RESEARCH ARTICLE

Copy Number Variation and Haplotype Analysis of 17q21.31 Reveals Increased Risk Associated with Progressive Supranuclear Palsy and Gene Expression Changes in Neuronal Cells

Hui Wang, PhD,[1,2] Timothy S. Chang, MD, PhD,[3] Beth A. Dombroski, PhD,[1,2] Po-Liang Cheng, PhD,[1,2] Ya-Qin Si, PhD,[4] Albert Tucci, PhD,[4] Vishakha Patil, MS,[3] Leopoldo Valiente-Banuet,[3] Chong Li, MS,[5] Kurt Farrell, PhD,[6] Catriona Mclean, MD,[7] Laura Molina-Porcel, MD, PhD,[8,9] Alex Rajput, MD,[10] Peter Paul De Deyn, MD, PhD,[11,12] Nathalie Le Bastard, PhD,[13] Marla Gearing, PhD,[14] Laura Donker Kaat, MD, PhD,[15] John C. Van Swieten, MD, PhD,[15] Elise Dopper, MD, PhD,[15] Bernardino F. Ghetti, MD,[16] Kathy L. Newell, MD,[16] Claire Troakes, PhD,[17] Justo G. de Yébenes, MD, PhD,[18] Alberto Rábano-Gutierrez, MD, PhD,[19] Tina Meller, PhD,[20] Wolfgang H. Oertel, MD, PhD,[20] Gesine Respondek, MD,[21] Maria Stamelou, MD,[22,23] Thomas Arzberger, MD,[24,25] Sigrun Roeber, MD,[25] Ulrich Müller, MD,[26] Franziska Hopfner, MD,[27] Pau Pastor, MD, PhD,[28,29] Alexis Brice, MD,[30] Alexandra Durr, MD, PhD,[30] Isabelle Le Ber, MD, PhD,[30] Thomas G. Beach, MD, PhD,[31] Geidy E. Serrano, PhD,[31] Lili-Naz Hazrati, MD, PhD,[32] Irene Litvan, MD,[33] Rosa Rademakers, PhD,[34,35] Owen A. Ross, PhD,[35] Douglas Galasko, MD,[33] Adam L. Boxer, MD, PhD,[36] Bruce L. Miller, MD,[36] Willian W. Seeley, MD,[36] Vivianna M. Van Deerlin, MD, PhD,[1] Edward B. Lee, MD, PhD,[1,37] Charles L. White, III, MD,[38] Huw R. Morris, MD, PhD,[39] Rohan de Silva, PhD,[40] John F. Crary, MD, PhD,[6] Alison M. Goate, PhD,[41] Jeffrey S. Friedman, MD, PhD,[42] Yaroslau Compta, MD, PhD,[43,44] Yuk Yee Leung, PhD,[1,2] Giovanni Coppola, MD,[3,45] Adam C. Naj, PhD,[1,2,46] Li-San Wang, PhD,[1,2] PSP Genetics Study Group, Clifton Dalgard, PhD,[47] Dennis W. Dickson, MD,[35] Günter U. Höglinger, MD,[21,27,48] Jung-Ying Tzeng, PhD,[4,49] Daniel H. Geschwind, MD, PhD,[3,50,51] Gerard D. Schellenberg, PhD,[1,2] and Wan-Ping Lee, PhD[1,2*]

ABSTRACT: Background: The 17q21.31 region with various structural forms characterized by the H1/H2 haplotypes and three large copy number variations (CNVs) represents the strongest risk locus in progressive supranuclear palsy (PSP).

Objective: To investigate the association between CNVs and structural forms on 17q.21.31 with the risk of PSP.

Methods: Utilizing whole genome sequencing data from 1684 PSP cases and 2392 controls, the three large CNVs (α, β, and γ) and structural forms within 17q21.31 were identified and analyzed for their association with PSP.

Results: We found that the copy number of γ was associated with increased PSP risk (odds ratio [OR] = 1.10, $P = 0.0018$). From H1β1γ1 (OR = 1.21) and H1β2γ1 (OR = 1.24) to H1β1γ4 (OR = 1.57), structural forms of H1 with additional copies of γ displayed a higher risk for PSP. The frequency of the risk sub-haplotype H1c rises from 1% in individuals with two γ copies to 88% in those with eight copies. Additionally, γ duplication up-regulates expression of *ARL17B*, *LRRC37A/LRRC37A2*, and *NSFP1*, while down-regulating *KANSL1*. Single-nucleus RNA-seq of the dorsolateral prefrontal cortex analysis reveals γ duplication primarily up-regulates *LRRC37A/LRRC37A2* in neuronal cells.

Conclusions: The copy number of γ is associated with the risk of PSP after adjusting for H1/H2, indicating that the complex structure at 17q21.31 is an important consideration when evaluating the genetic risk of PSP. © 2025 The Author(s). *Movement Disorders* published by Wiley Periodicals LLC on behalf of International Parkinson and Movement Disorder Society.

Key Words: progressive supranuclear palsy; H1 and H2 haplotypes; 17q21.31; copy number variations; single-cell gene expression

[1]Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA; [2]Penn Neurodegeneration Genomics Center, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA; [3]Movement Disorders Programs, Department of Neurology, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, California, USA; [4]Bioinformatics Research Center, North Carolina State University, Raleigh, North Carolina, USA; [5]Department of Computer and Information Sciences, College of Science and Technology, Temple University, Philadelphia, Pennsylvania, 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 and Research CoRE, Icahn School of Medicine at Mount Sinai, New York City, New York, 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

Progressive supranuclear palsy (PSP) is a neurodegenerative disease characterized by the accumulation of tau in the brain along with symptoms such as postural instability and ocular motor abnormalities.[1-3] Despite a number of other loci identified through association studies in the last decade,[4-7] the 17q21.31 of human

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, Antwerp, Belgium; [12]Department of Neurology, University Medical Center Groningen, Groningen, The Netherlands; [13]Fujirebio Europe NV, Ghent, Belgium; [14]Department of Pathology and Laboratory Medicine and Department of Neurology, Emory University School of Medicine, Atlanta, Georgia, USA; [15]Netherlands Brain Bank and Erasmus University, Rotterdam, The Netherlands; [16]Department of Pathology and Laboratory Medicine, Indiana University School of Medicine, Indianapolis, Indiana, USA; [17]London Neurodegenerative Diseases Brain Bank, King's College London, London, United Kingdom; [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, Munich, Germany; [25]Center for Neuropathology and Prion Research, Ludwig-Maximilians-University Munich, Munich, Germany; [26]Institute of Human Genetics, Justus-Liebig University Giessen, Giessen, Germany; [27]Department of Neurology, LMU University Hospital, Ludwig-Maximilians-Universität München, Munich, Germany; [28]Unit of Neurodegenerative Diseases, Department of Neurology, University Hospital Germans Trias i Pujol, Barcelona, Spain; [29]Neurosciences, The Germans Trias i Pujol Research Institute, Barcelona, Spain; [30]Sorbonne Université, Paris Brain Institute – Institut du Cerveau – ICM, Inserm U1127, CNRS UMR 7225, APHP - Hôpital Pitié-Salpêtrière, Paris, France; [31]Banner Sun Health Research Institute, Sun City, Arizona, USA; [32]Department of Pathology, University McGill, Montreal, Quebec, Canada; [33]Department of Neuroscience, University of California San Diego, La Jolla, California, USA; [34]VIB Center for Molecular Neurology, University of Antwerp, Antwerp, Belgium; [35]Department of Neuroscience, Mayo Clinic Jacksonville, Jacksonville, Florida, USA; [36]Memory and Aging Center, University of California San Francisco, San Francisco, California, USA; [37]Center for Neurodegenerative Disease Research, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania, USA; [38]University of Texas Southwestern Medical Center, Dallas, Texas, USA; [39]Departmento of Clinical and Movement Neuroscience, University College of London, London, United Kingdom; [40]Reta Lila Weston Institute, UCL Queen Square Institute of Neurology, London, United Kingdom; [41]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York City, New York, USA; [42]Friedman Bioventure, Inc., Del Mar, California, USA; [43]Parkinson's Disease & Movement Disorders Unit, Hospital Clínic de Barcelona; IDIBAPS, CIBERNED (CB06/05/0018-ISCIII), ERN-RND, UBNeuro - Maria de Maeztu Excellence Centre, Universitat de Barcelona, Barcelona, Spain; [44]Fellow of the Royal Academy of Medicine of Catalonia, Barcelona, Spain; [45]Department of Psychiatry, Semel Institute for Neuroscience and Human Behavior, University of California Los Angeles, Los Angeles, California, USA; [46]Department of Biostatistics, Epidemiology, and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA; [47]Department of Anatomy Physiology and Genetics, the American Genome Center, Uniformed Services University of the Health Sciences, Bethesda, Maryland, USA; [48]Munich Cluster for Systems Neurology (SyNergy), Munich, Germany; [49]Department of Statistics, North Carolina State University, Raleigh, North Carolina, USA; [50]Department of Human Genetics, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, California, USA; [51]Institute of Precision Health, University of California Los Angeles, Los Angeles, California, USA

