

# 1 Whole-Genome Sequencing Analysis Reveals New Susceptibility Loci and 2 Structural Variants Associated with Progressive Supranuclear Palsy

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

16

17 \*These authors contributed equally to this work.

18 #These authors are corresponding authors.

19

20 <sup>1</sup>Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of  
21 Pennsylvania, Philadelphia, PA, USA

22 <sup>2</sup>Penn Neurodegeneration Genomics Center, Perelman School of Medicine, University of  
23 Pennsylvania, Philadelphia, PA, USA

24 <sup>3</sup>Movement Disorders Programs, Department of Neurology, David Geffen School of Medicine,  
25 University of California, Los Angeles, Los Angeles, CA, USA

26 <sup>4</sup>Department of Pathology, Department of Artificial Intelligence & Human Health, Nash Family,  
27 Department of Neuroscience, Ronald M. Loeb Center for Alzheimer's Disease, Friedman Brain,  
28 Institute, Neuropathology Brain Bank & Research CoRE, Icahn School of Medicine at Mount Sinai,  
29 New York, NY, USA.

30 <sup>5</sup>Victorian Brain Bank, The Florey Institute of Neuroscience and Mental Health, Parkville, Victoria,  
31 Australia

32 <sup>6</sup>Alzheimer's disease and other cognitive disorders unit. Neurology Service, Hospital Clínic,  
33 Fundació Recerca Clínic Barcelona (FRCB). Institut d'Investigacions Biomediques August Pi i  
34 Sunyer (IDIBAPS), University of Barcelona, Barcelona, Spain

35 <sup>7</sup>Neurological Tissue Bank of the Biobanc-Hospital Clínic-IDIBAPS, Barcelona, Spain

36 <sup>8</sup>Movement Disorders Program, Division of Neurology, University of Saskatchewan, Saskatoon,  
37 Saskatchewan, Canada

38 <sup>9</sup>Laboratory of Neurochemistry and Behavior, Experimental Neurobiology Unit, University of  
39 Antwerp, Wilrijk (Antwerp), Belgium

40 <sup>10</sup>Department of Neurology, University Medical Center Groningen, NL-9713 AV Groningen,

- 41 Netherlands
- 42 <sup>11</sup>Fujirebio Europe NV, Technologiepark 6, 9052 Gent, Belgium
- 43 <sup>12</sup>Department of Pathology and Laboratory Medicine and Department of Neurology, Emory  
44 University School of Medicine, Atlanta, GA, USA
- 45 <sup>13</sup>Netherlands Brain Bank and Erasmus University, Netherlands
- 46 <sup>14</sup>Department of Pathology and Laboratory Medicine, Indiana University School of Medicine,  
47 Indianapolis, IN, USA
- 48 <sup>15</sup>London Neurodegenerative Diseases Brain Bank, King's College London, London, UK
- 49 <sup>16</sup>Autonomous University of Madrid, Madrid, Spain
- 50 <sup>17</sup>Fundación CIEN (Centro de Investigación de Enfermedades Neurológicas) - Centro Alzheimer  
51 Fundación Reina Sofía, Madrid, Spain
- 52 <sup>18</sup>Department of Neurology, Philipps-Universität, Marburg, Germany
- 53 <sup>19</sup>German Center for Neurodegenerative Diseases (DZNE), Munich, Germany
- 54 <sup>20</sup>Parkinson's disease and Movement Disorders Department, HYGEIA Hospital, Athens, Greece
- 55 <sup>21</sup>European University of Cyprus, Nicosia, Cyprus
- 56 <sup>22</sup>Department of Psychiatry and Psychotherapy, University Hospital Munich,  
57 Ludwig-Maximilians-University Munich, Germany
- 58 <sup>23</sup>Center for Neuropathology and Prion Research, Ludwig-Maximilians-University Munich,  
59 Germany
- 60 <sup>24</sup>German Brain Bank, Neurobiobank Munich, Germany
- 61 <sup>25</sup>Unit of Neurodegenerative diseases, Department of Neurology, University Hospital Germans Trias  
62 i Pujol, Badalona, Barcelona, Spain
- 63 <sup>26</sup>Neurosciences, The Germans Trias i Pujol Research Institute (IGTP) Badalona, Badalona, Spain
- 64 <sup>27</sup>Sorbonne Université, Paris Brain Institute – Institut du Cerveau – ICM, Inserm U1127, CNRS  
65 UMR 7225, APHP - Hôpital Pitié-Salpêtrière, Paris, France
- 66 <sup>28</sup>Banner Sun Health Research Institute, Sun City, AZ, USA
- 67 <sup>29</sup>University McGill, Montreal, Quebec, Canada
- 68 <sup>30</sup>Department of Neuroscience, University of California, San Diego, CA, USA
- 69 <sup>31</sup>VIB Center for Molecular Neurology, University of Antwerp, Belgium
- 70 <sup>32</sup>Department of Neuroscience, Mayo Clinic Jacksonville, FL, USA
- 71 <sup>33</sup>Memory and Aging Center, University of California, San Francisco, CA, USA
- 72 <sup>34</sup>Penn Center for Neurodegenerative Disease Research, University of Pennsylvania School of  
73 Medicine, Philadelphia, PA, USA
- 74 <sup>35</sup>University of Texas Southwestern Medical Center, Dallas, TX, USA
- 75 <sup>36</sup>Departamento of Clinical and Movement Neuroscience, University College of London, London, UK
- 76 <sup>37</sup>Reta Lila Weston Institute, UCL Queen Square Institute of Neurology, London, UK.
- 77 <sup>38</sup>Department of Genetics and Genomic Sciences, New York, NY, USA; Icahn School of Medicine at  
78 Mount Sinai, New York, NY, USA
- 79 <sup>39</sup>Friedman Bioventure, Inc., Del Mar, CA, USA
- 80 Department of Genetics and Genomic Sciences, New York, NY, USA
- 81 <sup>40</sup>Department of Psychiatry, Semel Institute for Neuroscience and Human Behavior, University of  
82 California, Los Angeles, CA, USA

83 <sup>41</sup>Department of Biostatistics, Epidemiology, and Informatics, Perelman School of Medicine,  
84 University of Pennsylvania, Philadelphia, PA, USA

85 <sup>42</sup>Department of Neurology, LMU University Hospital, Ludwig-Maximilians-Universität (LMU)  
86 München; German Center for Neurodegenerative Diseases (DZNE), Munich, Germany; and  
87 Munich Cluster for Systems Neurology (SyNergy), Munich, Germany

88 <sup>43</sup>Department of Human Genetics, David Geffen School of Medicine, University of California, Los  
89 Angeles, Los Angeles, CA, USA

90 <sup>44</sup>Institute of Precision Health, University of California, Los Angeles, Los Angeles, CA, USA  
91

92 **Key words:** Progressive Supranuclear Palsy (PSP), Whole-Genome Sequencing (WGS),  
93 Genome-Wide Association Study (GWAS), Structural Variants (SVs), Apolipoprotein E (APOE)

94 **Abstract**

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

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

104 **Results:** Our analysis of common SNVs and indels confirmed known genetic loci at *MAPT*, *MOBP*,  
105 *STX6*, *SLCO1A2*, *DUSP10*, and *SPI*, and further uncovered novel signals in *APOE*, *FCHO1/MAP1S*,  
106 *KIF13A*, *TRIM24*, *TNXB*, and *ELOVL1*. Notably, in contrast to Alzheimer's disease (AD), we  
107 observed the *APOE*  $\epsilon 2$  allele to be the risk allele in PSP. Analysis of rare SNVs and indels identified  
108 significant association in *ZNF592* and further gene network analysis identified a module of neuronal  
109 genes dysregulated in PSP. Moreover, seven common SVs associated with PSP were observed in the  
110 H1/H2 haplotype region (17q21.31) and other loci, including *IGH*, *PCMT1*, *CYP2A13*, and *SMCP*.  
111 In the H1/H2 haplotype region, there is a burden of rare deletions and duplications ( $P = 6.73 \times 10^{-3}$ ) in  
112 PSP.

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

## 115 **Background**

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

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

136 [15].

137 To date, no comprehensive analysis of single nucleotide variants (SNVs), small insertions and  
138 deletions (indels), and structural variants (SVs) in PSP by whole genome sequencing has been  
139 conducted. To gain a more comprehensive understanding of the genetic underpinnings of PSP, we  
140 performed whole genome sequencing (WGS) and analyzed SNVs, indels and SVs. As a result, we  
141 not only validated previously reported genes but also unveiled new loci that provide novel insights  
142 into the genetic basis of PSP.

