP = 1,511; Control = 1,447)</td>
<td>(PSP = 168; Control = 807)</td>
</tr>
</tbody>
</table>

Table: Association between the copy numbers of α, β, γ and risk of PSP
### Table 2: Associations of Copy Numbers of α, β, and γ with PSP

<table>
<thead>
<tr>
<th></th>
<th>OR (95% CI)</th>
<th>P</th>
<th>OR (95% CI)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>α</td>
<td>0.91 (0.81-1.02)</td>
<td>0.11</td>
<td>0.81 (0.58-1.11)</td>
<td>0.20</td>
</tr>
<tr>
<td>β</td>
<td>0.91 (0.81-1.02)</td>
<td>0.11</td>
<td>0.79 (0.53-1.15)</td>
<td>0.23</td>
</tr>
<tr>
<td>γ</td>
<td>1.08 (1.02-1.15)</td>
<td>0.014*</td>
<td>1.29 (1.06-1.56)</td>
<td>0.0096*</td>
</tr>
</tbody>
</table>

**Model: Sex, PC1-5**  
(N = 4,076)  
(PSP = 1,684; Control = 2,392)

<table>
<thead>
<tr>
<th></th>
<th>OR (95% CI)</th>
<th>P</th>
<th>OR (95% CI)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>α</td>
<td>0.57 (0.52-0.63)</td>
<td>&lt; 2×10⁻¹⁶*</td>
<td>0.90 (0.81-1.00)</td>
<td>0.061</td>
</tr>
<tr>
<td>β</td>
<td>1.14 (1.03-1.27)</td>
<td>0.011*</td>
<td>0.90 (0.81-1.01)</td>
<td>0.064</td>
</tr>
<tr>
<td>γ</td>
<td>0.98 (0.93-1.04)</td>
<td>0.60</td>
<td>1.10 (1.04-1.17)</td>
<td>0.0018*</td>
</tr>
</tbody>
</table>

*Represents statistical significance (< 0.05)

We found no significant association for the copy numbers of α (OR, 0.9; 95%CI 0.81-1.00; P = 0.061) and β (OR, 0.9; 95%CI 0.81-1.01; P =0.064) with PSP when adjusting the allele count (0, 1, or 2) of H2 (Table 2); however, individuals with more copies of α and β showed slightly lower odds ratio for PSP. Similarly, no significance was observed for the copy numbers of α and β when association analysis was stratified by H1H1 or H1H2 genotype. Only under the regression model without adjusting H1 and H2, we observed statistically significant association for the copy numbers of α (P < 2 × 10⁻¹⁶) and β (P = 0.011) with PSP. However, the observed significance mainly arises from their correlation with the H1 and H2 haplotypes, i.e., the increased copies (usually two copies) of α and the absence of β duplication in the H2 haplotype.
Structural forms of 17q21.31 and PSP risk

We then evaluated whether the structural forms of 17q21.31, characterized by the α, β, and γ, show distinct effect on the risk of PSP. We found that the odds ratio for PSP increases from 1.21 (H1β1γ1, 95%CI 1.10-1.33) to 1.57 (H1β1γ4, 95%CI 1.10-2.26) as the copy number of γ increases from one copy to four copies (Table 3). With an additional copy of β, H1β2γ1 (OR, 1.24; 95%CI 1.11-1.38; $P = 1.87 \times 10^{-4}$) displayed similar risk compared to H1β1γ1 (OR, 1.21; 95%CI 1.10-1.33; $P = 5.47 \times 10^{-5}$) regarding PSP. This reaffirmed our finding that the copy number of γ was associated with increased risk of PSP independently from H1 and H2, and β was not associated with the risk of PSP (Table 2; eFigure 7 in Supplement). Besides, all structural forms with H1 background displayed increased risk of PSP and all structural forms with H2 background displayed decreased risk of PSP (Table 3).

