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

Hui Wang[1,2], PhD; Timothy S Chang[3], MD, PhD; Beth A Dombroski[1,2], PhD; Po-Liang Cheng[1,2], PhD; Ya-Qin Si[4], PhD; Albert Tucci[4], PhD; Vishakha Patil[5], MS; Leopoldo Valiente-Banuet[5], MS; Kurt Farrell[6], PhD; Catriona Mclean[7], MD; Laura Molina-Porcel[8,9], MD, PhD; Alex Rajput[10], MD; Peter Paul De Deyn[11,12], MD, PhD; Nathalie Le Bastard[13], PhD; Marla Gearing[14], PhD; Laura Donker Kaat[15], MD, PhD; John C Van Swieten[15], MD, PhD; Elise Dopper[15], MD, PhD; Bernardino F Ghetti[16], MD; Kathy L Newell[16], MD; Claire Troakes[17], PhD; Justo G de Yébenes[18], MD, PhD; Alberto Rábano-Gutierrez[19], MD, PhD; Tina Meller[20], PhD; Wolfgang H Oertel[20], MD, PhD; Gesine Respondek[21], MD; Maria Stamelou[22,23], MD; Thomas Arzberger[24,25], MD; Sigrun Roeber[26], MD; Ulrich Müller[26], MD; Franziska Hopfner[44], MD; Pau Pastor[27,28], MD, PhD; Alexis Brice[29], MD; Alexandra Durr[29], MD, PhD; Isabelle Le Ber[29], MD, PhD; Thomas G Beach[30], MD, PhD; Geidy E Serrano[30], PhD; Lili-Naz Hazrati[31], MD, PhD; Irene Litvan[32], MD; Rosa Rademakers[33,34], PhD; Owen A Ross[34], PhD; Douglas Galasko[32], MD; Adam L Boxer[35], MD, PhD; Bruce L Miller[35], MD; Willian W Seeley[35], MD; Vivianna M Van Deerlin[1], MD, PhD; Edward B Lee[1,36], MD, PhD; Charles L White III[37], MD; Huw R Morris[38], MD, PhD; Rohan de Silva[39], PhD; John F Crary[6], MD, PhD; Alison M Goate[40], PhD; Jeffrey S Friedman[41], MD, PhD; Yuk Yee Leung[1,2], PhD; Giovanni Coppola[5,42], MD; Adam C Naj[1,2,43], PhD; Li-San Wang[1,2], PhD; PSP genetics study group; Dennis W Dickson[34], MD; Günter U Höglinger[44], MD; Jung-Ying Tzeng[4,45], PhD; Daniel H Geschwind[5,46,47], MD, PhD; Gerard D Schellenberg[1,2], PhD; Wan-Ping Lee[1,2], PhD

[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]Bioinformatics Research Center, North Carolina State University, NC, USA
[5]Movement Disorders Programs, Department of Neurology, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA

[6]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.
[7]Victorian Brain Bank, The Florey Institute of Neuroscience and Mental Health, Parkville, Victoria, Australia
[8]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
[9]Neurological Tissue Bank of the Biobanc-Hospital Clínic-IDIBAPS, Barcelona, Spain
[10]Movement Disorders Program, Division of Neurology, University of Saskatchewan, Saskatoon, Saskatchewan, Canada
[11]Laboratory of Neurochemistry and Behavior, Experimental Neurobiology Unit, University of Antwerp, Wilrijk (Antwerp), Belgium
[12]Department of Neurology, University Medical Center Groningen, NL-9713 AV Groningen, Netherlands
[13]Fujirebio Europe NV, Technologiepark 6, 9052 Gent, Belgium
[14]Department of Pathology and Laboratory Medicine and Department of Neurology, Emory University School of Medicine, Atlanta, GA, USA
[15]Netherlands Brain Bank and Erasmus University, Netherlands
[16]Department of Pathology and Laboratory Medicine, Indiana University School of Medicine, Indianapolis, IN, USA
[17]London Neurodegenerative Diseases Brain Bank, King's College London, London, UK
[18]Autonomous University of Madrid, Madrid, Spain
[19]Fundación CIEN (Centro de Investigación de Enfermedades Neurológicas) - Centro Alzheimer Fundación Reina Sofía, Madrid, Spain
[20]Department of Neurology, Philipps-Universität, Marburg, Germany
[21]German Center for Neurodegenerative Diseases (DZNE), Munich, Germany
[22]Parkinson's disease and Movement Disorders Department, HYGEIA Hospital, Athens, Greece
[23]European University of Cyprus, Nicosia, Cyprus
[24]Department of Psychiatry and Psychotherapy, University Hospital Munich, Ludwig-Maximilians-University Munich, Germany
[25]Center for Neuropathology and Prion Research, Ludwig-Maximilians-University Munich, Germany
[26]German Brain Bank, Neurobiobank Munich, Germany
[27]Unit of Neurodegenerative diseases, Department of Neurology, University Hospital Germans Trias i Pujol, Badalona, Barcelona, Spain
[28]Neurosciences, The Germans Trias i Pujol Research Institute (IGTP) Badalona, Badalona, Spain
[29]Sorbonne Université, Paris Brain Institute – Institut du Cerveau – ICM, Inserm U1127, CNRS UMR 7225, APHP - Hôpital Pitié-Salpêtrière, Paris, France