## 143 **Methods**

### 144 Study subjects

145 We performed WGS at 30x coverage (**Table S1**) for 1,834 PSP cases and 128 controls from  
146 the PSP-NIH-CurePSP-Tau, PSP-CurePSP-Tau, PSP-UCLA, and AMPAD-MAYO cohorts included  
147 in Alzheimer's Disease Sequencing Project (ADSP, NG00067.v7) and used 3,008 controls from other  
148 cohorts in ADSP [16]. Control subjects were self-identified as non-Hispanic white. WGS data is  
149 available on The National Institute on Aging Genetics of Alzheimer's Disease Data Storage Site  
150 (NIAGADS) [17]. We removed related subjects (identify by descent  $> 0.25$ ), five clinically  
151 diagnosed PSP who were not found to have PSP on autopsy, and non-Europeans (subjects that were  
152 eight standard deviations away from the 1000 Genomes Project European samples [18,19] using the  
153 first six principal components (PCs)), resulting in 1,718 individuals with PSP and 2,944 control  
154 subjects. Of the 1,718 PSP individuals, 1,441 were autopsy-confirmed and 277 were clinically  
155 diagnosed (**Table 1**).

156 Considering that our sample set incorporated external controls from ADSP, initially collected  
157 for Alzheimer's Disease (AD) studies, there was a potential selection biases for *APOE*  $\epsilon 4$  and  $\epsilon 2$  in  
158 controls. To rigorously validate our findings linked to *APOE*, we broke down the allele frequencies  
159 of *APOE*  $\epsilon 4$  and  $\epsilon 2$  by cohorts (**Table S2**), reviewed the study design of each cohort, and created an  
160 additional sample set by excluding those cohorts with selection bias against *APOE*  $\epsilon 4$  or  $\epsilon 2$   
161 (**Supplementary Methods**).

162

### 163 SNVs/indels quality controls

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

175

#### 176 Common SNVs/indels analysis

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

183 For H1/H2 region, fine-mapping were analyzed using SuSie [26]. We ran the analysis several  
184 times assuming the number of maximum causal variants were from 2 to 10. The only variant



185 (rs242561) robust to the choice of maximum causal variants was reported. To avoid potential  
186 confounding effects (particularly for *APOE* alleles), we also performed association analysis (**Table**  
187 **S4, Table S5**) for suggestive and genome-wide significant signals when excluding subjects from the  
188 three cohorts with selection bias against *APOE* alleles (ADSP-FUS1-APOEextremes,  
189 ADSP-FUS1-StEPAD1, and CacheCounty) along with cohorts with less than 10 subjects  
190 (NACC-Genentech, FASe-Families-WGS, and KnightADRC-WGS) (**Table S2**). We also performed  
191 additional experimental validation using TaqMan assay/Sanger sequencing to confirm the genotype  
192 of *APOE* observed from WGS (**Supplementary Methods, Table S6**).

193

#### 194 Rare SNVs/indels analysis

195 For aggregated tests of rare variants, we considered rare protein truncating variants (PTVs)  
196 and PTVs/damaging missense variants. Variant were annotated with ANNOVAR (version  
197 2020-06-07) [27] and Variant Effect Predictor (VEP, version 104.3) [28]. PTVs were in protein  
198 coding genes (Ensembl version 104) [29] and had VEP consequence as stop gained, splice acceptor,  
199 splice donor or frameshift. Damaging missense variants were in protein coding genes (Ensembl  
200 version 104) and had a VEP consequence as missense, CADD score  $\geq 15$ , and PolyPhen-2 HDIV of  
201 probably damaging. Rare variants were selected based on a MAF  $< 0.01\%$  from gnomAD and a  
202 MAF  $< 1\%$  in our dataset. The number of alternative allele variants in PTVs and PTVs/damaging  
203 missense variants was similar across sequencing centers and when evaluated for loss of function  
204 intolerant genes (observed/expected score upper confidence interval  $< 0.35$  [30]) (**Fig. S14**)

205 We performed SKAT-O and gene burden testing (SKATBinary, method='burden') for PTVs

206 and PTV/damaging missense variants (**Supplementary Methods**). We also considered only PTVs or  
207 PTVs/missense variants in loss of function intolerant genes (observed/expected score upper  
208 confidence interval  $< 0.35$ [30]) when performing the tests.  $P$ -values were FDR corrected for the  
209 number of genes with a total minor allele count (MAC)  $\geq 10$ . As SKAT-O does not calculate an odds  
210 ratio, we calculated the odds ratio of significant genes using logistic regression with the same  
211 covariates as SKAT-O and burden testing, and the same variant weights.

212 We evaluated the C1 module, a gene set, which was previously shown to be composed of  
213 neuronal genes and enriched for common variants in PSP [31]. We performed a permutation test  
214 ( $N=1000$ ) of random gene set modules from brain expressed genes that contained the same number  
215 of genes as C1. From the human protein atlas ([www.proteinatlas.org](http://www.proteinatlas.org)) [32], brain expressed genes  
216 were defined as the union of unique proteins from the cerebral cortex, basal ganglia and midbrain  
217 ( $N=15,638$ ). We calculated SKAT-O  $P$ -values from these random gene modules to determine the null  
218 distribution. We calculated the unadjusted odds ratio of significant genes or gene sets by summing  
219 the number of alternate alleles in the gene set among the total number alleles in cases and controls.  
220 Normalized quantification (TPM) gene expression across tissues was obtained from Genotype-Tissue  
221 Expression (GTEx) [33]. The expression of ZNF592 and C1 module (summarized as an eigengene  
222 [34]) were plotted.

223

#### 224 *SV detection and filtering*

225 For each sample, SVs were called by Manta (v1.6.0) [35] and Smoove (v0.2.5) [36] with default  
226 parameters. Calls from Manta and Smoove were merged by Svimmer [37] to generate a union of two  
227 call sets for a sample. Then, all individual sample VCF files were merged together by Svimmer as

228 input to GraphTyper2 (v2.7.3) [37] for joint genotyping. SV calls after joint-genotyping are  
229 comparable across the samples, therefore, can be used directly in genome-wide association analysis  
230 [37]. A subset of SV calls was defined as high-quality calls [37]. Details of SV calling pipeline were  
231 in our previous study [38]. For each individual SV reported, Samplot [39] or IGV [40] were used to  
232 keep only high-confident CNVs and inversions that are supported by read depth or split reads; for  
233 insertions, we kept high-confident insertions that are high-quality and not in the masked regions  
234 (**Supplementary Methods**).

235

### 236 SV analysis

237 For SV association, more strict sample filtering was applied: outlier samples with too many  
238 (larger than median + 4\*MAD) CNV/insertion calls or too little (smaller than median - 4\*MAD)  
239 high-quality CNV/insertion calls were removed. There were 4,432 samples (1,703 cases and 2,729  
240 controls) remaining for PSP SV association analysis. Due to more false positives being picked up,  
241 the genomic inflation would be high ( $\lambda = 1.89$ , **Fig. S9**) if all SVs were included in the analysis.  
242 Therefore, we restricted our analysis to high-quality SVs only, making the genomic inflation drop to  
243 1.27 (**Fig. S9**). The 14,792 high-quality common SVs (MAF > 0.1) with call rate > 0.5 were included  
244 in the analysis. Mixed model implemented in R Genesis were used for association. Sex, PCR  
245 information, SV PCs 1-5, and SNV PCs 1-5 were adjusted in the mixed model. After association, we  
246 manually inspect deletions, duplications, and inversions by Samplot or IGV to keep only those with  
247 support from read depth, split read or insert size. For insertions, those not on masked regions were  
248 reported.

249 For SVs inside the H1/H2 region, all SVs those that are not high-quality are included. Then, we  
250 removed SVs with missing rate  $> 0.5$  and manual inspect deletions, duplications, and inversions by  
251 Samplot or IGV to keep only those with support from read depth, split read or insert size. For  
252 insertions, those high-quality ones not on masked regions were kept for analysis. LD between SVs  
253 was calculated using PLINK (V1.90 beta) [41]. Rare SV burden on H1/H2 region was evaluated by  
254 SKAT-O [42] adjusting for gender and PCs 1-5. As SKAT-O does not calculate an odds ratio, we  
255 calculated the odds ratio using logistic regression with the same covariates.

## 256 **Results**

### 257 Common SNVs and indels associated with PSP

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

266

### 267 *MAPT*, *MOBP* and *STX6*

268 In the *MAPT* region, a multitude of SNVs and indels in high linkage disequilibrium (LD) with  
269 the H1/H2 haplotype remains the most significant association with PSP (**Fig. S2A**). From our  
270 analysis, the prominent signal within the *MAPT* region is rs62057121 ( $P = 7.45 \times 10^{-78}$ ,  $\beta = -1.32$ ,  
271 MAF = 0.15). Fine mapping suggests that rs242561 ( $P = 4.49 \times 10^{-74}$ ,  $\beta = -1.23$ , MAF = 0.16) is  
272 likely to be a causal SNV underlying the statistical significance. The SNP rs242561 is located in an  
273 enhancer region, containing an antioxidant response element that binds with NRF2/sMAF protein  
274 complex. The T allele of rs242561 showed a stronger binding affinity for NRF2/sMAF in ChIP-seq  
275 analysis, therefore inducing a significantly higher transactivation of the *MAPT* gene [43]. rs242561  
276 and rs62057151 were both in high LD ( $r^2 > 0.9$ ) with H1/H2 (defined by the 238 bp deletion in

277 *MAPT* intron 9) and represented the same association signal as the H1/H2. However, in previous  
278 studies [8,44], the H1c tagging SNV (rs242557) inside the H1/H2 region was found to be significant  
279 when conditioning on H1/H2. We confirmed that rs242557 was genome-wide significant after  
280 adjusting for H1/H2 ( $P = 3.68 \times 10^{-15}$ ,  $\beta = 0.39$ , MAF = 0.42) though in weak LD with H1/H2 ( $r^2 =$   
281 0.14). To pinpoint the causal genes underlying the association in H1/H2 requires additional  
282 functional study. For example, Cooper *et al.* [11] analyzed transcriptional regulatory activity of  
283 SNVs and suggested *PLEKHM1* and *KANSL1* were probable causal genes in H1/H2 besides *MAPT*.  
284 In *MOBP* (rs11708828,  $P = 7.04 \times 10^{-12}$ ,  $\beta = -0.35$ , MAF = 0.46, **Fig. S2B**) and *STX6* (rs10753232,  
285  $P = 6.79 \times 10^{-10}$ ,  $\beta = 0.31$ , MAF = 0.44, **Fig. S2C**), the associated variants were of high allele  
286 frequency and exhibited moderate effect size.