Table 3. Structural forms of 17q21.31 and the risk of PSP

<table>
<thead>
<tr>
<th>Structural forms</th>
<th>Frequency (%)</th>
<th>Odds ratio</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1β1γ1</td>
<td>32.57</td>
<td>1.21 (1.10-1.33)</td>
<td>5.47 × 10^{-5}</td>
</tr>
<tr>
<td>H1β1γ2</td>
<td>21.94</td>
<td>1.29 (1.16-1.43)</td>
<td>1.35 × 10^{-6}</td>
</tr>
<tr>
<td>H1β1γ3</td>
<td>14.73</td>
<td>1.45 (1.27-1.65)</td>
<td>3.94 × 10^{-8}</td>
</tr>
<tr>
<td>H1β1γ4</td>
<td>1.93</td>
<td>1.57 (1.10-2.26)</td>
<td>1.35 × 10^{-2}</td>
</tr>
<tr>
<td>H1β2γ1</td>
<td>22.98</td>
<td>1.24 (1.11-1.38)</td>
<td>1.87 × 10^{-4}</td>
</tr>
<tr>
<td>H2α1γ2</td>
<td>0.42</td>
<td>0.23 (0.12-0.40)</td>
<td>5.94 × 10^{-7}</td>
</tr>
<tr>
<td>H2α2γ2</td>
<td>4.78</td>
<td>0.19 (0.16-0.23)</td>
<td>&lt; 2 × 10^{-16}</td>
</tr>
</tbody>
</table>

Haplotypes in less than 1% of individuals were excluded.

Odds ratio and P value were from logistic regression adjusting for PC1-5 and Sex.
Structural forms of 17q21.31 and MAPT sub-haplotypes

According to the LD structure in MAPT gene (~150 Kb), 26 MAPT sub-haplotypes can be determined by six haplotype-tagging SNVs\textsuperscript{11–13} (eTable 2 in Supplement). We confirmed the association of H1c (OR 1.79; 95%CI 1.58-2.04; \( P = 1.84 \times 10^{-19} \)), H1d (OR 1.52; 95%CI 1.29-1.79; \( P = 3.89 \times 10^{-7} \)), and H1o (OR 2.88; 95%CI 2.15-3.89; \( P = 2.77 \times 10^{-12} \)) with PSP (Table 4). Moreover, H1g (OR 1.46; 95%CI 1.07-1.98; \( P = 0.016 \)), which was significant, and H1h (OR 1.36; 95%CI 1.10-1.69; \( P = 0.0053 \)), which was not significant in the previous study\textsuperscript{11}, were both nominal significant in our analysis.