[30]Banner Sun Health Research Institute, Sun City, AZ, USA
[31]University McGill, Montreal, Quebec, Canada
[32]Department of Neuroscience, University of California, San Diego, CA, USA
[33]VIB Center for Molecular Neurology, University of Antwerp, Belgium
[34]Department of Neuroscience, Mayo Clinic Jacksonville, FL, USA
[35]Memory and Aging Center, University of California, San Francisco, CA, USA
[36]Center for Neurodegenerative Disease Research, University of Pennsylvania School of Medicine, Philadelphia, PA, USA
[37]University of Texas Southwestern Medical Center, Dallas, TX, USA
[38]Departmento of Clinical and Movement Neuroscience, University College of London, London, UK
[39]Reta Lila Weston Institute, UCL Queen Square Institute of Neurology, London, UK.
[40]Department of Genetics and Genomic Sciences, New York, NY, USA; Icahn School of Medicine at Mount Sinai, New York, NY, USA
[41]Friedman Bioventure, Inc., Del Mar, CA, USA
Department of Genetics and Genomic Sciences, New York, NY, USA
[42]Department of Psychiatry, Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles, CA, USA
[43]Department of Biostatistics, Epidemiology, and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
[44]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
[45]Department of Statistics, North Carolina State University, NC, USA
[46]Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA
[47]Institute of Precision Health, University of California, Los Angeles, Los Angeles, CA, USA

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

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 17q.21.31, 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

176 measures of severity of tau pathology in PSP cases[11]. However, these sub-haplotypes were inferred from
177 LD structure limited to *MAPT* gene (~150 Kb) and do not represent the intricacies of structural forms on
178 17q21.31 (~1.5 Mb). To address this limitation, we constructed the ten structural forms of 17q21.31,
179 characterized by the copy numbers of α, β, and γ (150Kb, 300Kb, and 218Kb; eFigure 1 in
180 **Supplement**)[14,15], and evaluated their association with the risk of PSP in this study. We found that the
181 H1 haplotype with more copies γ duplication showed an increased risk of PSP. The result is in
182 accordance with the higher portion of H1c and other *MAPT* sub-haplotypes associated with increased
183 risk of PSP in the H1 haplotypes with more copies of γ. Through this study, we gained a better
184 understanding how the haplotypic structure of 17q.21.31 correlates with the risk of PSP.
185
186 **Methods**
187 *Study subjects and quality control*
188 All study subjects and WGS data are available on The National Institute on Aging Genetics of
189 Alzheimer's Disease Data Storage Site (NIAGADS)[16] under Alzheimer's Disease Sequencing Project
190 (ADSP) Umbrella NG00067.v7[17]. We inferred ancestry of subjects by GRAF-pop[18] and selected 4,618
191 subjects (1,797 cases and 2,821 controls) of European ancestry for analysis. WGS were performed at
192 30x coverage (e**Table 1**).
193 Among 4,618 samples, we filtered 183 samples with abnormally low reads mapped (aligned read
194 depth < 1.7x) to α, β or γ region (eFigure 2 in **Supplement**) and ten samples with high genotyping
195 missing rate (> 0.05). Next, 244 related samples inferred by KING[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*[7] 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[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 α, β and γ, 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 α, β and γ. This discordance might be due to subjects carrying undiscovered structural forms or genotyping errors on the copy numbers of α, β and γ. 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.