287

#### 288 *APOE* and risk of PSP

289 One newly identified significant locus from our analysis is the well-known AD risk gene, *APOE*.  
290 We observed a significant association between the *APOE*  $\epsilon 2$  haplotype and an elevated risk of PSP  
291 ( $P = 9.57 \times 10^{-16}$ ,  $\beta = 0.87$ , MAF = 0.06, **Table 3, Fig. S3B**). The *APOE*  $\epsilon 2$  haplotype is encoded by  
292 rs429358-T and rs4712-T, which is considered a protective allele in AD. The increased risk of *APOE*  
293  $\epsilon 2$  in PSP has been previously reported in a Japanese cohort, albeit with a relatively small sample  
294 size [45]. Furthermore, Zhao *et al.* [46] confirmed that *APOE*  $\epsilon 2$  is linked to increased tau pathology  
295 in the brains of individuals with PSP and reported a higher frequency of homozygosity of *APOE*  $\epsilon 2$   
296 in PSP with an odds ratio of 4.41. Consistent with these findings, our dataset exhibited a higher  
297 frequency of homozygosity of rs7412-T in PSP, yielding an odds ratio of 3.91.

298 For *APOE*  $\epsilon 4$  allele, contrary to its association with AD, we observed that rs429358-C exhibits a  
299 protective effect against PSP ( $P = 5.71 \times 10^{-18}$ ,  $\beta = -0.60$ , MAF = 0.16, **Table 3**). The lead SNV  
300 demonstrating this protective association from our analysis is rs4420638 ( $P = 2.91 \times 10^{-19}$ ,  $\beta = -0.57$ ,  
301 MAF = 0.20, **Fig. S3A**), which is in LD ( $r^2 = 0.74$ ) with rs429358. In a previous PSP GWAS  
302 conducted by Hoglinger *et al.* [8], another *APOE*  $\epsilon 4$  tagging SNV (rs2075650,  $r^2 = 0.52$  with  
303 rs429358) was also found to be diminished (MAF<sub>case</sub> = 0.11 and MAF<sub>control</sub> = 0.15) in PSP,  
304 although not reaching significance ( $P = 1.28 \times 10^{-5}$ ). Notably, in our analysis, rs2075650 reached  
305 genome-wide significance ( $P = 3.39 \times 10^{-13}$ ,  $\beta = -0.51$ , MAF = 0.15). *APOE*  $\epsilon 4$  or  $\epsilon 2$  displayed an  
306 independent effect for PSP risk without a significant epistatic interaction with H1/H2 haplotype ( $P >$   
307 0.05) (**Fig. S4**).

308 Given that our dataset included external controls from ADSP collected for Alzheimer's disease  
309 studies, there were a potential selection biases for *APOE*  $\epsilon 4$  and  $\epsilon 2$  in controls. To address this  
310 concern, we broke down the allele frequencies of *APOE*  $\epsilon 4$  and  $\epsilon 2$  by cohorts (**Table S2**) and  
311 indicated cohorts with potential selection bias. The association analysis excluding these cohorts  
312 shows the  $\epsilon 2$  SNV (rs7412,  $P = 1.23 \times 10^{-12}$ ,  $\beta = 0.70$ , MAF = 0.06) remained genome-wide  
313 significant and  $\epsilon 4$  SNV (rs429358,  $P = 0.02$ ,  $\beta = -0.16$ , MAF = 0.14) was nominal significant (**Table**  
314 **S4, Table S5**).

315

#### 316 *Suggestive significant loci*

317 Eight loci were suggestive of significance in our analysis of which three, *SLCO1A2*, *DUSP10*,  
318 and *SPI*, were previously reported [12,13]. In *SLCO1A2*, the lead SNV rs74651308 ( $P = 2.86 \times 10^{-7}$ ,

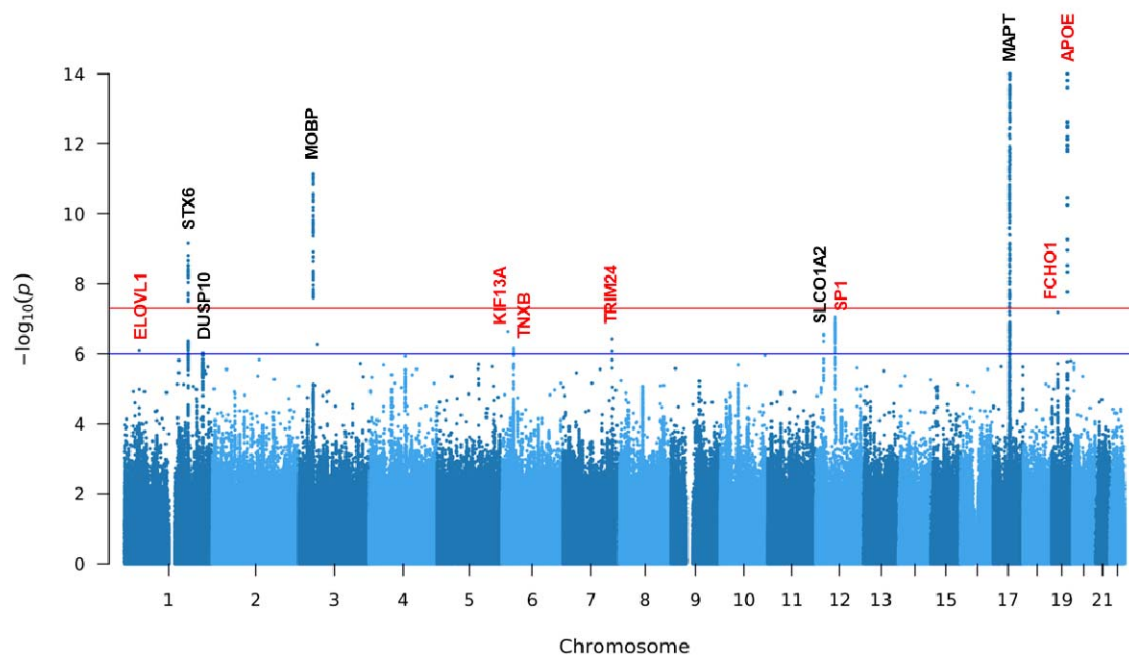
319  $\beta = 0.51$ , MAF = 0.07, **Fig. S5A**) is intronic and in LD ( $r^2 = 0.98$ ) with missense SNV rs11568563 ( $P$   
320  $= 1.45 \times 10^{-6}$ ,  $\beta = 0.47$ , MAF = 0.07), which was reported in a previous study [12]. About 250 kb  
321 upstream of *DUSP10* lies the previously reported SNV rs6687758 [12] ( $P = 3.36 \times 10^{-6}$ ,  $\beta = 0.29$ ,  
322 MAF = 0.21), which is in LD ( $r^2 = 0.98$ ) with the lead SNV rs12026659 in our analysis ( $P = 9.48 \times$   
323  $10^{-7}$ ,  $\beta = 0.31$ , MAF = 0.21, **Fig. S5B**). In *SPI*, the reported indel rs147124286 [13] ( $P = 4.39 \times 10^{-7}$ ,  
324  $\beta = -0.35$ , MAF = 0.16) is in LD ( $r^2 = 0.995$ ) with the lead SNV rs12817984 ( $P = 8.91 \times 10^{-8}$ ,  $\beta =$   
325  $-0.37$ , MAF = 0.16, **Fig. S5C**). Notably, disruption of a transcriptional network centered on *SPI* by  
326 causal variants has been implicated previously in PSP [11].