Table 4. MAPT sub-haplotypes defined by six SNVs and the risk of PSP

<table>
<thead>
<tr>
<th>Haplotype</th>
<th>Sub-haplotype Frequency (%)</th>
<th>PSP (N = 1,684)</th>
<th>Control (N = 2,392)</th>
<th>OR</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1b</td>
<td>16.92</td>
<td>16.99</td>
<td>1.01 (0.90-1.14)</td>
<td>0.84</td>
<td></td>
</tr>
<tr>
<td>H1c**</td>
<td>19.6</td>
<td>11.98</td>
<td>1.79 (1.58-2.04)</td>
<td>( 1.84 \times 10^{-19} )</td>
<td></td>
</tr>
<tr>
<td>H1d**</td>
<td>10.39</td>
<td>7.09</td>
<td>1.52 (1.29-1.79)</td>
<td>( 3.89 \times 10^{-7} )</td>
<td></td>
</tr>
<tr>
<td>H1e</td>
<td>8.76</td>
<td>8.07</td>
<td>1.12 (0.95-1.32)</td>
<td>0.17</td>
<td></td>
</tr>
<tr>
<td>H1g*</td>
<td>2.67</td>
<td>1.80</td>
<td>1.46 (1.07-1.98)</td>
<td>0.016</td>
<td></td>
</tr>
<tr>
<td>H1h*</td>
<td>5.23</td>
<td>4.03</td>
<td>1.36 (1.10-1.69)</td>
<td>0.0053</td>
<td></td>
</tr>
<tr>
<td>H1i</td>
<td>4.69</td>
<td>4.08</td>
<td>1.17 (0.94-1.46)</td>
<td>0.15</td>
<td></td>
</tr>
<tr>
<td>H1j</td>
<td>1.28</td>
<td>1.21</td>
<td>1.09 (0.72-1.64)</td>
<td>0.68</td>
<td></td>
</tr>
<tr>
<td>H1l</td>
<td>2.91</td>
<td>3.22</td>
<td>0.92 (0.71-1.20)</td>
<td>0.55</td>
<td></td>
</tr>
<tr>
<td>H1m</td>
<td>2.11</td>
<td>2.07</td>
<td>1.08 (0.78-1.47)</td>
<td>0.64</td>
<td></td>
</tr>
<tr>
<td>H1o**</td>
<td>4.13</td>
<td>1.53</td>
<td>2.88 (2.15-3.89)</td>
<td>( 2.77 \times 10^{-12} )</td>
<td></td>
</tr>
<tr>
<td>H1p</td>
<td>1.07</td>
<td>1.36</td>
<td>0.80 (0.52-1.21)</td>
<td>0.30</td>
<td></td>
</tr>
<tr>
<td>H1q</td>
<td>1.22</td>
<td>0.86</td>
<td>1.45 (0.92-2.26)</td>
<td>0.11</td>
<td></td>
</tr>
<tr>
<td>H1r</td>
<td>1.16</td>
<td>1.30</td>
<td>0.92 (0.60-1.39)</td>
<td>0.69</td>
<td></td>
</tr>
<tr>
<td>H1u</td>
<td>2.64</td>
<td>2.34</td>
<td>1.16 (0.86-1.54)</td>
<td>0.33</td>
<td></td>
</tr>
<tr>
<td>H1x</td>
<td>1.43</td>
<td>1.36</td>
<td>0.98 (0.67-1.44)</td>
<td>0.93</td>
<td></td>
</tr>
<tr>
<td>H1y</td>
<td>1.54</td>
<td>1.51</td>
<td>1.01 (0.70-1.44)</td>
<td>0.98</td>
<td></td>
</tr>
<tr>
<td>H1z</td>
<td>1.40</td>
<td>1.00</td>
<td>1.41 (0.93-2.14)</td>
<td>0.11</td>
<td></td>
</tr>
</tbody>
</table>

MAPT sub-haplotypes in less than 1% of individuals were excluded.
*Represents MAPT sub-haplotypes with a $P \leq 0.05$ (nominal significant)

**Represents MAPT sub-haplotypes with a $P \leq 0.0028$ (Bonferroni corrected $P$ cutoff for the 18 test performed)

We examined the relationship between the structural forms of 17q21.31 (~1.5 Mb) and MAPT sub-haplotypes (~150 Kb) through LD analysis. We identified two pairs with $R^2 > 0.1$ (eTable 4 in Supplement): H1β1γ3 and H1c ($R^2 = 0.31$) and H1β2γ1 and H1b ($R^2 = 0.29$). This was confirmed by the fact that 70% of H1β1γ3 are H1c and 56% of H1β2γ1 are H1b (Fig. 1A). Moreover, with additional copies of γ in the H1, the proportion of MAPT sub-haplotypes with increased risk of PSP (i.e., H1c, H1d, H1g, H1h, and H1o) increases from 34% (H1β1γ1) to 77% (H1β1γ4) (Fig. 1A). In contrast, the proportion of MAPT sub-haplotypes with increased risk of PSP drops from 34% (H1β1γ1) to 11% (H1β2γ1) with an additional copy of β (Fig. 1A). In individuals with H1H2 genotypes, no phasing is needed before comparison, therefore, the association between the structural forms of 17q21.31 and MAPT sub-haplotypes can be observed directly without phasing (Fig. 1B).
Fig. 1: Structural forms of 17q21.31 and MAPT sub-haplotypes