*__Correspondence to:__ Asst. Prof. W.-P. Lee, Richards Medical Research Laboratories, D101, 3700 Hamilton Walk, Philadelphia, PA 19104, USA. E-mail: wan-ping.lee@pennmedicine.upenn.edu

Investigators of the PSP Genetics Study Group are listed in the Appendix.

Relevant conflicts of interest/financial disclosures: L.M.-P. received income from Biogen as a consultant in 2022. G.R. has been employed by Roche (Hoffmann-La Roche, Basel, Switzerland) since 2021. Her affiliation while completing her contribution to this manuscript was German Center for Neurodegenerative Diseases (DZNE), Munich, Germany. T.G.B. is a consultant for Aprinoia Therapeutics and a scientific advisor and stock option holder for Vivid Genomics. H.R.M. is employed by University College London (UCL). In the last 12 months he reports paid consultancy from Roche, Aprinoia, AI Therapeutics, and Amylyx; and lecture fees/honoraria from BMJ, Kyowa Kirin, and the Movement Disorder Society. H.R.M. is a co-applicant on a patent application related to C9ORF72: Method for diagnosing a neurodegenerative disease (PCT/GB2012/052140). G.C. is currently an employee of Regeneron Pharmaceuticals. A.M.G. serves on the scientific advisory board for Genentech and Muna Therapeutics.

Funding agencies: This work was supported by National Institutes of Health (NIH) 5UG3NS104095, the Rainwater Charitable Foundation, and CurePSP. H.W. and P.-L.C. are supported by RF1-AG074328, P30-AG072979, U54-AG052427, and U24-AG041689. T.S.C. is supported by NIH K08AG065519 and the Larry L. Hillblom Foundation 2021-A-005-SUP. Y.-Q.S., A.T., and J.-Y.T. are supported by RF1-AG074328. K.F. was supported by CurePSP 685-2023-06-Pathway and K01 AG070326. M.G. is supported by P30 AG066511. B.F.G. and K.L.N. are supported by P30 AG072976 and R01 AG080001. T.G.B. and G.E.S. are supported by U24 NS072026, P30 AG019610, P30AG072980, the State of Arizona, and The Michael J. Fox Foundation for Parkinson's Research. I.L. is supported by 2R01AG038791-06A, U01NS100610, R25NS098999, U19 AG063911-1, and 1R21NS114764-01A1. O.A.R. is supported by U54 NS100693. D.G. is supported by P30AG062429. A.L.B. is supported by U19AG063911, R01AG073482, R01AG038791, and R01AG071756. B.L.M. is supported by P01 AG019724, R01 AG057234, and P0544014. V.M.V.D. is supported by P01-AG-066597 and P01-AG-017586. H.R.M. is supported by CurePSP, PSPA, MRC, and The Michael J. Fox Foundation. R.D.S. is supported by CurePSP, PSPA, and Reta Lila Weston Trust. J.F.C. 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. A.M.G. is supported by the Tau Consortium and U54-NS123746. Y.C. is supported by CIBERNED (CB06/05/0018-ISCIII), Maria de Maeztu Excellence Center, CERCA Generalitat de Catalunya. Y.Y.L. is supported by U54-AG052427 and U24-AG041689. L.-S.W. is supported by U01AG032984, U54AG052427, and U24AG041689. G.U.H. 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). D.H.G. is supported by 3UH3NS104095 and Tau Consortium. W.-P.L. is supported by RF1-AG074328, P30-AG072979, U54-AG052427, and 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 National Institute on Aging (NIA) grants P50 AG016574, CurePSP Foundation, and support from Mayo Foundation.

Received: 10 July 2024; **Revised:** 24 January 2025; **Accepted:** 29 January 2025

Published online 8 March 2025 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/mds.30150

genome, which presents two haplotypes H1 and H2 (distinguished by a ~1 Mb inversion, Figure S1A in Data S1), remains the most prominent genetic risk factor for PSP. The *MAPT* gene, which encodes the microtubule-associated protein tau, is the most prominent risk factor within the 17q21.31 region.[8-10] In addition, recent functional studies using multiple parallel reporter assays coupled to CRISPR interference (CRISPRi) have identified other risk genes in this locus, including *KANSL1* and *PLEKHM1*.[11]

The 17q21.31 is one of the most structurally complex regions in the human genome, featuring multiple rearrangements throughout the evolutional history. At least 10 structural forms within 17q21.31 can be characterized by H1 and H2 along with three large duplications (ie, α, β, γ; Figure S1B in Data S1).[12,13] However, the impact of these structural forms and copy number variations (CNVs) on PSP risk has not been systematically assessed. To assess the impact of these structural forms and CNVs on PSP risk, the copy numbers of α, β, and γ and structural forms of 17q21.31 were called from whole genome sequencing (WGS) data (Figure S1C in Data S1). Case–control analysis was performed to identify CNVs significantly associated with PSP and single nucleus RNA-seq analysis was employed to evaluate the regulatory role of CNVs on gene expression.

Methods

Study Subjects

All study subjects and WGS data are available on The National Institute on Aging Genetics of Alzheimer's Disease Data Storage Site (NIAGADS)[14] under Alzheimer's Disease Sequencing Project (ADSP) Umbrella NG00067.v7.[15] All human subjects provided informed consent. We inferred the ancestry of subjects by GRAF-pop (Version 1.0, https://github.com/ncbi/graf)[16] and selected 4618 subjects (1797 cases and 2821 controls) of European ancestry for analysis. WGS were performed at 30× coverage (Table S1 in Data S2).