327 Five newly discovered suggestive loci are in *FCHO1/MAP1S*, *KIF13A*, *TRIM24*, *TNXB*, and  
328 *ELOVL1*. Within *FCHO1/MAP1S*, the most significant signal (rs56251816,  $P = 6.57 \times 10^{-8}$ ,  $\beta = 0.35$ ,  
329 MAF = 0.22, **Fig. S6A**) is in the intron of *FCHO1*. rs56251816 is a significant expression  
330 quantitative trait locus (eQTL) for both *FCHO1* and *MAP1S* (13 kb upstream of *FCHO1*) in the  
331 Genotype-Tissue Expression (GTEx) project [47]. *MAP1S* encodes a microtubule associated protein  
332 that is involved in microtubule bundle formation, aggregation of mitochondria and autophagy [48],  
333 and therefore, is more relevant than *FCHO1* regarding PSP. *KIF13A*, which encodes a  
334 microtubule-based motor protein was also suggestive of significance (rs4712314,  $P = 2.37 \times 10^{-7}$ ,  $\beta$   
335  $= 0.27$ , AF = 0.51, **Fig. S6B**). The significance in genes involved in microtubule-based processes,  
336 such as *MAPT*, *MAP1S* and *KIF13A*, implicates the neuronal cytoskeleton as a convergent aspect of  
337 PSP etiology.

338 Other variants with suggestive association evidence include *TRIM24* (rs111593852,  $P = 3.75 \times$   
339  $10^{-7}$ ,  $\beta = 0.87$ , MAF = 0.02, **Fig. S7A**). *TRIM24* is involved in transcriptional initiation and shows



340 differential expression in individuals with Parkinson disease [49,50]. Another suggestive locus is  
341 *TNXB*, located in the major histocompatibility complex (MHC) region on chromosome 6, with the  
342 lead SNV rs367364 ( $P = 7.07 \times 10^{-7}$ ,  $\beta = -0.37$ , MAF = 0.13, **Fig. S7B**). Finally, *ELOVL1* yields  
343 suggestive evidence of association (rs839764,  $P = 7.94 \times 10^{-7}$ ,  $\beta = 0.27$ , MAF = 0.41, **Fig. S7C**).  
344 This gene encodes an enzyme that elongates fatty acids and can cause a neurological disorder with  
345 ichthyotic keratoderma, spasticity, hypomyelination and dysmorphic features [51]. Furthermore, we  
346 found a few SNV/indels that reached genome-wide or suggestive significance without other  
347 supporting variants in LD (**Fig.S1, Table S3**). These signals could be due to sequencing errors and  
348 need further experimental validation.



349

350 **Fig. 1: Manhattan plot of SNVs/indels for PSP.**

351 Loci with a suggestive or genome-wide significant signal are annotated (novel loci in red and known  
352 loci in black). Variants with a  $P$ -value below  $1 \times 10^{-14}$  are not shown. The red horizontal line  
353 represents genome-wide significance level ( $5 \times 10^{-8}$ ). The blue horizontal line represents suggestive  
354 significance level ( $1 \times 10^{-6}$ ).

355 **Table 1. Characteristics of study participants.**

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

<sup>a</sup>APOE ε4 is represented by the genotypes of rs429358-C.

<sup>b</sup>APOE ε2 is represented by the genotypes of rs7412-T.

<sup>c</sup>H2 haplotype is determined by the genotypes of rs8070723-G.

SD, standard deviation.

356 **Table 2. Genome-wide and suggestive significant loci.**

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

Chr, chromosome; Ref, reference allele; Alt, alternative allele; AF, allele frequency.

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

<sup>a</sup>SNVs with significant eQTL hits.

---

<sup>b</sup>SNVs with significant sQTL hits.

<sup>c</sup>SNVs with both eQTL and sQTL hits.

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

Studies	rs429358		rs7412	
	AF (Case)	AF (Control)	AF (Case)	AF (Control)
PSP WGS (This study)	0.1279	0.1742	0.0844	0.0414
PSP GWAS [52]	0.1159	0.1366	0.0826	0.0794
1000 Genomes Project [18]		0.1512		0.0771
ExAC		0.2078		0.1060
European (non-Finnish) [53]				
gnomAD V4		0.1506		0.0783
European (non-Finnish) [54]				
TOPMed Freeze 8		0.1501		0.0752
NFE (Non-Finnish European)				
ADSP R3 Non-Hispanic White [55]	0.3139 (AD as cases)	0.1803	0.0244 (AD as cases)	0.0406

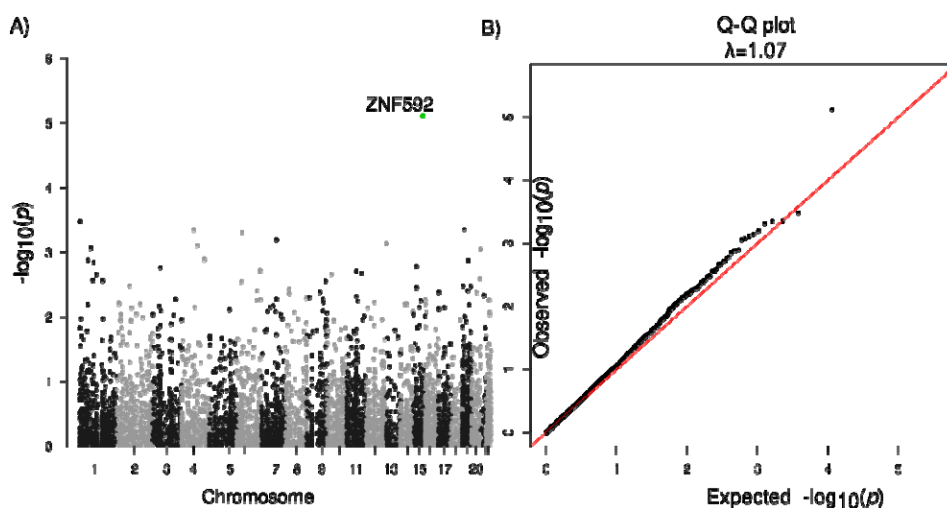
358

359 Rare SNVs/indels and network analysis

360 The heritability of PSP for common SNVs and indels (MAF > 0.01) was estimated to be 20%,  
361 while common plus rare SNVs/indels was estimated to be 23% from our analysis using  
362 GCTA-LDMS [22]. Therefore, we performed aggregated tests for rare SNVs and indels, and  
363 identified *ZNF592* (SKAT-O FDR=0.043, burden test FDR=0.041) with an of OR = 1.08 (95% CI:  
364 1.008-1.16) (**Fig. 2, Table 4, Table S7**) for protein truncating or damaging missense variants. There  
365 was no genomic inflation with a  $\lambda=1.07$  (**Fig. 2**). Risk in *ZNF592* was imparted by 16 unique  
366 variants, with one splice donor and 15 damaging missense variants (**Table S7**). *ZNF592* has not been  
367 previously associated with PSP but showed moderate RNA expression in the cerebellum compared to  
368 other tissues from GTEx (**Fig. S8**). There were no significant genes identified when evaluating PTVs  
369 only or when restricting to loss of function intolerant genes.

370 Considering that genes do not operate along, but rather within signaling pathways and networks,

371 we and others have shown that better understanding of disease mechanisms can be achieved through  
372 gene network analysis [56–58]. Therefore, we scrutinized rare variants within a network framework,  
373 focusing on co-expression network analysis performed in PSP post mortem brain that had previously  
374 identified a brain co-expression module, C1, which was conserved at the protein interaction level and  
375 enriched for common variants in PSP [31]. We found this C1 neuronal module was significantly  
376 enriched with PSP rare variants ( $P = 0.006$ , OR [95% CI] = 1.31 [1.01-1.70], **Table 4; Table S8**).  
377 Genes from the C1 module were more likely to be loss of function intolerant compared to the  
378 background of all brain expressed genes (**Fig. S8**). To ensure that this was association not spurious,  
379 we performed permutation testing using random gene modules of brain expressed genes with the  
380 same number of genes as C1. The C1 module remains significant (Permutation  $P = 0.078$ ). Exploring  
381 GTEx, we found that C1 genes are highly expressed in brain tissues including the cerebellum, frontal  
382 cortex, and basal ganglia (**Fig. S8**), consistent with regions affected in this disorder.  
383



384  
385 **Fig. 2: Association analysis of rare SNVs/indels.**  
386 **A.** Manhattan plot for genes with protein truncating variants or damaging missense variants. **B.** Q-Q  
387 plot of gene  $P$ -values with protein truncating variants or damaging missense variants.