207 **Table 1. Characteristics of PSP cases and controls**

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

208 [a]1,130 PSP cases and 111 controls have missing age.

209 [b]H1/H2 status was determined by the genotype of a 238-bp H2 tagging deletion[7].

210 [c]Structural 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 studies[14,15] (eFigure 1 in **Supplement**). Segmental duplications can introduce mapping challenges and thus inaccurate calling of the number of copies[21–23]. 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-means[25] 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: TTAGTCAATTTCTTAGCCAACCCAT; chr17:46171656-46171680, GRCh38) was determined using a predesigned TaqMan CNV assay (Applied Biosystems Part No. Hs01788222_cn).

Copy number at the γ region (Context Sequence: AGAAAAAAAGCATTGACTCCAACCC; chr17:46419867-46419891 and chr17:46637445-46637469, GRCh38) was determined using a custom designed TaqMan CNV assay (Applied Biosystems; Forward Primer: TGGCACAATGACCATCGAGATT, Reverse Primer: CTGCCATCTTGTCGGTGTCA, 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 α, β and γ. 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 α, β, and γ. 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)[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[26]. Variants were removed if they were monomorphic, did not pass variant quality score recalibration, had an average read depth ≥ 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[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 α, β and γ as multi-allelic CNVs by a series of surrogate bi-allelic markers with 0/1 alleles[14] (eTable 3 in **Supplement**). Then, SHAPEIT4[27] (Version 4.2.2) were used for phasing the copy numbers of α, β and γ together with SNVs/indels. SNVs and indels insides α, β, and γ 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_{PSP} = 1,684, $N_{control}$ = 2,392). For the association of the copy numbers of α, β, and γ with PSP, logistic regression adjusting for sex and PC1-5 for population substracture 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[14,15]. 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

Model: Sex, PC1-5 (H1H1 carriers, N=2,958 (PSP = 1,511; Control = 1,447)	**Model: Sex, PC1-5 (H1H2 carriers, N=975) (PSP = 168; Control = 807)**

	OR (95% CI)	P	OR (95% CI)	P
α	0.91 (0.81-1.02)	0.11	0.81 (0.58-1.11)	0.20
β	0.91 (0.81-1.02)	0.11	0.79 (0.53-1.15)	0.23
γ	1.08 (1.02-1.15)	0.014*	1.29 (1.06-1.56)	0.0096*
	Model: Sex, PC1-5 (N = 4,076) (PSP = 1,684; Control = 2,392)		**Model: Sex, PC1-5, H2 (N = 4,076) (PSP = 1,684; Control = 2,392)**	
	OR (95% CI)	**P**	**OR (95% CI)**	**P**
α	0.57 (0.52-0.63)	$< 2\times10^{-16}$*	0.90 (0.81-1.00)	0.061
β	1.14 (1.03-1.27)	0.011*	0.90 (0.81-1.01)	0.064
γ	0.98 (0.93-1.04)	0.60	1.10 (1.04-1.17)	0.0018*

*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 ($P < 2 \times 10^{-16}$) 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

	Frequency (%)			
Structural forms	**PSP (N = 1,684)**	**Control (N = 2,392)**	**Odds ratio**	**P**
H1β1γ1	32.57	28.20	1.21 (1.10-1.33)	5.47×10^{-5}
H1β1γ2	21.94	16.99	1.29 (1.16-1.43)	1.35×10^{-6}
H1β1γ3	14.73	10.26	1.45 (1.27-1.65)	3.94×10^{-8}
H1β1γ4	1.93	1.28	1.57 (1.10-2.26)	1.35×10^{-2}
H1β2γ1	22.98	19.69	1.24 (1.11-1.38)	1.87×10^{-4}
H2α1γ2	0.42	1.78	0.23 (0.12-0.40)	5.94×10^{-7}
H2α2γ2	4.78	20.09	0.19 (0.16-0.23)	$< 2 \times 10^{-16}$

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[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[11], were both nominal significant in our analysis.