Among 4618 samples, we filtered 183 samples with abnormally low reads mapped (aligned read depth <1.7×) to α, β, or γ region (Figure S2 in Data S1) and 10 samples with high genotyping missing rate (>0.05). Next, 244 related samples inferred by KING (Version 2.3.1, https://www.kingrelatedness.com/)[17] (duplicates, monozygotic twins, parent-offsprings, full-siblings, and second-degree relatives) were removed while retaining one sample from each related group. We used the 238-base pair (bp) deletion between exons 9 and 10 of *MAPT*[18] to determine the H1 and H2 haplotypes of each sample. The genotype calls of the 238-bp deletion were obtained from our previous structural variant work.[19] Some 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 10 structural forms (Figure S1B in Data S1) in each individual. We removed 30 individuals (Figure S3 in Data S1) 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, 4076 subjects (Table 1; $N_{PSP} = 1684$, $N_{control} = 2392$) remained for statistical analyses in this study. Among them, 1684 PSP cases and 145 controls were sourced from the PSP-NIH-CurePSP-Tau, PSPCurePSP-Tau, PSP-UCLA, and AMPAD-MAYO cohorts included in ADSP (NG0067.v7), while an additional 2247 controls were drawn from other ADSP cohorts (Table S2 in Data S2). Detailed information about each cohort is available through NIAGADS.[14] Of the 1684 individuals diagnosed with PSP, 1386 were autopsy-confirmed. Clinical diagnosis criteria are outlined in the Supplementary Methods in Data S1. Age was missing for 1130 PSP cases as autopsy-confirmed cases determined at brain banks did not always have the age of symptom onset when brain tissue was sent from outside the brain bank's health system. The mean age of onset for PSP cases was 68.03 years and the mean age at the last visit for controls was 81.04 years (Table 1).

Determine the Copy Number of α, β, and γ and Structural Forms of 17q21.31

The genomic coordinates on HG38 for α (chr17:46,135,415–46,289,349), β (chr17:46,087,894–46,356,512), and γ (chr17:46,289,349–46,707,123) were obtained from two previous studies[12,13] (Figure S1A in Data S1). Segmental duplications can introduce mapping challenges and thus inaccurate calling of the number of copies.[20-22] To address this, we removed segmental duplicated regions inside the α, β, and γ (Figure S4 in Data S1) when calculating aligned read depth. Subsequently, the copy numbers of α, β, and γ were obtained based on the aligned read depth on chr17:46,135,415–46,203,287, chr17:46,106,189–46,135,415, and chr17:46,356,512-46,489,410/chr17:46,565,081–46,707,123, 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, https://github.com/abyzovlab/CNVpytor).[23] Then, we employed K-means[24] to assign an integer copy number for α, β, and γ for the 4076 individuals. Each individual was found to have up to six copies of α or β and up to eight copies of γ (Figure S1 in Data S1). On the H1 background, the β region, which includes α, can duplicate up to four copies, whereas on the H2 background, only

TABLE 1 *Characteristics of progressive supranuclear palsy cases and controls*

Characteristic	Overall (N = 4076)	PSP (N = 1684)	Control (N = 2392)
Age, years (SD)[a]	78.49 (8.50)	68.03 (8.17)	81.04 (6.37)
Sex, n (%)			
Female	2168 (53.19)	739 (43.88)	1429 (59.74)
Male	1908 (46.81)	945 (56.12)	963 (40.26)
H1/H2 status, n (%)[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, n (%)[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)

Abbreviations: PSP, progressive supranuclear palsy; SD, standard deviation.
[a]1130 PSP cases and 111 controls have missing age. Age for PSP refers to the age at disease onset, while age for controls indicates the age at last visit.
[b]H1/H2 status was determined by the genotype of a 238-bp H2 tagging deletion.[5]
[c]Structural forms of 17q21.31 were inferred by the H1/H2 status and the copy numbers of α, β, and γ (see Methods).

the α region duplicates, with a maximum of two copies of γ (Figure S1 in Data S1).

To validate the copy numbers of α, β, and γ called from WGS, 65 samples were genotyped using TaqMan CNV assay. For α and β, we utilized the same TaqMan primer, given that β shares largely the same region with α and has the same copy number in H1 haplotypes. To assess the accuracy of β copy number calls from WGS, we focused on 60 of the 65 samples with an H1/H1 genotype, as β is not duplicated in H2 haplotypes. Overall, the copy number of α, β, or γ inferred by aligned read depth from WGS were highly consistent (α, R = 0.87; β, R = 0.85; γ, R = 0.96) with that from TaqMan assay (Figure S5 in Data S1). Notably, for high-confident calls from the TaqMan assay: all γ copy numbers matched those obtained from WGS; only two individuals showed discrepancies between WGS and TaqMan assay in α and β copy numbers, including one case with an improbable single copy of both α and β detected by TaqMan. The experimental procedure is detailed in the Supplementary Methods in Data S1.

For approximately 60% of the samples, only one combination of the structural forms (Figure S1B in Data S1) 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 remainder of the samples, multiple haplotypic combinations were possible. The expectation–maximization (EM) algorithm[12] (Supplementary Methods in Data S1) were employed to infer the two structural forms of 17q21.31 in each individual. The allele frequency of each structural form of 17q21.31 after EM convergence are shown in Figure S1B in Data S1. Overall, H2α2γ2 dominates the structural forms of H2 while several structural forms of H1 (H1β1γ1, H1β1γ2, H1β1γ3, and H1β2γ1) showed an allele frequency >10%.

Genetic Analysis of MAPT Sub-Haplotypes and Structural Forms of 17q21.31

The six single nucleotide variants (SNVs) (rs1467967, rs242557, rs3785883, rs2471738, rs8070723, and rs7521)[25-27] on *MAPT* were employed to define the

26 *MAPT* sub-haplotypes (Table S3 in Data S2). We phased the six SNVs with other SNVs and indels in chr17:43,000,000–48,000,000 to determine the *MAPT* sub-haplotypes. The SNV genotypes for the study subjects were called in our previous work.[28] Variants were removed if they were monomorphic, did not pass variant quality score recalibration, had an average read depth ≥500, or if all calls had DP < 10 and 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[29] (Version 4.2.2) with default parameters.

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[12] (Table S4 in Data S2). Then, SHAPEIT4[29] (Version 4.2.2, https://odelaneau.github.io/shapeit4/) with default parameters were used for phasing the copy numbers of α, β, and γ together with SNVs/indels. SNVs and indels inside α, β, and γ regions (chr17:46,087,000–46,708,000) were not included when phasing. After phasing, we calculated the linkage disequilibrium (LD) between structural forms of 17q21.31 and *MAPT* sub-haplotypes.

Association Analysis

Association analyses were performed for the 4076 individuals ($N_{PSP} = 1684$, $N_{control} = 2392$). For the association of the copy numbers of α, β, and γ with PSP, the default logistic regression model was adjusted for sex and principal components (PCs) 1–5. We also tested the models when the allele count of H2 was added as an additional covariate, as the β region can only duplicate in the H1 haplotype, the smaller α region but not the entire β duplicates in the H2 haplotype, and the γ region usually duplicates only once in the H2 haplotype (Figure S1B in Data S1). Then, association analysis was performed separately for individuals with H1H1 and H1H2 genotypes. Individuals with the H2H2 genotype are imbalanced and with few cases (5 cases, 138 controls), 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% was compared with the rest of structural forms using logistic regression model adjusting for sex and PCs 1–5.