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

Gene	Variants	Total MAC	Case MAC	Control MAC	Fraction Case	Fraction Control	OR (95% CI)	SKAT-O		Burden	
								FDR	P	FDR	P
ZNF592	16	19	8	11	0.0023	0.0018	1.08 (1.01-1.16)	0.044	7.60×10 <sup>-6</sup>	0.041	7.30×10 <sup>-06</sup>
Module	Variants	Total MAC	Case MAC	Control MAC	Fraction Case	Fraction Control	OR (95% CI)	Permutation test	P	Permutation test	P
C1	180	234	101	133	0.029	0.022	1.31 (1.01-1.70)	0.19	0.048	0.078	0.006

389 SVs associated with PSP

390 Seven high-confident SVs achieved genome-wide significance with PSP (**Table 5, Fig. S9**),  
391 including three deletions tagging the H2 haplotype. The most significant signal is a 238 bp deletion  
392 in *MAPT* intron 9 (**Fig. S10A**, chr17:46009357-46009595,  $P = 3.14 \times 10^{-50}$ , AF = 0.16) that has been  
393 reported on the H2 haplotype [59,60] and is in LD ( $r^2 = 0.99$ ) with the lead SNV, rs62057121  
394 (chr17:45823394,  $P = 7.45 \times 10^{-78}$ ,  $\beta = -1.32$ , MAF = 0.15), in the *MAPT* region. Adding to this, two  
395 other deletions, one spanning 314 bp (**Fig. S10B**, chr17:46146541-46146855, AF = 0.19) and the  
396 other covering 323 bp (**Fig. S10C**, chr17:46099028-46099351, AF = 0.22), both are Alu elements  
397 and in LD ( $r^2 > 0.8$ ) with the top signal (the 238 bp deletion). This observation indicates that  
398 transposable elements may play an important role in the evolution of H1/H2 haplotype structure.

399 Beyond the identified SVs in the H1/H2 region, we uncovered a significant deletion  
400 (chr14:105864208-105916743,  $P = 4.74 \times 10^{-14}$ , AF = 0.01) within the immunoglobulin heavy locus  
401 (*IGH*), which is a complex SV region (**Fig. S11**) related to antigen recognition. Moreover, a 619 bp  
402 deletion (chr6:149762615-149763234,  $P = 8.60 \times 10^{-12}$ , AF = 0.55; **Fig. S10D**) in *PCMT1* displayed  
403 increased risk of PSP with an odds ratio of 4.19. The odds ratio increased to 8.38 when comparing  
404 1,244 individuals with homozygous deletions in *PCMT1* with the rest of sample set. *PCMT1* encodes  
405 a type  $\square$  class of protein carboxyl methyltransferase enzyme that is highly expressed in the brain [61]  
406 and is able to ameliorate  $A\beta_{25-35}$  induced neuronal apoptosis [62,63]. Additionally, we found a  
407 deletion between *CYP2F1* and *CYP2A13* (chr19:41102802-41104285, AF = 0.17) and an insertion in  
408 *SMCP* (chr1:152880979-152880979, AF = 0.74) which were also significant (**Table 5**). The 1.5 kb  
409 deletion (chr19:41102802-41104285) almost completely overlaps the SINE-VNTR-Alus (SVA)



410 transposon region annotated by RepeatMasker [64].

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

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

\*Represents SVs with DNA samples available and PCR validated

412 *SVs in H1/H2 haplotype region*

413 The H1/H2 region stands out as the pivotal genetic risk factor for PSP [8,65]. The H2 haplotype  
414 exhibits a reduced odds ratio of 0.19, as we observed the allele frequency of the 238 bp H2-tagging  
415 deletion is 23% in PSP and only 5% in control ( $P < 2.2 \times 10^{-16}$ ). Moreover, our analysis pointed out  
416 five common (MAF > 0.01) and 12 rare deletions and duplications in the region (**Table 6**), ranging  
417 from 88 bp to 47 kb. Additionally, one common and four rare high-confidence insertions were  
418 reported in the region.

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

423 Among the 12 rare deletions and duplications (**Fig. S13**), five are located in potentially  
424 functional regions, such as splice sites, exons, and transcription factor binding sites (**Table 6**).  
425 Particularly, one deletion (chr17:45993882-45993970) in exon 9 of *MAPT* was identified in a PSP  
426 patient, adding to previous reports of exonic deletions in the *MAPT* in frontotemporal dementia, such  
427 as deletion of exon 10 [66] and exons 6-9 [67] in *MAPT*. Using the SKAT-O test ( $N = 4,432$ ), the 12  
428 rare CNVs displayed a significantly higher burden in PSP than controls ( $P = 0.01$ , OR = 1.64).–

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

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

---

AF, allele frequency; N, number of individuals with non-missing genotypes.

\*High-quality SVs that were included in association analysis.

<sup>a</sup>Represents SVs with DNA samples available and PCR validated.

430

## 431 Discussion

432 Through comprehensive analysis of whole genome sequence, we identified SNVs, indels and  
433 SVs contributing to the risk of PSP. For common SNVs, previously reported regions, including  
434 *MAPT*, *MOBP*, *STX6*, *SLCO1A2*, *DUSP10*, and *SPI* [8,12,13] were replicated in our analysis and  
435 novel loci in *APOE*, *FCHO1/MAP1S*, *KIF13A*, *TRIM24*, *ELOVL1*, and *TNXB* were discovered.  
436 *EIF2AK3* which was significantly associated with PSP in a previous GWAS [8] did not reach  
437 significance in our study. In the current study, the SNV with the lowest *P* around *EIF2AK3* was  
438 rs13003510 ( $P = 8.30 \times 10^{-5}$ ,  $\beta = 0.22$ , MAF = 0.3).

439 The *APOE*  $\epsilon 4$  haplotype was of particular interest as it is a common risk factor for AD,  
440 explaining more than a 1/3 of population attributable risk [68,69]. Typically, individuals with one  
441 copy of the *APOE*  $\epsilon 4$  allele (rs429358-C and rs4712-G) have approximately a threefold increased  
442 risk of developing AD, while those with two copies of the allele have an approximately a 12-fold  
443 increase in risk [70]. In striking contrast, the  $\epsilon 4$  tagging allele rs429358 was protective in PSP and  
444 the  $\epsilon 2$  tagging allele rs7412 was deleterious. This observation is particularly intriguing since both AD  
445 and PSP have intracellular aggregated tau as a prominent neuropathologic feature. Notably, both  $\epsilon 2$   
446 allele and  $\epsilon 4$  allele have been associated with tau pathology burden in the brain of mice models  
447 [46,71], which raises the question of distinct tau species in 4R-PSP versus 3R-4R-AD. It is also  
448 notable that the  $\epsilon 2$  allele is also associated with increased risk for age-related macular degeneration  
449 (AMD), and the  $\epsilon 4$  allele was associated with decreased risk [72,73]. These results demonstrate that  
450 the same variant may have opposite effects in different degenerative diseases. This is especially  
451 important, given the advent of gene editing as a therapeutic modality, and programs focused on

452 changing *APOE*  $\epsilon 4$  to  $\epsilon 2$ . Although this therapy would likely decrease risk for AD, our results  
453 indicate that it would increase risk for PSP, in addition to AMD. From this standpoint, caution is  
454 warranted in germ-line genome editing until the broad spectrum of phenotypes associated with  
455 human genetic variation is understood.

456 Burden association tests are an highly valuable for addressing sample size limitations in  
457 analyzing rare variants [74]. Indeed, burden testing allowed us to identify *ZNF592*, a classical C2H2  
458 zinc finger protein (ZNF) [75,76], as a candidate risk gene. ZNF proteins have been causative or  
459 strongly associated with large numbers of neurodevelopmental disease [77,78] and  
460 neurodegenerative disease including Parkinson's disease [79] and Alzheimer's disease [80,81].  
461 *ZNF592* was initially thought to be responsible for autosomal recessive spinocerebellar ataxia 5 from  
462 a consanguineous family with neurodevelopmental delay including cerebellar ataxia and intellectual  
463 disability due to a homozygous G1046R substitution [82]. However, further analysis of this family  
464 identified *WDR73* to be the most likely causative gene, consistent with Galloway-Mowat syndrome,  
465 although *ZNF592* may have contributed to the phenotype [83].

466 We also extended classical gene-based burden analysis to consider rare risk burden in the  
467 context of a gene set defined by co-expression networks [31,84]. We leveraged combined previous  
468 proteomic and transcriptomic analysis of post-mortem brain from patients afflicted with PSP, and  
469 showed that rare variants enrich in the C1 neuronal module, which was the same module enriched  
470 with common variants [31]. This, along with our recent work identifying a neuronally-enriched  
471 transcription factor network centered around SP1 disrupted by PSP common genetic risk, suggests  
472 that although PSP neuropathologically is defined by tufted astrocytes and oligodendroglial coiled

473 bodies [6,85,86], initial causal drivers of PSP appear to be primarily neuronal.

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

480 There were limitations to this study. Not all PSP were pathologically confirmed, although  
481 pathological confirmation was available in a significant subset (of the 1,718 PSP individuals, 1,441  
482 were autopsy-confirmed and 277 were clinically-diagnosed). Additionally, the majority of control  
483 samples in this study were from ADSP and were initially collected as controls for AD studies. As  
484 ADSP is a dataset composed of multiple cohorts from diverse sources, it is imperative to ensure that  
485 any observed allele frequency differences between controls and cases can be attributed to the disease  
486 itself rather than sample selection biases arising from technical artifacts or batch effects. To mitigate  
487 the risk of false reports, we meticulously examined the allele frequencies of both cases and controls,  
488 especially in relation to novel and significant signals.