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

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

MAPT sub-haplotypes in less than 1% of individuals were excluded.

*Represents *MAPT* sub-haplotypes with a $P <= 0.05$ (nominal significant)
**Represents *MAPT* sub-haplotypes with a $P <= 0.0028$ (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**).

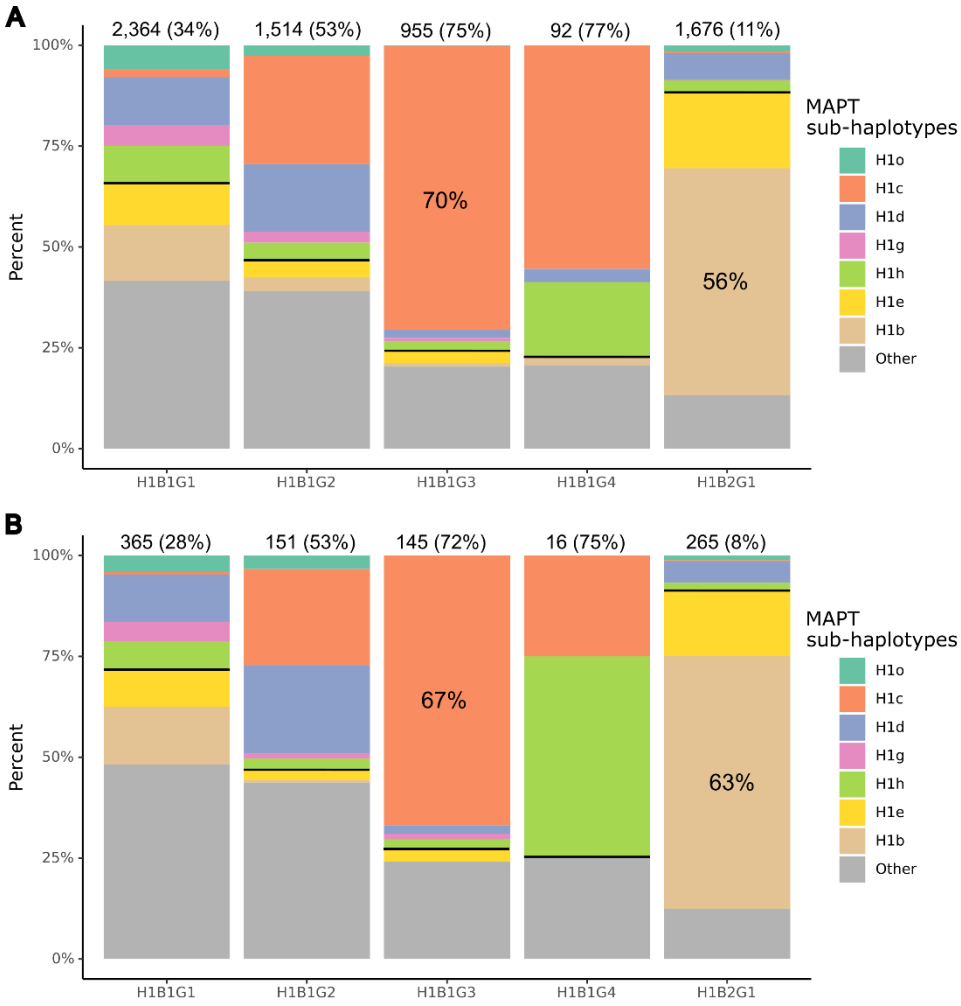

Fig. 1: Structural forms of 17q21.31 and *MAPT* sub-haplotypes

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-haplotypes 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[29,30]. Therefore, previous studies have been focused on variants and haplotypic structures inside *MAPT* and identified sub-haplotypes H1c, H1d, H1o, and H1g, which displayed increased risk of PSP[11,13]. In this study, we went beyond *MAPT* gene (~150 Kb) and evaluated the association of the structural forms of 17q21.31 (~1.5 Mb), characterized by large duplications α, β, and γ, with the risk PSP. We found that the copy number of γ duplication was associated with increased risk of PSP and structural forms with γ duplication (such as H1β1γ2, H1β1γ3, and H1β1γ4) had a higher odds ratio for PSP compared to H1β1γ1. This is in accordance with the fact that structural forms with additional copies of γ (from H1β1γ1 to H1β1γ4) tended to have more *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 *MAPT* sub-haplotypes.