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 (Table S5 in Data S2). Two logistic regression models were used: one adjusted for sex and PCs 1–5, and the other included H2 allele count as an additional covariate. All statistical analyses were performed using R (Version 4.2.1).[30]

Bulk and Single-Nucleus RNA-Seq Analysis

We used RNA-seq data from Mayo RNA-seq study[31-33] and snRNA-seq data from the Religious Order Study and the Rush Memory and Aging Project (ROSMAP).[34] To calculate the association between CNVs and gene expression, we only included overlapping samples in Mayo RNS-seq data (N = 211, Table S6 in Data S2) and ROSMAP snRNA-seq data (N = 276, Table S7 in Data S2) that had available WGS data from ADSP. For bulk RNA-seq, 191 individuals with RNA extracted from cerebellum and 189 individuals with RNA extracted from temporal cortex were used. Library preparation was performed by the TruSeq RNA Sample Prep Kit V2 (Illumina, San Diego, CA, USA). Illumina HiSeq 4000 sequencers (Illumina) were used for 100-bp paired-end sequencing. Read alignments were performed by SNAPR software (https://github.com/PriceLab/snapr)[35] and counts per million were calculated using edgeR.[36] Detailed methods for bulk RNA-seq can be found in previous studies.[31-33] For single-nucleus RNA-seq, 276 individuals with nucleus RNA from the dorsolateral prefrontal cortex were used. Single nuclei samples were isolated and profiled by the 10X Single Cell RNA-seq Platform using the Chromium Single Cell 3' Reagent Kits Version 3 (10X Genomics, Pleasanton, CA, USA). Libraries were aligned to the GRCh38 using CellRanger.[37] Pseudobulk gene expression for CUX2+ neurons, CUX2− neurons, inhibitory neurons, astrocytes, microglia, oligodendrocytes, oligodendrocytes precursor cells, and vascular cells were aggregated and log-normalized by Seurat (Version 5.0.3, https://github.com/satijalab/seurat).[38] Detailed methods for single-nucleus RNA-seq can be found in a previous study.[34]

To analyze the effect of γ on gene expression on 17q21.31 (42 genes, chr17:44,800,000–47,000,000), linear regression model adjusting for the allele count of H2, sex, and PCs 1–5 were employed and a Bonferroni-corrected *P* cutoff of 0.001 (0.05/42) was applied. The rs17660065[12] was used to tag H2 when the genotype for the 238-bp deletion[18] was unavailable. All statistical analyses were performed using R (Version 4.2.1).[30]

Results

Copy Number of γ and PSP Risk

Our initial analysis focused on whether the copies of α, β, or γ are associated to the risk of PSP and if these associations are due to correlation with the H1 and H2 haplotypes. Adjusting for sex, PCs 1–5, and allele count (0, 1, or 2) of the H2 haplotype, we observed that copy number of γ was associated with 1.10-fold of increased risk of PSP (95% CI 1.04–1.17; $P = 0.0018$; Table 2). As H2α2γ2 is predominant in H2, the observed increased risk of γ was mainly due to variations in H1. Without adjusting for H2, the higher risk of PSP

TABLE 2 *Association between the copy numbers of α, β, γ and risk of progressive supranuclear palsy*

	N = 4076 (PSP = 1684; Control = 2392)			
	Default model (sex and five PCs)		+H2 in the model (sex, five PCs, and H2)	
CNV	OR (95% CI)	*P*	OR (95% CI)	*P*
γ	0.98 (0.93–1.04)	0.60	1.10 (1.04–1.17)	0.0018★
β	1.14 (1.03–1.27)	0.011★	0.90 (0.81–1.01)	0.064
α	0.57 (0.52–0.63)	$<2 \times 10^{-16}$★	0.90 (0.81–1.00)	0.061
	H1H1 carriers, N = 2958 (PSP = 1511; Control = 1447)		H1H2 carriers, N = 975 (PSP = 168; Control = 807)	
CNV	OR (95% CI)	*P*	OR (95% CI)	*P*
γ	1.08 (1.02–1.15)	0.014★	1.29 (1.06–1.56)	0.0096★
β	0.91 (0.81–1.02)	0.11	0.79 (0.53–1.15)	0.23
α	0.91 (0.81–1.02)	0.11	0.81 (0.58–1.11)	0.20

Note: Association was not analyzed in H2H2 individuals as there were only five H2H2 PSP cases.
Abbreviations: CNV, copy number variation; CPM, counts per million PSP, progressive supranuclear palsy; PC, principal component; OR, odds ratio; CI, confidence interval.
★Represents statistical significance ($P<0.05$).

conferred by γ would be obscured (OR = 0.98; 95% CI 0.93–1.04; $P = 0.60$; Table 2) because H2 haplotype usually has two copies of γ and is protective against PSP (OR = 0.19; 95% CI, 0.16–0.22; $P = 3.00 \times 10^{-79}$) while the most common structural forms of H1 (H1β1γ1, allele frequency = 30%) has only one copy of γ. Another way to eliminate the confounding effects of H1 and H2 is to conduct the association separately on individuals with H1H1, H1H2, or H2H2 genotypes. We found that each additional copy of γ was associated with 1.08-fold (95% CI 1.02–1.15; $P = 0.014$) of increased risk of PSP in H1H1 individuals and 1.29-fold (95% CI 1.06–1.56; $P = 0.0096$) of increased risk of PSP in H1H2 individuals (Table 2; Figure S6 in Data S1). Among H2H2 individuals, who could have two, three, or four copies of γ, all five PSP cases in our data had four copies of γ. Therefore, association analysis was not possible due to insufficient samples in this group.

For α and β, only under the regression model without adjusting H1 and H2, we observed statistically significant association with PSP (Table 2). However, the observed significance mainly arises from their correlation with the H1 and H2 haplotypes, ie, the increased copies (usually two copies) of α and the absence of β duplication in the H2 haplotype. The association, adjusting for sex, PCs 1–5, and allele count (0, 1, or 2) of the H2 haplotype, shows 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 (Table 2). Although individuals with more copies of α and β showed slightly lower odds ratio (OR) for PSP (Table 2).

Structural Forms of 17q21.31 and PSP Risk

For a further analysis, we investigated the structural forms of 17q21.31, characterized by the α, β, and γ duplications along with H1/H2, and their impact on PSP risk. We tested seven structural forms of 17q21.31 with allele frequency >0.01 (Table 3). On the H1 background, the OR for PSP increases from 1.21 (95% CI 1.10–1.33; $P = 5.47 \times 10^{-5}$) for H1β1γ1 to 1.57 (95% CI 1.10–2.26; $P = 1.35 \times 10^{-2}$) for H1β1γ4 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; $Pc\ 1.87 \times 10^{-4}$) displayed similar risk of PSP compared with H1β1γ1 (OR = 1.21; 95% CI 1.10–1.33; $P = 5.47 \times 10^{-5}$). This finding reaffirmed that the copy number of γ was associated with increased risk of PSP, and β was not associated with the risk of PSP (Table 2; Figure S6 in Data S1). On the H2 background, it was not practical to evaluate the effect of γ as H2α2γ2 dominates (Figure S1B in Data S1).

Copy Number of γ and MAPT Sub-Haplotypes

Besides the 10 structural forms, there are 26 *MAPT* sub-haplotypes (Table S3 in Data S2) based on six tagging SNVs[25-27] representing the smaller LD structure in *MAPT* gene (~150 kb). We observed the association with the risk of PSP in 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; $P = 2.77 \times 10^{-12}$) (Table S5 in Data S2). H1g (OR = 1.46; 95% CI 1.07–1.98; $P = 0.016$) and H1h (OR = 1.36; 95% CI 1.10–1.69; $P = 0.0053$) were nominal significant in our analysis

TABLE 3 *Structural forms of 17q21.31 and the risk of progressive supranuclear palsy*

Structural form	Frequency (%)		OR (95% CI)	*P*
	PSP (N = 1684)	Control (N = 2392)		
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)	$<2 \times 10^{-16}$

Note: Haplotypes in less than 1% of individuals were excluded.
OR and *P* value were from logistic regression adjusting for PCs 1–5 and sex.
Abbreviations: OR, odds ratio; CI, confidence interval; PSP, progressive supranuclear palsy; PC, principal component.