489 This work represents an important first step; future work is necessary to further delineate the rare  
490 genetic risk in PSP harbored in coding and noncoding regions. These results may come to fruition as  
491 additional genomic analytical methods are developed, sample size increased, and orthogonal  
492 genomic data are integrated. While PSP is rare, it is the most common primary tauopathy, and  
493 studying this disease is critical to understanding common pathological mechanisms across



494 tauopathies. Further work to include individuals with diverse ancestry background will also improve  
495 our understanding of genetic architecture of the disease.

## 496 **Conclusion**

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

## 503 **Declarations**

504 *Ethics approval and consent to participate*

505

506 *Consent for publication*

507 Not applicable.

508

509 *Availability of data and materials*

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

511 <https://github.com/whtop/PSP-Whole-Genome-Sequencing-Analysis>

512

513 *Competing interests*

514 Laura Molina-Porcel received income from Biogen as a consultant in 2022. Gesine Respondek is  
515 now employed by Roche (Hoffmann-La Roche, Basel, Switzerland) since 2021. Her affiliation whilst  
516 completing her contribution to this manuscript was German Center for Neurodegenerative Diseases

517 (DZNE), Munich, Germany. Thomas G Beach is a consultant for Aprinoia Therapeutics and a  
518 Scientific Advisor and stock option holder for Vivid Genomics. Huw Morris is employed by UCL. In  
519 the last 12 months he reports paid consultancy from Roche, Aprinoia, AI Therapeutics and Amylyx;  
520 lecture fees/honoraria - BMJ, Kyowa Kirin, Movement Disorders Society. Huw Morris is a  
521 co-applicant on a patent application related to C9ORF72 - Method for diagnosing a  
522 neurodegenerative disease (PCT/GB2012/052140). Giovanni Coppola is currently an employee of  
523 Regeneron Pharmaceuticals. Alison Goate serves on the SAB for Genentech and Muna Therapeutics.

524

525 Funding

526 This work was supported by NIH 5UG3NS104095, the Rainwater Charitable Foundation, and  
527 CurePSP. HW and PLC are supported by RF1-AG074328, P30-AG072979, U54-AG052427 and  
528 U24-AG041689. TSC is supported by NIH K08AG065519 and the Larry L Hillblom Foundation  
529 2021-A-005-SUP. KF was supported by CurePSP 685-2023-06-Pathway and K01 AG070326. MG is  
530 supported by P30 AG066511. BFG and KLN are supported by P30 AG072976 and R01 AG080001.  
531 TGB and GES are supported by P30AG072980. IR is supported by 2R01AG038791-06A,  
532 U01NS100610, R25NS098999, U19 AG063911-1 and 1R21NS114764-01A1. OR is support by U54  
533 NS100693. DG is supported by P30AG062429. ALB is supported by U19AG063911,  
534 R01AG073482, R01AG038791, and R01AG071756. BLM is supported by P01 AG019724, R01  
535 AG057234 and P0544014. VMV is supported by P01-AG-066597, P01-AG-017586. HRM is  
536 supported by CurePSP, PSPA, MRC, and Michael J Fox Foundation. RDS is supported by CurePSP,  
537 PSPA, and Reta Lila Weston Trust. JFC is supported by R01 AG054008, R01 NS095252, R01

538 AG060961, R01 NS086736, R01 AG062348, P30 AG066514, the Rainwater Charitable Foundation /  
539 Tau Consortium, Karen Strauss Cook Research, and Scholar Award, Stuart Katz & Dr. Jane Martin.  
540 AMG is supported by the Tau Consortium and U54-NS123746. YYL is supported by  
541 U54-AG052427; U24-AG041689. LSW is supported by U01AG032984, U54AG052427, and  
542 U24AG041689. GUH was funded by the Deutsche Forschungsgemeinschaft (DFG, German  
543 Research Foundation) under Germany's Excellence Strategy within the framework of the Munich  
544 Cluster for Systems Neurology (EXC 2145 SyNergy – ID 390857198); Deutsche  
545 Forschungsgemeinschaft (DFG, HO2402/18-1 MSAomics); German Federal Ministry of Education  
546 and Research (BMBF, 01KU1403A EpiPD; 01EK1605A HitTau; 01DH18025 TauTherapy). DHG is  
547 supported by 3UH3NS104095, Tau Consortium. WPL is supported by RF1-AG074328;  
548 P30-AG072979; U54-AG052427; U24-AG041689. Cases from Banner Sun Health Research  
549 Institute were supported by the NIH (U24 NS072026, P30 AG19610 and P30AG072980), the  
550 Arizona Department of Health Services (contract 211002, Arizona Alzheimer's Research Center), the  
551 Arizona Biomedical Research Commission (contracts 4001, 0011, 05-901 and 1001 to the Arizona  
552 Parkinson's Disease Consortium) and the Michael J. Fox Foundation for Parkinson's Research. The  
553 Mayo Clinic Brain Bank is supported through funding by NIA grants P50 AG016574, CurePSP  
554 Foundation, and support from Mayo Foundation.

555

#### 556 Acknowledgements

557 This project is supported by CurePSP, courtesy of a donation from the Morton and Marcine Friedman  
558 Foundation. We are indebted to the Biobanc-Hospital Clinic-FRCB-IDIBAPS and Center for

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

580 valuable contribution to this work.

581 AMP-AD (sa000011) data: Mayo RNAseq Study- Study data were provided by the following  
582 sources: The Mayo Clinic Alzheimer's Disease Genetic Studies, led by Dr. Nilufer Ertekin-Taner and  
583 Dr. Steven G. Younkin, Mayo Clinic, Jacksonville, FL using samples from the Mayo Clinic Study of  
584 Aging, the Mayo Clinic Alzheimer's Disease Research Center, and the Mayo Clinic Brain Bank. Data  
585 collection was supported through funding by NIA grants P50 AG016574, R01 AG032990, U01  
586 AG046139, R01 AG018023, U01 AG006576, U01 AG006786, R01 AG025711, R01 AG017216,  
587 R01 AG003949, NINDS grant R01 NS080820, CurePSP Foundation, and support from Mayo  
588 Foundation. Study data includes samples collected through the Sun Health Research Institute Brain  
589 and Body Donation Program of Sun City, Arizona. The Brain and Body Donation Program is  
590 supported by the National Institute of Neurological Disorders and Stroke (U24 NS072026 National  
591 Brain and Tissue Resource for Parkinson's Disease and Related Disorders), the National Institute on  
592 Aging (P30 AG19610 Arizona Alzheimer's Disease Core Center), the Arizona Department of Health  
593 Services (contract 211002, Arizona Alzheimer's Research Center), the Arizona Biomedical Research  
594 Commission (contracts 4001, 0011, 05-901 and 1001 to the Arizona Parkinson's Disease Consortium)  
595 and the Michael J. Fox Foundation for Parkinson's Research.

596 PSP-NIH-CurePSP-Tau (sa000015) data: This project was funded by the NIH grant UG3NS104095  
597 and supported by grants U54NS100693 and U54AG052427. Queen Square Brain Bank is supported  
598 by the Reta Lila Weston Institute for Neurological Studies and the Medical Research Council UK.  
599 The Mayo Clinic Florida had support from a Morris K. Udall Parkinson's Disease Research Center of  
600 Excellence (NINDS P50 #NS072187), CurePSP and the Tau Consortium. The samples from the

601 University of Pennsylvania are supported by NIA grant P01AG017586.

602 PSP-CurePSP-Tau (sa000016) data: This project was funded by the Tau Consortium, Rainwater  
603 Charitable Foundation, and CurePSP. It was also supported by NINDS grant U54NS100693 and NIA  
604 grants U54NS100693 and U54AG052427. Queen Square Brain Bank is supported by the Reta Lila  
605 Weston Institute for Neurological Studies and the Medical Research Council UK. The Mayo Clinic  
606 Florida had support from a Morris K. Udall Parkinson's Disease Research Center of Excellence  
607 (NINDS P50 #NS072187), CurePSP and the Tau Consortium. The samples from the University of  
608 Pennsylvania are supported by NIA grant P01AG017586. Tissues were received from the Victorian  
609 Brain Bank, supported by The Florey Institute of Neuroscience and Mental Health, The Alfred and  
610 the Victorian Forensic Institute of Medicine and funded in part by Parkinson's Victoria and MND  
611 Victoria. We are grateful to the Sun Health Research Institute Brain and Body Donation Program of  
612 Sun City, Arizona for the provision of human biological materials (or specific description, e.g. brain  
613 tissue, cerebrospinal fluid). The Brain and Body Donation Program is supported by the National  
614 Institute of Neurological Disorders and Stroke (U24 NS072026 National Brain and Tissue Resource  
615 for Parkinson's Disease and Related Disorders), the National Institute on Aging (P30 AG19610  
616 Arizona Alzheimer's Disease Core Center), the Arizona Department of Health Services (contract  
617 211002, Arizona Alzheimer's Research Center), the Arizona Biomedical Research Commission  
618 (contracts 4001, 0011, 05-901 and 1001 to the Arizona Parkinson's Disease Consortium) and the  
619 Michael J. Fox Foundation for Parkinson's Research. Biomaterial was provided by the Study Group  
620 DESCRIBE of the Clinical Research of the German Center for Neurodegenerative Diseases (DZNE).