It should be noted that the copy numbers of α, β, and γ are correlated with H1 and H2. When performing the association of the copy numbers of α, β, and γ with PSP in individuals with European ancestry, previous studies demonstrated that the 205-Kb β region can only duplicate in the H1 haplotype, the smaller 155-Kb α region but not the entire β region duplicates in the H2 haplotype, and the 210-Kb γ region usually duplicates only once in the H2 haplotype (eFigure 1 in **Supplement**)[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 α, β, and γ with PSP. In terms of genomic position, the α region is included in β region, and parts of β and γ are overlapping with each other (eFigure 1 in **Supplement**). When performing the association of α with PSP, we considered whenever there was a β duplication there was also a α duplication as α region is included in β. Considering these factors, we

found that only γ is independently associated with risk of PSP, and the associations of α and β 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 of PSP. Particularly, H1 with additional γ copies (H1β1γ2, H1β1γ3, and H1β1γ4) displayed a higher odds ratio for PSP compared to H1β1γ1. 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[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 γ 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 H1β1γ1 to 77% in H1β1γ4.

Limitations

Despite our finding of the importance of the copy number of γ 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 α, β, and γ 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 γ 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 γ (H1β1γ2 [OR, 1.29; 95%CI 1.16-1.43; *P* = 1.35×10^{-6}], H1β1γ3 [OR, 1.45; 95%CI 1.27-1.65; P = 3.94×10^{-8}], H1β1γ4[OR, 1.57; 95%CI 1.10-2.26; P = 1.35×10^{-2}]) showed higher odds ratio for PSP compared to H1β1γ1 (OR, 1.21; 95%CI 1.10-1.33; *P* = 5.47×10^{-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, 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.

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

454 supported by U19AG063911, R01AG073482, R01AG038791, and R01AG071756. BLM is supported
455 by P01 AG019724, R01 AG057234 and P0544014. VMV is supported by P01-AG-066597,
456 P01-AG-017586. HRM is supported by CurePSP, PSPA, MRC, and Michael J Fox Foundation. RDS is
457 supported by CurePSP, PSPA, and Reta Lila Weston Trust. JFC is supported by R01 AG054008, R01
458 NS095252, R01 AG060961, R01 NS086736, R01 AG062348, P30 AG066514, the Rainwater Charitable
459 Foundation / Tau Consortium, Karen Strauss Cook Research, and Scholar Award, Stuart Katz & Dr. Jane
460 Martin. AMG is supported by the Tau Consortium and U54-NS123746. YYL is supported by
461 U54-AG052427; U24-AG041689. LSW is supported by U01AG032984, U54AG052427, and
462 U24AG041689. GUH was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research
463 Foundation) under Germany's Excellence Strategy within the framework of the Munich Cluster for
464 Systems Neurology (EXC 2145 SyNergy - ID 390857198); Deutsche Forschungsgemeinschaft (DFG,
465 HO2402/18-1 MSAomics); German Federal Ministry of Education and Research (BMBF, 01KU1403A
466 EpiPD; 01EK1605A HitTau; 01DH18025 TauTherapy). DHG is supported by 3UH3NS104095, Tau
467 Consortium. WPL is supported by RF1-AG074328; P30-AG072979; U54-AG052427; U24-AG041689.
468 Cases from Banner Sun Health Research Institute were supported by the NIH (U24 NS072026, P30
469 AG19610 and P30AG072980), the Arizona Department of Health Services (contract 211002, Arizona
470 Alzheimer's Research Center), the Arizona Biomedical Research Commission (contracts 4001, 0011,
471 05-901 and 1001 to the Arizona Parkinson's Disease Consortium) and the Michael J. Fox Foundation for
472 Parkinson's Research. The Mayo Clinic Brain Bank is supported through funding by NIA grants P50
473 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.