(Table S5 in Data S2). As the observed increased risk of those H1 sub-haplotypes could be due to the protective effect of H2, we performed additional tests adjusting for the allele count of H2. Despite the observed lower OR, H1c (OR = 1.40; 95% CI 1.22–1.59; $P = 6.78 \times 10^{-7}$) and H1o (OR = 2.37; 95% CI 1.75–3.24; $P = 4.32 \times 10^{-8}$) remained significant (Table S5 in Data S2). This confirmed the previous genome-wide association study (GWAS) findings that the H1c tagging SNV rs242557 still contributes to the risk of PSP after controlling for H1/H2.[4,25] Moreover, individuals with H1b showed a lower risk of PSP compared with other H1 sub-haplotypes when the allele count of H2 was adjusted in the regression model (OR = 0.79; 95% CI 0.70–0.90; $P = 4.10 \times 10^{-4}$) (Table S5 in Data S2). These results further refined the association between H1 and PSP through the H1 sub-haplotypes.

In line with the increased risk of PSP in individuals carrying H1c and extra copies of γ, we observed an association between γ and H1c. The proportion of H1c increased from 1% in individuals with two copies of γ to 88% in individuals with eight copies of γ (Fig. 1). Furthermore, 96% of H1c sub-haplotypes corresponded to structural forms of 17q21.31 with more than one copy of γ (H1β1γ2, H1β1γ3, and H1β1γ4) (Figure S7 in Data S1). When compared with structural forms of 17q21.31 with exactly one copy of γ (H1β1γ1 and H1β2γ1), structural forms with additional copies of γ (H1β1γ2, H1β1γ3, and H1β1γ4) were more likely to be H1c or other *MAPT* sub-haplotypes associated with increased risk of PSP (ie, H1o, H1d, H1g, and H1g) (Figure S7 in Data S1). We then phased the CNVs (Table S4 in Data S2; Methods) together with SNVs to examine the LD between structural forms of 17q21.31 and *MAPT* sub-haplotypes. Two structural forms were in LD ($R^2 > 0.1$) with *MAPT* sub-haplotypes (Table S8 in Data S2): H1β1γ3 was in LD with H1c ($R^2 = 0.31$) with 70% of H1β1γ3 being H1c and H1β2γ1 was in LD with H1b ($R^2 = 0.29$) with 56% of H1β2γ1 being H1b (Figure S7 in Data S1).

Copy Number of γ and Gene Expression on 17q21.31

Finally, we examined the function impact of γ duplication on gene expression (Fig. 2A). Based on RNA-seq of the cerebellum, we observed that the expression of three genes located on γ region, ie, *ARL17B* (β = 0.63; $P = 1.16 \times 10^{-20}$), *LRRC37A* (β = 0.48; $P = 4.21 \times 10^{-14}$), and *NSFP1* (β = 0.81; $P = 4.31 \times 10^{-47}$) showed the strongest correlation with the copy number of γ (Fig. 2B; Figure S8A in Data S1; Table S9 in Data S2). This increased expression with higher γ copy numbers was also observed from RNA-seq of the temporal cortex (Fig. 2C; Figure S8B in Data S1; Table S9 in Data S2). We also found higher expression of *LRRC37A2* accompanying γ duplication in the temporal cortex (β = 0.30; $P = 1.56 \times 10–9$) but not the cerebellum (β = 0.07; $P = 0.16$). Further analysis of single-nucleus RNA-seq (snRNA) of cells from the dorsolateral prefrontal cortex revealed that the association between the higher expression of LRRC37A/LRRC37A2 and the increased copy number of γ was mainly driven by neuronal cells (Fig. 2D,E; Table S10 in Data S2). Specifically, the association between *LRRC37A* expression and the copy number γ was not significant ($P > 0.001$) in astrocytes, microglia, oligodendroglia, and vascular cells while it was strongly presented in CUX2+ (β = 0.70; $P = 1.67 \times 10^{-54}$), CUX2− (β = 0.60; $P = 5.25 \times 10^{-40}$), and inhibitory neurons (β = 0.63; $P = 2.77 \times 10^{-47}$) (Table S10 in Data S2). For *LRRC37A2*, the increased copy of γ not only strongly up-regulated its expression in CUX2+ (β = 0.56; $P = 2.72 \times 10^{-50}$), CUX2− (β = 0.46; $P = 2.96 \times 10^{-40}$), and inhibitory neurons (β = 0.58; $P = 1.52 \times 10^{-54}$) but also down-regulated its

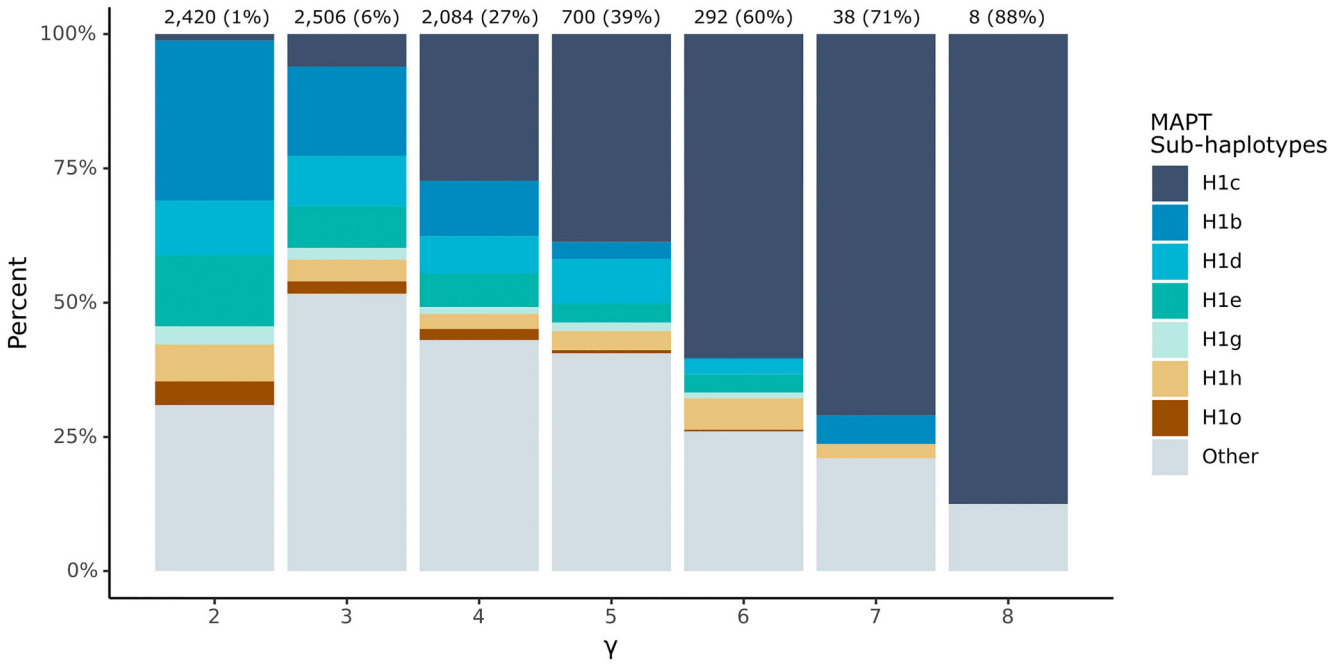

FIG. 1. The association between the copy number of γ and MAPT sub-haplotypes. The number of haplotypes (2 × the number of individuals) are showed on each bar. The percentage of H1c is showed in brackets. The MAPT sub-haplotypes on H1 that were associated with the risk of progressive supranuclear palsy or have an allele frequency >0.05 were color coded. All the other MAPT sub-haplotypes were included in the 'Other' category. The color information: H1c (#45526C), H1b (#2B8CBE), H1d (#4EB3D3), H1e (#5AB4AC), H1g (#C7EAE5), H1h (#DFC27D), H1o (#8C510A), and Other (#D6DCE5). [Color figure can be viewed at wileyonlinelibrary.com]

expression in astrocytes ($\beta = -0.17$; $P = 3.54 \times 10^{-4}$) and oligodendroglia ($\beta = -0.21$; $P = 2.37 \times 10^{-4}$) (Table S10 in Data S2). The over-expression of *LRRC37A* in HeLa cells could cause deformation of plasma membrane shape and the generation of filopodia-like protrusions, followed by apoptosis.[39] This suggests that the cell type-specific up-regulation of *LRRC37A/LRRC37A2* by γ duplication might contribute to the neurodegeneration in PSP. In addition to genes on the γ region, we also observed decreased expression of *KANSL1* associated with increased γ duplications in both bulk RNA-seq and snRNA-seq data (Fig. 2B,C; Tables S7 and S8 in Data S2).