621 PSP UCLA (sa000017) data: Thank to the AL-108-231 investigators, Adam L Boxer, Anthony E

622 Lang, Murray Grossman, David S Knopman, Bruce L Miller, Lon S Schneider, Rachelle S Doody,  
623 Andrew Lees, Lawrence I Golbe, David R Williams, Jean-Cristophe Corvol, Albert Ludolph,  
624 David Burn, Stefan Lorenzl, Irene Litvan, Erik D Roberson, Günter U Höglinger, Mary Koestler,  
625 Clifford R Jack Jr, Viviana Van Deerlin, Christopher Randolph, Iryna V Lobach, Hilary W Heuer,  
626 Illana Gozes, Lesley Parker, Steve Whitaker, Joe Hirman, Alistair J Stewart, Michael Gold, and  
627 Bruce H Morimoto.

628

629 *Authors' contribution*

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

## 639 **References**

- 640 1. Hauw JJ, Daniel SE, Dickson D, Horoupian DS, Jellinger K, Lantos PL, et al. Preliminary  
641 NINDS neuropathologic criteria for Steele-Richardson-Olszewski syndrome (progressive  
642 supranuclear palsy). *Neurology*. 1994;44(11):2015–2015.
- 643 2. Stamelou M, Respondek G, Giagkou N, Whitwell JL, Kovacs GG, Höglinger GU. Evolving  
644 concepts in progressive supranuclear palsy and other 4-repeat tauopathies. *Nat Rev Neurol*.  
645 2021 Oct;17(10):601–20.
- 646 3. Höglinger GU, Respondek G, Stamelou M, Kurz C, Josephs KA, Lang AE, et al. Clinical  
647 Diagnosis of Progressive Supranuclear Palsy: The Movement Disorder Society Criteria. *Mov*  
648 *Disord Off J Mov Disord Soc*. 2017 Jun;32(6):853–64.
- 649 4. Lukic MJ, Respondek G, Kurz C, Compta Y, Gelpi E, Ferguson LW, et al. Long-Duration  
650 Progressive Supranuclear Palsy: Clinical Course and Pathological Underpinnings. *Ann Neurol*.  
651 2022;92(4):637–49.
- 652 5. Ali F, Martin PR, Botha H, Ahlskog JE, Bower JH, Masumoto JY, et al. Sensitivity and  
653 specificity of diagnostic criteria for progressive supranuclear palsy. *Mov Disord*.  
654 2019;34(8):1144–53.
- 655 6. Kovacs GG, Lukic MJ, Irwin DJ, Arzberger T, Respondek G, Lee EB, et al. Distribution  
656 patterns of tau pathology in progressive supranuclear palsy. *Acta Neuropathol (Berl)*. 2020  
657 Aug;140(2):99–119.
- 658 7. Wen Y, Zhou Y, Jiao B, Shen L. Genetics of progressive supranuclear palsy: a review. *J Park*  
659 *Dis*. 2021;11(1):93–105.
- 660 8. Höglinger GU, Melhem NM, Dickson DW, Sleiman PM, Wang LS, Klei L, et al. Identification  
661 of common variants influencing risk of the tauopathy progressive supranuclear palsy. *Nat Genet*.  
662 2011;43(7):699–705.
- 663 9. Borroni B, Agosti C, Magnani E, Di Luca M, Padovani A. Genetic bases of Progressive  
664 Supranuclear Palsy: the MAPT tau disease. *Curr Med Chem*. 2011;18(17):2655–60.
- 665 10. Rademakers R, Cruts M, Van Broeckhoven C. The role of tau (MAPT) in frontotemporal  
666 dementia and related tauopathies. *Hum Mutat*. 2004;24(4):277–95.
- 667 11. Cooper YA, Teyssier N, Dräger NM, Guo Q, Davis JE, Sattler SM, et al. Functional regulatory  
668 variants implicate distinct transcriptional networks in dementia. *Science*. 2022 Aug  
669 19;377(6608):eabi8654.
- 670 12. Sanchez-Contreras MY, Kouri N, Cook CN, Heckman MG, Finch NA, Caselli RJ, et al.  
671 Replication of progressive supranuclear palsy genome-wide association study identifies



- 672 SLCO1A2 and DUSP10 as new susceptibility loci. *Mol Neurodegener.* 2018;13(1):1–10.
- 673 13. Chen JA, Chen Z, Won H, Huang AY, Lowe JK, Wojta K, et al. Joint genome-wide association  
674 study of progressive supranuclear palsy identifies novel susceptibility loci and genetic  
675 correlation to neurodegenerative diseases. *Mol Neurodegener.* 2018;13(1):1–11.
- 676 14. Jabbari E, Koga S, Valentino RR, Reynolds RH, Ferrari R, Tan MM, et al. Genetic determinants  
677 of survival in progressive supranuclear palsy: a genome-wide association study. *Lancet Neurol.*  
678 2021;20(2):107–16.
- 679 15. Jabbari E, Woodside J, Tan MM, Shoai M, Pittman A, Ferrari R, et al. Variation at the TRIM11  
680 locus modifies progressive supranuclear palsy phenotype. *Ann Neurol.* 2018;84(4):485–96.
- 681 16. Beecham GW, Bis JC, Martin ER, Choi SH, DeStefano AL, Van Duijn CM, et al. The  
682 Alzheimer’s Disease Sequencing Project: study design and sample selection. *Neurol Genet.*  
683 2017;3(5).
- 684 17. Kuzma A, Valladares O, Cweibel R, Greenfest-Allen E, Childress DM, Malamon J, et al.  
685 NIAGADS: The NIA Genetics of Alzheimer’s Disease Data Storage Site. *Alzheimers Dement.*  
686 2016 Nov;12(11):1200–3.
- 687 18. Consortium 1000 Genomes Project. A global reference for human genetic variation. Vol. 526,  
688 *Nature.* Nature Publishing Group; 2015. p. 68.
- 689 19. Lowy-Gallego E, Fairley S, Zheng-Bradley X, Ruffier M, Clarke L, Flicek P. Variant calling on  
690 the GRCh38 assembly with the data from phase three of the 1000 Genomes Project. *Wellcome*  
691 *Open Res.* 2019 Dec 30;4:50.
- 692 20. Genome Reference Consortium. GRCh38 reference 000001405.15 [Internet]. [cited 2022 Jun  
693 22]. Available from:  
694 [https://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000/001/405/GCA\\_000001405.15\\_GRCh38/seqs](https://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000/001/405/GCA_000001405.15_GRCh38/seqs_for_alignment_pipelines.ucsc_ids/GCA_000001405.15_GRCh38_no_alt_analysis_set.fna.gz)  
695 [\\_for\\_alignment\\_pipelines.ucsc\\_ids/GCA\\_000001405.15\\_GRCh38\\_no\\_alt\\_analysis\\_set.fna.gz](https://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000/001/405/GCA_000001405.15_GRCh38/seqs_for_alignment_pipelines.ucsc_ids/GCA_000001405.15_GRCh38_no_alt_analysis_set.fna.gz)
- 696 21. Schneider VA, Graves-Lindsay T, Howe K, Bouk N, Chen HC, Kitts PA, et al. Evaluation of  
697 GRCh38 and de novo haploid genome assemblies demonstrates the enduring quality of the  
698 reference assembly. *Genome Res.* 2017 May;27(5):849–64.
- 699 22. Yang J, Bakshi A, Zhu Z, Hemani G, Vinkhuyzen AA, Lee SH, et al. Genetic variance  
700 estimation with imputed variants finds negligible missing heritability for human height and  
701 body mass index. *Nat Genet.* 2015;47(10):1114–20.
- 702 23. Gogarten SM, Sofer T, Chen H, Yu C, Brody JA, Thornton TA, et al. Genetic association testing  
703 using the GENESIS R/Bioconductor package. *Bioinformatics.* 2019;35(24):5346–8.
- 704 24. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM. Robust relationship  
705 inference in genome-wide association studies. *Bioinformatics.* 2010;26(22):2867–73.