494 The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of
495 the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and
496 NINDS. The data used for the analyses described in this manuscript were obtained from:
497 https://gtexportal.org/home/datasets the GTEx Portal on 1/27/2022. We also thank to Drs. Murray
498 Grossman and Hans Kretzschmar for their valuable contribution to this work.

499 AMP-AD (sa000011) data: Mayo RNAseq Study- Study data were provided by the following sources:
500 The Mayo Clinic Alzheimer's Disease Genetic Studies, led by Dr. Nilufer Ertekin-Taner and Dr. Steven
501 G. Younkin, Mayo Clinic, Jacksonville, FL using samples from the Mayo Clinic Study of Aging, the
502 Mayo Clinic Alzheimer's Disease Research Center, and the Mayo Clinic Brain Bank. Data collection
503 was supported through funding by NIA grants P50 AG016574, R01 AG032990, U01 AG046139, R01
504 AG018023, U01 AG006576, U01 AG006786, R01 AG025711, R01 AG017216, R01 AG003949,
505 NINDS grant R01 NS080820, CurePSP Foundation, and support from Mayo Foundation. Study data
506 includes samples collected through the Sun Health Research Institute Brain and Body Donation Program
507 of Sun City, Arizona. The Brain and Body Donation Program is supported by the National Institute of
508 Neurological Disorders and Stroke (U24 NS072026 National Brain and Tissue Resource for Parkinson's
509 Disease and Related Disorders), the National Institute on Aging (P30 AG19610 Arizona Alzheimer's
510 Disease Core Center), the Arizona Department of Health Services (contract 211002, Arizona Alzheimer's
511 Research Center), the Arizona Biomedical Research Commission (contracts 4001, 0011, 05-901 and
512 1001 to the Arizona Parkinson's Disease Consortium) and the Michael J. Fox Foundation for Parkinson's
513 Research.

514 PSP-NIH-CurePSP-Tau (sa000015) data: This project was funded by the NIH grant UG3NS104095 and
515 supported by grants U54NS100693 and U54AG052427. Queen Square Brain Bank is supported by the
516 Reta Lila Weston Institute for Neurological Studies and the Medical Research Council UK. The Mayo
517 Clinic Florida had support from a Morris K. Udall Parkinson's Disease Research Center of Excellence
518 (NINDS P50 #NS072187), CurePSP and the Tau Consortium. The samples from the University of
519 Pennsylvania are supported by NIA grant P01AG017586.
520 PSP-CurePSP-Tau (sa000016) data: This project was funded by the Tau Consortium, Rainwater
521 Charitable Foundation, and CurePSP. It was also supported by NINDS grant U54NS100693 and NIA
522 grants U54NS100693 and U54AG052427. Queen Square Brain Bank is supported by the Reta Lila
523 Weston Institute for Neurological Studies and the Medical Research Council UK. The Mayo Clinic
524 Florida had support from a Morris K. Udall Parkinson's Disease Research Center of Excellence (NINDS
525 P50 #NS072187), CurePSP and the Tau Consortium. The samples from the University of Pennsylvania
526 are supported by NIA grant P01AG017586. Tissues were received from the Victorian Brain Bank,
527 supported by The Florey Institute of Neuroscience and Mental Health, The Alfred and the Victorian
528 Forensic Institute of Medicine and funded in part by Parkinson’s Victoria and MND Victoria. We are
529 grateful to the Sun Health Research Institute Brain and Body Donation Program of Sun City, Arizona for
530 the provision of human biological materials (or specific description, e.g. brain tissue, cerebrospinal
531 fluid). The Brain and Body Donation Program is supported by the National Institute of Neurological
532 Disorders and Stroke (U24 NS072026 National Brain and Tissue Resource for Parkinson's Disease and
533 Related Disorders), the National Institute on Aging (P30 AG19610 Arizona Alzheimer's Disease Core