Discussion

In summary, we evaluated the association of the structural forms of 17q21.31, characterized by large duplications α, β, and γ along with H1/H2 haplotype with the risk PSP. We found that the copy number of γ was associated with increased risk of PSP and structural forms with additional γ copies (ie, H1β1γ2, H1β1γ3, and H1β1γ4) exhibited a higher OR for PSP compared with H1β1γ1. This aligns with the observation that individuals with additional copies of γ tended to carry *MAPT* sub-haplotypes with a higher risk of PSP, such as H1c.

We assessed the association between H1c, γ, and PSP risk, adjusting for sex, PCs 1–5, and H2 allele count. Individuals with more γ copies, such as >5 copies (OR = 1.58; 95% CI 1.13–2.22; $P = 7.45 \times 10^{-3}$), >6 copies (OR = 2.61; 95% CI 1.07–7.34; $P = 4.7 \times 10^{-2}$), and >7 copies (4 individuals, 3 with PSP), showed a higher risk for PSP compared with H1c (OR = 1.40). Notably, individuals carrying at least one H1c allele and more than five copies of γ (N = 141; 72 of whom are H1c heterozygotes) demonstrated a PSP risk (OR = 1.88; 95% CI 1.30–2.74; $P = 8.85 \times 10^{-4}$) equivalent to that of H1c homozygotes (N = 104; OR = 1.88; 95% CI 1.24–2.92; $P = 3.74 \times 10^{-3}$). However, due to their strong correlation, H1c remained significant (OR = 1.43; 95% CI 1.19–1.71; $P = 1.04 \times 10^{-4}$) and γ did not reach significance (OR = 0.99; 95% CI 0.91–1.07; $P = 0.74$) under the same regression model. This suggests several possible scenarios: (1) the increased risk associated with H1c is due to γ combined with other unknown factors; (2) H1c is a causal factor, and γ is irrelevant; or (3) another hidden collider variable may be driving the association. The first scenario is more plausible from a genomic perspective, as H1c is inferred by LD structure using SNVs, which likely capture structural changes in 17q21.31, including the additional γ copies. Further studies with larger sample sizes are needed to clarify the causal relationship between γ and H1c, as well as to explore the impact of extreme γ values and the co-occurrence of H1c with elevated γ copy numbers.

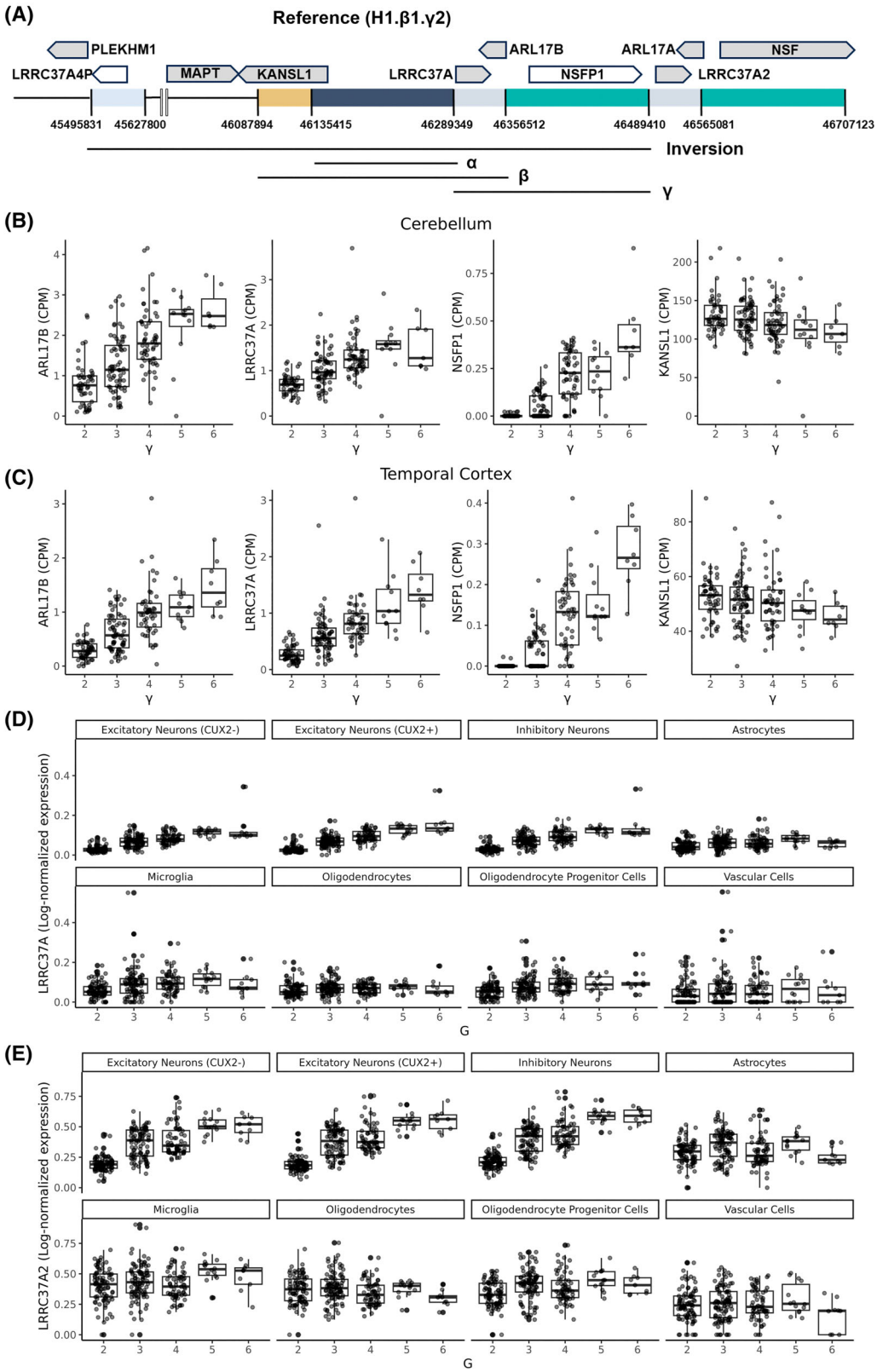

FIG. 2. The association between the copy number of γ and gene expression. (**A**) Schematic plot of gene locations at 17q.21.31. (**B**) Gene expression values for three genes on the γ duplication. Total RNA was isolated from the cerebellum of 191 samples. (**C**) Gene expression values for three genes on the γ duplication. Total RNA was isolated from the temporal cortex of 189 samples. (**D–E**) *LRRC37A/LRRC37A2* pseudobulk expression for different cell types in dorsolateral prefrontal cortex stratified by the number of γ duplication. Pseudobulk counts were log-normalized using AggregateExpression function from Seurat.[38] CPM, counts per million. [Color figure can be viewed at wileyonlinelibrary.com]