Structural forms of 17q.21.31 in less than 1% of individuals were excluded. Structural forms with H2 background are also excluded since there is only one MAPT sub-haplotype for H2. The number above each bar is the total number of haplotypes for the specific structural form of 17q21.31 and the percentage that are also MAPT sub-haplotypes (H1o, H1c, H1d, H1g, H1h) associated with increased risk of PSP. Besides, the percentage of H1β1γ3 that is H1c and the percentage of of H1β2γ1 that is H1b are displayed. A. In phased haplotypes from all samples, the proportion of MAPT sub-haplotypes in each structural form of 17q21.31. B. In H1H2 individuals, the proportion of MAPT sub-haplotypes in each structural form of 17q21.31.

Discussion

The H1 and H2 haplotypes on 17q21.31 is the strongest genetic risk factor for PSP\(^7\). The MAPT gene inside H1 and H2 showed haplotypic-specific expression and was considered as the possible cause
of the association\textsuperscript{29,30}. Therefore, previous studies have been focused on variants and haplotypic structures inside \textit{MAPT} and identified sub-haplotypes H1c, H1d, H1o, and H1g, which displayed increased risk of PSP\textsuperscript{11,13}. In this study, we went beyond \textit{MAPT} gene (~150 Kb) and evaluated the association of the structural forms of 17q21.31 (~1.5 Mb), characterized by large duplications \(\alpha\), \(\beta\), and \(\gamma\), with the risk PSP. We found that the copy number of \(\gamma\) duplication was associated with increased risk of PSP and structural forms with \(\gamma\) duplication (such as H1\(\beta\)1\(\gamma\)2, H1\(\beta\)1\(\gamma\)3, and H1\(\beta\)1\(\gamma\)4) had a higher odds ratio for PSP compared to H1\(\beta\)1\(\gamma\)1. This is in accordance with the fact that structural forms with additional copies of \(\gamma\) (from H1\(\beta\)1\(\gamma\)1 to H1\(\beta\)1\(\gamma\)4) tended to have more \textit{MAPT} sub-haplotypes with a significant risk to PSP (i.e., H1c, H1d, H1o, H1g, H1h). Overall, this study provided a first analysis on the association of structural forms of 17q21.31 with the risk of PSP and \textit{MAPT} sub-haplotypes.

It should be noted that the copy numbers of \(\alpha\), \(\beta\), and \(\gamma\) are correlated with H1 and H2. When performing the association of the copy numbers of \(\alpha\), \(\beta\), and \(\gamma\) with PSP in individuals with European ancestry, previous studies demonstrated that the 205-Kb \(\beta\) region can only duplicate in the H1 haplotype, the smaller 155-Kb \(\alpha\) region but not the entire \(\beta\) region duplicates in the H2 haplotype, and the 210-Kb \(\gamma\) region usually duplicates only once in the H2 haplotype (eFigure 1 in \textbf{Supplement})\textsuperscript{15,31}. Therefore, it is important to adjust the status of H1 and H2 and split analysis by individuals with H1H1 and H1H2 genotypes when performing the association of \(\alpha\), \(\beta\), and \(\gamma\) with PSP. In terms of genomic position, the \(\alpha\) region is included in \(\beta\) region, and parts of \(\beta\) and \(\gamma\) are overlapping with each other (eFigure 1 in \textbf{Supplement}). When performing the association of \(\alpha\) with PSP, we considered whenever there was a \(\beta\) duplication there was also a \(\alpha\) duplication as \(\alpha\) region is included in \(\beta\). Considering these factors, we
found that only \( \gamma \) is independently associated with risk of PSP, and the associations of \( \alpha \) and \( \beta \) with PSP are due to their correlation with H1 and H2.