534 Center), the Arizona Department of Health Services (contract 211002, Arizona Alzheimer's Research
535 Center), the Arizona Biomedical Research Commission (contracts 4001, 0011, 05-901 and 1001 to the
536 Arizona Parkinson's Disease Consortium) and the Michael J. Fox Foundation for Parkinson's Research.
537 Biomaterial was provided by the Study Group DESCRIBE of theClinical Research of the German
538 Center for Neurodegenerative Diseases (DZNE).
539 PSP_UCLA (sa000017) data: Thank to the AL-108-231 investigators, Adam L Boxer, Anthony E Lang,
540 Murray Grossman, David S Knopman, Bruce L Miller, Lon S Schneider, Rachelle S Doody, Andrew
541 Lees, Lawrence I Golbe, David R Williams, Jean-Cristophe Corvol, Albert Ludolph, David Burn,
542 Stefan Lorenzl, Irene Litvan, Erik D Roberson, Günter U Höglinger, Mary Koestler, Cliff ord R Jack Jr,
543 Viviana Van Deerlin, Christopher Randolph, Iryna V Lobach, Hilary W Heuer, Illana Gozes, Lesley
544 Parker, Steve Whitaker, Joe Hirman, Alistair J Stewart, Michael Gold, and Bruce H Morimoto.

545 *Authors' information*

546

547 *Authors' contribution*

548 Study design: TSC, DD, GUH, JYT, DHG, GDS, and WPL. Sample collection, brain biospecimens,
549 and neuropathological examinations: TSC, CM, LM, AR, PPDD, NLB, MG, LDK, JCVS, ED, BFG,
550 KLN, CT, JGdY, ARG, TM, WHO, GR, UM, FH, TA, SR, PP, AB, AD, ILB, TGC, GES, LNH, IL, RR,
551 OR, DG, ALB, BLM, WWS, VMVD, EBL, CLW, HM, JH, RdS, JFC, AMG, GC, and DHG. Genotype
552 or phenotype acquisition: HW, TSC, VP, LVB, KF, AN, LSW, DHG, GDS, and WPL. Variant detection
553 and variant quality check: HW, TSC, VP, LVB, KF, YYL, and WPL. Statistical analyses and

554 interpretation of results: HW, YQS, AT, TSC, KF, AN, JYT, DHG, GDS, and WPL. Experimental

555 validation: BAD and PLC. Draft of the manuscript: HW, GDS, and WPL. All authors read, critically

556 revised, and approved the manuscript.

References

1. Armstrong RA. Visual signs and symptoms of progressive supranuclear palsy. *Clin Exp Optom*. 2011;94(2):150-160. doi:10.1111/j.1444-0938.2010.00504.x

2. Lukic MJ, Respondek G, Kurz C, et al. Long-Duration Progressive Supranuclear Palsy: Clinical Course and Pathological Underpinnings. *Ann Neurol*. 2022;92(4):637-649.

3. Ali F, Martin PR, Botha H, et al. Sensitivity and specificity of diagnostic criteria for progressive supranuclear palsy. *Mov Disord*. 2019;34(8):1144-1153.

4. Hauw JJ, Daniel SE, Dickson D, et al. Preliminary NINDS neuropathologic criteria for Steele-Richardson-Olszewski syndrome (progressive supranuclear palsy). *Neurology*. 1994;44(11):2015-2015.

5. Kovacs GG, Lukic MJ, Irwin DJ, et al. Distribution patterns of tau pathology in progressive supranuclear palsy. *Acta Neuropathol (Berl)*. 2020;140(2):99-119. doi:10.1007/s00401-020-02158-2

6. 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):601-620. doi:10.1038/s41582-021-00541-5

7. Baker M, Litvan I, Houlden H, et al. Association of an extended haplotype in the tau gene with progressive supranuclear palsy. *Hum Mol Genet*. 1999;8(4):711-715.

8. Höglinger GU, Melhem NM, Dickson DW, et al. Identification of common variants influencing risk of the tauopathy progressive supranuclear palsy. *Nat Genet*. 2011;43(7):699-705.

9. Wang H, Chang TS, Dombroski BA, et al. Whole-Genome Sequencing Analysis Reveals New Susceptibility Loci and Structural Variants Associated with Progressive Supranuclear Palsy. *medRxiv*. Published online 2023:2023-12.

10. Conrad C, Andreadis A, Trojanowski JQ, et al. Genetic evidence for the involvement of τ in progressive supranuclear palsy. *Ann Neurol*. 1997;41(2):277-281. doi:10.1002/ana.410410222

11. Heckman MG, Brennan RR, Labbé C, et al. Association of MAPT subhaplotypes with risk of progressive supranuclear palsy and severity of tau pathology. *JAMA Neurol*. 2019;76(6):710-717.