Bulk RNA-seq of the cerebellum and temporal cortex revealed higher expression of *ARL17B*, *LRRC37A*, *LRRC37A2*, and *NSFP1* and lower expression of *KANSL1* in individuals with more copies of γ (Table S9 in Data S2). Notably, *LRRC37A2* and *KANSL1* are located outside the γ region, suggesting that their altered expression is likely driven by the gain of enhancers or three-dimensional chromatin structure changes accompanying the γ duplication.[11,40] snRNA-seq of the dorsolateral prefrontal cortex analysis revealed γ duplication primarily up-regulates *LRRC37A/LRRC37A2* in neuronal cells, down-regulates *KANSL1* across all cell types, and upregulates *ARL17B* in microglia (Table S10 in Data S2). For *ARL17B*, a nominally significant ($P < 0.05$) association with γ duplication was observed in neuronal cells using snRNA-seq (Table S10 in Data S2). In snRNA-seq, CellRanger[37] was employed for read alignment and used an abridged version of Ensembl annotations. Therefore, most pseudogenes, including *NSFP1*, were removed, which might shift counts towards the normal genes and explain the observed increased *NSF* expression associated with γ duplication in inhibitory/CUX2+ neurons and oligodendrocyte precursor cells (Table S10 in Data S2).

From bulk RNA-seq, we also observed significantly lower expression of *ARL17A* and higher expression of *FAM215B* accompanying increased copy number of γ (Table S9 in Data S2). However, similar expression changes were not observed across different cell types in snRNA-seq (Table S10 in Data S2), potentially reflecting differences in gene expression profiles between the cerebellum and temporal cortex (bulk RNA-seq) versus the dorsolateral prefrontal cortex (snRNA-seq). It is also important to note that both bulk and single-cell RNA-seq in this study utilized poly(A) selection, which targets mRNA with polyadenylation and may not fully capture expression changes of lncRNAs and pseudogenes (such as *FAM215B* and *NSFP1* on γ duplication).[41,42] To more accurately assess the expression of these genes, future studies based on rRNA depletion methods without poly(A) selection are necessary.

Age is a recognized risk factor for PSP, with the condition typically affecting individuals in their 60s.[43] However, age was not included as a covariate in the regression model due to missing data for more than half of the PSP cases (Table 1). To evaluate the potential impact of age on our analysis, we used 2835 individuals with available age data and found no significant associations between age and the copy number of α, β, or γ ($P > 0.05$) after adjusting for H2 allele count, sex, and the first five PCs. Nonetheless, as more PSP cases with available age data become accessible, it will be important to reassess the influence of age on our findings to ensure the robustness of our conclusions.

Variants within the H1/H2 haplotypes likely contribute to PSP risk by interacting with *MAPT*, the gene that encodes tau and is directly linked to PSP pathology. Consequently, it is essential to understand how structural forms of 17q21.31, including changes in γ duplications, might alter the regulatory landscape in this region and impact MAPT function and PSP risk. There is already evidence suggesting that genes on the γ region play a significant role in regulating MAPT function. For instance, Radford and colleagues[44] identified NSF as a p-Tau interactor using a proteomic approach that combines antibody-mediated biotinylation and mass spectrometry. Rogers and colleagues[40] reported multiple regulatory elements on genes at 17q21.31, including those on γ (*LRRC37A*, *ARL17B*, and *NSFP1*), supported by ATAC-seq, H3K27ac, and CTCF ChIP-seq data. CRISPR interference experiments further demonstrated that these regulatory elements could influence multiple genes within the H1 and H2 haplotypes, such as *MAPT*.[40] In addition, Hi-C analyses have revealed that *FMNL1*, located more than 650 kb upstream of the *MAPT* promoter, may interact with *MAPT* as well.[40] Together, these findings suggest a complex network of gene interactions within the 17q21.31 region. In future studies, it is important to perform additional functional studies to explore how these structural forms of 17q21.31 affect the complex regulatory dynamics and *MAPT* function, thereby influencing PSP risk. ■

Author Roles: Study design: T.S.C., D.W.D., G.U.H., J.-Y.T., D.H.G., G.D.S., and W.-P.L. Sample collection, brain biospecimens, and neuropathological examinations: T.S.C., C.M., L.M.-P, A.R., P.P.D.D., N.L.B., M.G., L.D.K., J.C.V.S., E.D., B.F.G., K.L.N., C.T., J.G.d.Y., A.R.-G., T.M., W.H.O., G.R., M.S., T.A., S.R., U.M., F.H., P.P., A.B., A.D., I.L.B., T.G.B., G.E.S., L.-N.H., I.L., R.R., O.A.R., D.G., A.L.B., B.L.M., W.W.S., V.M.V.D., E.B.L., C.L.W., H.R.M., R.d.S., J.F.C., A.M.G., J.S.F., G.C., C.D., and D.H.G. Genotype or phenotype acquisition: H.W., T.S.C., V.P., L.V.-B., K.F., A.C.N., L.-S.W., D.H.G., G.D.S., and W.-P.L. Variant detection and variant quality check: H.W., T.S.C., V.P., L.V.-B., K.F., Y.Y.L., and W.-P.L. Statistical analyses and interpretation of results: H.W., Y.-Q.S., A.T., C.L., T.S.C., K.F., A.C.N., J.-Y.T., D.H.G., G.D.S., and W.-P.L. Experimental validation: B.A.D. and P.-L.C. Draft of the manuscript: H.W., G.D.S., and W.-P.L. All authors read, critically revised, and approved the manuscript.

Acknowledgments: This project was 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). We also thank Drs Murray Grossman and Hans Kretzschmar for their valuable contribution to this work. The acknowledgement of PSP

cohorts is listed below, whereas the acknowledgement of ADSP cohorts for control samples can be found in the Supplementary Materials.
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 the National Institute on Aging (NIA) grants P50 AG016574, R01 AG032990, U01 AG046139, R01 AG018023, U01 AG006576, U01 AG006786, R01 AG025711, R01 AG017216, R01 AG003949, National Institute of Neurological Disorders and Stroke (NINDS) grant R01 NS080820, CurePSP Foundation, and support from the 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 NINDS (U24 NS072026 National Brain and Tissue Resource for Parkinson's Disease and Related Disorders), the NIA (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 grants P01AG017586 and P01AG066597.
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, eg, brain tissue, cerebrospinal fluid). The Brain and Body Donation Program is supported by the NINDS (U24 NS072026 National Brain and Tissue Resource for Parkinson's Disease and Related Disorders), the NIA (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: Thanks to the AL-108-231 investigators. A list of the investigators appears in the Appendix.

Financial Disclosures: L.M.-P. received income from Biogen as a consultant in 2022. G.R. has been employed by Roche (Hoffmann-La Roche, Basel, Switzerland) since 2021. Her affiliation while completing her contribution to this manuscript was German Center for Neurodegenerative Diseases (DZNE), Munich, Germany. T.G.B. is a consultant for Aprinoia Therapeutics and a scientific advisor and stock option holder for Vivid Genomics. H.R.M. is employed by University College London (UCL). In the last 12 months he reports paid consultancy from Roche, Aprinoia, AI Therapeutics, and Amylyx; and lecture fees/honoraria from BMJ, Kyowa Kirin, and the Movement Disorder Society. H. R.M. is a co-applicant on a patent application related to C9ORF72: Method for diagnosing a neurodegenerative disease (PCT/GB2012/052140). G.C. is currently an employee of Regeneron Pharmaceuticals. A.M.G. serves on the scientific advisory board for Genentech and Muna Therapeutics.