The influence of structural forms of 17q21.31 on the risk of PSP has never been examined before. In this study, we found that all structural forms with H1 background were association with increased with odd of PSP. Particularly, H1 with additional \( \gamma \) copies (H1\( \beta \)1\( \gamma \)2, H1\( \beta \)1\( \gamma \)3, and H1\( \beta \)1\( \gamma \)4) displayed a higher odds ratio for PSP compared to H1\( \beta \)1\( \gamma \)1. In previous studies of MAPT sub-haplotypes, only a few sub-haplotypes (i.e., H1c, H1d, H1o, and H1g) were associated with increased risk of PSP while other sub-haplotypes displayed no differences against the rest of the population\textsuperscript{11,13}. We replicated the association of H1c, H1d, H1o and H1g with PSP and identified H1h with nominal significance. Furthermore, we showed that the H1 haplotype with additional \( \gamma \) copies had more MAPT sub-haplotypes that have higher risk of PSP (H1c, H1d, H1o, H1g and H1h), the proportion of MAPT sub-haplotypes associated with PSP increased from 34\% in H1\( \beta \)1\( \gamma \)1 to 77\% in H1\( \beta \)1\( \gamma \)4.

**Limitations**

Despite our finding of the importance of the copy number of \( \gamma \) duplication in the risk of PSP, there are a few limitations in the current study. First, all our PSP samples are of European ancestry. It is important to collect samples from other ancestries to validate the discovery in a genetically diverse background. Second, not all PSP cases are pathologically confirmed: of the 1,684 individuals with PSP, 1,386 were autopsy confirmed. Finally, the association analysis needed to be replicated in an independent dataset. Currently, the available resource for whole genome sequencing data in PSP is limited. Therefore, future study focusing on \( \alpha \), \( \beta \), and \( \gamma \) regions using CNV array to replicate the findings
in a large independent cohorts might be helpful.

Conclusions

In this study of 4076 subjects (N_{PSP} = 1,684, N_{control} = 2,392), we found that \( \gamma \) duplication was associated with increased risk (OR, 1.10; 95% CI, 1.04-1.17; \( P = 0.0018 \)) of PSP. In accordance with this finding, the structural forms with additional copies of \( \gamma \) (H1\( \beta \)1\( \gamma \)2 [OR, 1.29; 95%CI 1.16-1.43; \( P = 1.35 \times 10^{-6} \)], H1\( \beta \)1\( \gamma \)3 [OR, 1.45; 95%CI 1.27-1.65; \( P = 3.94 \times 10^{-9} \)], H1\( \beta \)1\( \gamma \)4 [OR, 1.57; 95%CI 1.10-2.26; \( P = 1.35 \times 10^{-2} \)]) showed higher odds ratio for PSP compared to H1\( \beta \)1\( \gamma \)1 (OR, 1.21; 95% CI 1.10-1.33; \( P = 5.47 \times 10^{-5} \)) and had higher proportion of MAPT sub-haplotypes associated with increased risk of PSP (i.e., H1c, H1d, H1g, H1o, and H1g).

Declarations

Ethics approval and consent to participate

Consent for publication

Not applicable.

Availability of data and materials

NIAGADS Data Sharing Service (https://dss.niagads.org/)

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, Al 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.

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. YQS, AT, and JYT are supported by RF1-AG074328. 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.
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 the Clinical 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 Lorenz, Irene Litvan, Erik D Roberson, Günter U Högl, 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' information

Authors' contribution

Study design: TSC, DD, GUH, JYT, DHG, GDS, 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, DHG, GDS, and WPL. Variant detection and variant quality check: HW, TSC, VP, LVB, KF, YYL, and WPL. Statistical analyses and
interpretation of results: HW, YQS, AT, TSC, KF, AN, JYT, DHG, GDS, and WPL. Experimental validation: BAD and PLC. Draft of the manuscript: HW, GDS, and WPL. All authors read, critically revised, and approved the manuscript.
References


association analysis of the tau gene in progressive supranuclear palsy and corticobasal degeneration. 


26. Lee WP, Choi SH, Shea MG, 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. Published online 2023:2023-09.