12. Heckman MG, Kasanuki K, Brennan RR, et al. Association of MAPT H1 subhaplotypes with neuropathology of lewy body disease. *Mov Disord*. 2019;34(9):1325-1332. doi:10.1002/mds.27773

13. Pittman AM, Myers AJ, Abou-Sleiman P, et al. Linkage disequilibrium fine mapping and haplotype

association analysis of the tau gene in progressive supranuclear palsy and corticobasal degeneration. *J Med Genet*. 2005;42(11):837-846.

14. Boettger LM, Handsaker RE, Zody MC, McCarroll SA. Structural haplotypes and recent evolution of the human 17q21. 31 region. *Nat Genet*. 2012;44(8):881-885.

15. Steinberg KM, Antonacci F, Sudmant PH, et al. Structural diversity and African origin of the 17q21. 31 inversion polymorphism. *Nat Genet*. 2012;44(8):872-880.

16. Kuzma A, Valladares O, Cweibel R, et al. NIAGADS: The NIA Genetics of Alzheimer's Disease Data Storage Site. *Alzheimers Dement*. 2016;12(11):1200-1203. doi:10.1016/j.jalz.2016.08.018

17. Beecham GW, Bis JC, Martin ER, et al. The Alzheimer's Disease Sequencing Project: study design and sample selection. *Neurol Genet*. 2017;3(5).

18. Jin Y, Schaffer AA, Feolo M, Holmes JB, Kattman BL. GRAF-pop: a fast distance-based method to infer subject ancestry from multiple genotype datasets without principal components analysis. *G3 Bethesda Md*. 2019;9(8):2447-2461.

19. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM. Robust relationship inference in genome-wide association studies. *Bioinformatics*. 2010;26(22):2867-2873.

20. Wang H, Dombroski BA, Cheng PL, et al. Structural Variation Detection and Association Analysis of Whole-Genome-Sequence Data from 16,905 Alzheimer's Diseases Sequencing Project Subjects. *medRxiv*. Published online 2023.

21. Cantsilieris S, Western PS, Baird PN, White SJ. Technical considerations for genotyping multi-allelic copy number variation (CNV), in regions of segmental duplication. *BMC Genomics*. 2014;15(1):329. doi:10.1186/1471-2164-15-329

22. Handsaker RE, Van Doren V, Berman JR, et al. Large multiallelic copy number variations in humans. *Nat Genet*. 2015;47(3):296-303.

23. Sharp AJ, Locke DP, McGrath SD, et al. Segmental duplications and copy-number variation in the human genome. *Am J Hum Genet*. 2005;77(1):78-88.

24. Suvakov M, Panda A, Diesh C, Holmes I, Abyzov A. CNVpytor: a tool for copy number variation detection and analysis from read depth and allele imbalance in whole-genome sequencing. *Gigascience*. 2021;10(11):giab074.

25. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine learning in Python. *J Mach Learn Res*. 2011;12:2825-2830.

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.

27. Delaneau O, Zagury JF, Robinson MR, Marchini JL, Dermitzakis ET. Accurate, scalable and integrative haplotype estimation. *Nat Commun*. 2019;10(1):5436.

28. Rajput A, Rajput AH. Progressive Supranuclear Palsy: Clinical Features, Pathophysiology and Management. *Drugs Aging*. 2001;18(12):913-925. doi:10.2165/00002512-200118120-00003

29. Rademakers R, Melquist S, Cruts M, et al. High-density SNP haplotyping suggests altered regulation of tau gene expression in progressive supranuclear palsy. *Hum Mol Genet*. 2005;14(21):3281-3292.

30. Caffrey TM, Joachim C, Paracchini S, Esiri MM, Wade-Martins R. Haplotype-specific expression of exon 10 at the human MAPT locus. *Hum Mol Genet*. 2006;15(24):3529-3537.

31. Sudmant PH, Kitzman JO, Antonacci F, et al. Diversity of human copy number variation and multicopy genes. *Science*. 2010;330(6004):641-646.