Data Availability Statement

The whole genome sequencing data and phenotypic information for all the PSP cases and controls can be accessed through the National Institute on Aging Genetics of Alzheimer's Disease Data Storage Site (NIAGADS, https://www.niagads.org). Copy number calls of α, β, and γ, structural forms of 17q21.31, and MAPT sub-haplotypes for the study subjects will be available through the NIAGADS. Bulk RNA-seq data from temporal cortex and cerebellum can be accessed through AD Knowledge Portal (https://www.synapse.org/#!Synapse:syn20818651). Single-nucleus RNA-seq data from dorsolateral prefrontal cortex can be accessed through AD Knowledge Portal (https://www.synapse.org/#!Synapse:syn31512863).

References

1. Hauw JJ, Daniel SE, Dickson D, et al. Preliminary NINDS neuropathologic criteria for Steele–Richardson–Olszewski syndrome (progressive supranuclear palsy). Neurology 1994;44(11):2015–2019.
2. 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. https://doi.org/10.1007/s00401-020-02158-2
3. Armstrong RA. Visual signs and symptoms of progressive supranuclear palsy. Clin Exp Optom 2011;94(2):150–160. https://doi.org/10.1111/j.1444-0938.2010.00504.x
4. 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.
5. Chen JA, Chen Z, Won H, et al. Joint genome-wide association study of progressive supranuclear palsy identifies novel susceptibility loci and genetic correlation to neurodegenerative diseases. Mol Neurodegener 2018;13(1):1–11.
6. Sanchez-Contreras MY, Kouri N, Cook CN, et al. Replication of progressive supranuclear palsy genome-wide association study identifies SLCO1A2 and DUSP10 as new susceptibility loci. Mol Neurodegener 2018;13(1):1–10.
7. Wang H, Chang TS, Dombroski BA, et al. Whole-genome sequencing analysis reveals new susceptibility loci and structural variants associated with progressive supranuclear palsy. Mol Neurodegener 2024;19(1):61.
8. Wen Y, Zhou Y, Jiao B, Shen L. Genetics of progressive supranuclear palsy: a review. J Parkinsons Dis 2021;11(1):93–105.
9. Borroni B, Agosti C, Magnani E, Di Luca M, Padovani A. Genetic bases of progressive supranuclear palsy: the MAPT tau disease. Curr Med Chem 2011;18(17):2655–2660.
10. Rademakers R, Cruts M, Van Broeckhoven C. The role of tau (MAPT) in frontotemporal dementia and related tauopathies. Hum Mutat 2004;24(4):277–295.
11. Cooper YA, Teyssier N, Dräger NM, et al. Functional regulatory variants implicate distinct transcriptional networks in dementia. Science 2022;377(6608):eabi8654. https://doi.org/10.1126/science.abi8654
12. 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.
13. 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.
14. 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. https://doi.org/10.1016/j.jalz.2016.08.018
15. Beecham GW, Bis JC, Martin ER, et al. The Alzheimer's disease sequencing project: study design and sample selection. Neurol Genet 2017;3(5):e194.
16. 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) 2019;9(8):2447–2461.

17. 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.

18. 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.

19. 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 2023. https://doi.org/10.1101/2023.09.13.23295505

20. 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. https://doi.org/10.1186/1471-2164-15-329

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

22. 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.

23. 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.

24. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in python. J Mach Learn Res 2011;12:2825–2830.

25. 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.

26. 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. https://doi.org/10.1002/mds.27773

27. 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.

28. Lee WP, Choi SH, Shea MG, et al. Association of common and rare variants with Alzheimer's disease in more than 13,000 diverse individuals with whole-genome sequencing from the Alzheimer's Disease Sequencing Project. Alzheimers Dement 2024;20(12):8470–8483.

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

30. R Core Team R. R: a language and environment for statistical computing; Published online 2013. Accessed May 21, 2024 https://apps.dtic.mil/sti/citations/AD1039033.

31. Allen M, Wang X, Serie DJ, et al. Divergent brain gene expression patterns associate with distinct cell-specific tau neuropathology traits in progressive supranuclear palsy. Acta Neuropathol 2018;136(5):709–727. https://doi.org/10.1007/s00401-018-1900-5

32. Allen M, Burgess JD, Ballard T, et al. Gene expression, methylation and neuropathology correlations at progressive supranuclear palsy risk loci. Acta Neuropathol 2016;132(2):197–211. https://doi.org/10.1007/s00401-016-1576-7

33. Allen M, Carrasquillo MM, Funk C, et al. Human whole genome genotype and transcriptome data for Alzheimer's and other neurodegenerative diseases. Sci Data 2016;3(1):1–10.

34. Green GS, Yang H, Fujita M, et al. Cellular dynamics across aged human brains uncover a multicellular cascade leading to Alzheimer's disease. Alzheimers Dement 2023;19(S24):e083212. https://doi.org/10.1002/alz.083212

35. Magis AT, Funk CC, Price ND. SNAPR: a bioinformatics pipeline for efficient and accurate RNA-seq alignment and analysis. IEEE Life Sci Lett 2015;1(2):22–25.

36. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 2010;26(1):139–140.

37. Zheng GX, Terry JM, Belgrader P, et al. Massively parallel digital transcriptional profiling of single cells. Nat Commun 2017;8(1):14049.

38. Hao Y, Stuart T, Kowalski MH, et al. Dictionary learning for integrative, multimodal and scalable single-cell analysis. Nat Biotechnol 2024;42(2):293–304.

39. Giannuzzi G, Siswara P, Malig M, et al. Evolutionary dynamism of the primate LRRC37 gene family. Genome Res 2013;23(1):46–59.

40. Rogers BB, Anderson AG, Lauzon SN, et al. Neuronal MAPT expression is mediated by long-range interactions with cis-regulatory elements. Am J Hum Genet 2024;111(2):259–279.

41. Zhao S, Zhang Y, Gamini R, Zhang B, Von Schack D. Evaluation of two main RNA-seq approaches for gene quantification in clinical RNA sequencing: polyA+ selection versus rRNA depletion. Sci Rep 2018;8(1):4781.

42. Yang L, Duff MO, Graveley BR, Carmichael GG, Chen LL. Genomewide characterization of non-polyadenylated RNAs. Genome Biol 2011;12(2):R16. https://doi.org/10.1186/gb-2011-12-2-r16

43. Arena JE, Weigand SD, Whitwell JL, et al. Progressive supranuclear palsy: progression and survival. J Neurol 2016;263(2):380–389. https://doi.org/10.1007/s00415-015-7990-2

44. Radford RAW, Rayner SL, Szwaja P, et al. Identification of phosphorylated tau protein interactors in progressive supranuclear palsy (PSP) reveals networks involved in protein degradation, stress response, cytoskeletal dynamics, metabolic processes, and neurotransmission. J Neurochem 2023;165(4):563–586. https://doi.org/10.1111/jnc.15796

APPENDIX

Investigators of the PSP Genetics Study Group: Adam L. Boxer, Anthony E. Lang, Murray Grossman, David S. Knopman, Bruce L. Miller, Lon S. Schneider, Rachelle S. Doody, Andrew Lees, Lawrence I. Golbe, David R. Williams, Jean-Cristophe Corvol, Albert Ludolph, David Burn, Stefan Lorenzl, Irene Litvan, Erik D. Roberson, Günter U. Höglinger, Mary Koestler, Clifford 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, Bruce H. Morimoto, Franziska Hopfner, Sigrun Roeber, Jochen Herms, Ulrich Müller, Claire Troakes, Ellen Gelpi, Yaroslau Compta, John C. van Swieten, Alex Rajput.

Supporting Data

Additional Supporting Information may be found in the online version of this article at the publisher's web-site.