- 706 25. Conomos MP, Miller MB, Thornton TA. Robust inference of population structure for ancestry  
707 prediction and correction of stratification in the presence of relatedness. *Genet Epidemiol.*  
708 2015;39(4):276–93.
- 709 26. Zou Y, Carbonetto P, Wang G, Stephens M. Fine-mapping from summary data with the “Sum of  
710 Single Effects” model. *PLoS Genet.* 2022;18(7):e1010299.
- 711 27. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from  
712 high-throughput sequencing data. *Nucleic Acids Res.* 2010 Sep;38(16):e164.
- 713 28. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, et al. The Ensembl Variant  
714 Effect Predictor. *Genome Biol.* 2016 Jun 6;17(1):122.
- 715 29. Cunningham F, Allen JE, Allen J, Alvarez-Jarreta J, Amode MR, Armean IM, et al. Ensembl  
716 2022. *Nucleic Acids Res.* 2022 Jan 7;50(D1):D988–95.
- 717 30. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. The mutational  
718 constraint spectrum quantified from variation in 141,456 humans. *Nature.* 2020  
719 May;581(7809):434–43.
- 720 31. Swarup V, Chang TS, Duong DM, Dammer EB, Dai J, Lah JJ, et al. Identification of Conserved  
721 Proteomic Networks in Neurodegenerative Dementia. *Cell Rep.* 2020 Jun 23;31(12):107807.
- 722 32. Sjöstedt E, Zhong W, Fagerberg L, Karlsson M, Mitsios N, Adori C, et al. An atlas of the  
723 protein-coding genes in the human, pig, and mouse brain. *Science.* 2020 Mar  
724 6;367(6482):eaay5947.
- 725 33. Melé M, Ferreira PG, Reverter F, DeLuca DS, Monlong J, Sammeth M, et al. Human genomics.  
726 The human transcriptome across tissues and individuals. *Science.* 2015 May  
727 8;348(6235):660–5.
- 728 34. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis.  
729 *BMC Bioinformatics.* 2008 Dec 29;9:559.
- 730 35. Chen X, Schulz-Trieglaff O, Shaw R, Barnes B, Schlesinger F, Källberg M, et al. Manta: rapid  
731 detection of structural variants and indels for germline and cancer sequencing applications.  
732 *Bioinformatics.* 2016;32(8):1220–2.
- 733 36. Layer RM, Chiang C, Quinlan AR, Hall IM. LUMPY: a probabilistic framework for structural  
734 variant discovery. *Genome Biol.* 2014;15(6):1–19.
- 735 37. Eggertsson HP, Kristmundsdottir S, Beyter D, Jonsson H, Skuladottir A, Hardarson MT, et al.  
736 GraphTyper2 enables population-scale genotyping of structural variation using pangenome  
737 graphs. *Nat Commun.* 2019;10(1):1–8.
- 738 38. Wang H, Dombroski BA, Cheng PL, Tucci A, Si Y qin, Farrell JJ, et al. Structural Variation

- 739 Detection and Association Analysis of Whole-Genome-Sequence Data from 16,905  
740 Alzheimer's Diseases Sequencing Project Subjects. medRxiv. 2023;
- 741 39. Belyeu JR, Chowdhury M, Brown J, Pedersen BS, Cormier MJ, Quinlan AR, et al. Samplot: a  
742 platform for structural variant visual validation and automated filtering. *Genome Biol.*  
743 2021;22(1):1–13.
- 744 40. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV):  
745 high-performance genomics data visualization and exploration. *Brief Bioinform.*  
746 2013;14(2):178–92.
- 747 41. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set  
748 for whole-genome association and population-based linkage analyses. *Am J Hum Genet.*  
749 2007;81(3):559–75.
- 750 42. Lee S, Emond MJ, Bamshad MJ, Barnes KC, Rieder MJ, Nickerson DA, et al. Optimal Unified  
751 Approach for Rare-Variant Association Testing with Application to Small-Sample Case-Control  
752 Whole-Exome Sequencing Studies. *Am J Hum Genet.* 2012 Aug 10;91(2):224–37.
- 753 43. Wang X, Campbell MR, Lacher SE, Cho HY, Wan M, Crowl CL, et al. A polymorphic  
754 antioxidant response element links NRF2/sMAF binding to enhanced MAPT expression and  
755 reduced risk of Parkinsonian disorders. *Cell Rep.* 2016;15(4):830–42.
- 756 44. Anaya F, Lees A, Silva R. Tau gene promoter rs242557 and allele-specific protein binding.  
757 *Transl Neurosci* [Internet]. 2011 Jan 1 [cited 2023 Nov 9];2(2). Available from:  
758 <https://www.degruyter.com/document/doi/10.2478/s13380-011-0021-6/html>
- 759 45. Sawa A, Amano N, Yamada N, Kajio H, Yagishita S, Takahashi T, et al. Apolipoprotein E in  
760 progressive supranuclear palsy in Japan. *Mol Psychiatry.* 1997;2(4):341–2.
- 761 46. Zhao N, Liu CC, Van Ingelgom AJ, Linares C, Kurti A, Knight JA, et al. APOE  $\epsilon$ 2 is associated  
762 with increased tau pathology in primary tauopathy. *Nat Commun.* 2018;9(1):4388.
- 763 47. Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, et al. The genotype-tissue  
764 expression (GTEx) project. *Nat Genet.* 2013;45(6):580.
- 765 48. Xie R, Nguyen S, McKeenan K, Wang F, McKeenan WL, Liu L. Microtubule-associated  
766 protein 1S (MAP1S) bridges autophagic components with microtubules and mitochondria to  
767 affect autophagosomal biogenesis and degradation. *J Biol Chem.* 2011;286(12):10367–77.
- 768 49. Shi L, Huang C, Luo Q, Xia Y, Liu H, Li L, et al. Pilot study: molecular risk factors for  
769 diagnosing sporadic Parkinson's disease based on gene expression in blood in MPTP-induced  
770 rhesus monkeys. *Oncotarget.* 2017;8(62):105606.
- 771 50. Pan M, Li X, Xu G, Tian X, Li Y, Fang W. Tripartite Motif Protein Family in Central Nervous  
772 System Diseases. *Cell Mol Neurobiol.* 2023;1–23.

- 773 51. Kutkowska-Kaźmierczak A, Rydzanicz M, Chlebowski A, Kłosowska-Kosicka K, Mika A,  
774 Gruchota J, et al. Dominant ELOVL1 mutation causes neurological disorder with ichthyotic  
775 keratoderma, spasticity, hypomyelination and dysmorphic features. *J Med Genet.*  
776 2018;55(6):408–14.
- 777 52. Farrell K, Humphrey J, Chang T, Zhao Y, Leung YY, Kuksa PP, et al. Genetic, transcriptomic,  
778 histological, and biochemical analysis of progressive supranuclear palsy implicates glial  
779 activation and novel risk genes. *bioRxiv.* 2023;2023–11.
- 780 53. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Analysis of  
781 protein-coding genetic variation in 60,706 humans. *Nature.* 2016;536(7616):285–91.
- 782 54. Chen S, Francioli LC, Goodrich JK, Collins RL, Kanai M, Wang Q, et al. A genome-wide  
783 mutational constraint map quantified from variation in 76,156 human genomes. *bioRxiv.*  
784 2022;2022–03.
- 785 55. Lee WP, Choi SH, Shea MG, Cheng PL, Dombroski BA, Pitsillides AN, et al. Association of  
786 Common and Rare Variants with Alzheimer’s Disease in over 13,000 Diverse Individuals with  
787 Whole-Genome Sequencing from the Alzheimer’s Disease Sequencing Project. *medRxiv.*  
788 2023;2023–09.
- 789 56. Swarup V, Hinz FI, Rexach JE, Noguchi K ichi, Toyoshiba H, Oda A, et al. Identification of  
790 evolutionarily conserved gene networks mediating neurodegenerative dementia. *Nat Med.*  
791 2019;25(1):152–64.
- 792 57. Swarup V, Chang TS, Duong DM, Dammer EB, Dai J, Lah JJ, et al. Identification of conserved  
793 proteomic networks in neurodegenerative dementia. *Cell Rep [Internet].* 2020 [cited 2023 Nov  
794 9];31(12). Available from: [https://www.cell.com/cell-reports/pdf/S2211-1247\(20\)30788-9.pdf](https://www.cell.com/cell-reports/pdf/S2211-1247(20)30788-9.pdf)
- 795 58. Parikshak NN, Gandal MJ, Geschwind DH. Systems biology and gene networks in  
796 neurodevelopmental and neurodegenerative disorders. *Nat Rev Genet.* 2015;16(8):441–58.
- 797 59. Baker M, Litvan I, Houlden H, Adamson J, Dickson D, Perez-Tur J, et al. Association of an  
798 extended haplotype in the tau gene with progressive supranuclear palsy. *Hum Mol Genet.*  
799 1999;8(4):711–5.
- 800 60. Wang H, Wang LS, Schellenberg G, Lee WP. The role of structural variations in Alzheimer’s  
801 disease and other neurodegenerative diseases. *Front Aging Neurosci.* 2023;
- 802 61. Mizobuchi M, Murao K, Takeda R, Kakimoto Y. Tissue-specific expression of isoaspartyl  
803 protein carboxyl methyltransferase gene in rat brain and testis. *J Neurochem.* 1994;62(1):322–8.
- 804 62. Wu X, Jia G, Yang H, Sun C, Liu Y, Diao Z. Neural stem cell-conditioned medium upregulated  
805 the PCMT1 expression and inhibited the phosphorylation of MST1 in SH-SY5Y cells induced  
806 by A $\beta$  25-35. *Biocell.* 2022;46(2):471.

- 807 63. Shi L, Al-Baadani A, Zhou K, Shao A, Xu S, Chen S, et al. PCMT1 ameliorates neuronal  
808 apoptosis by inhibiting the activation of MST1 after subarachnoid hemorrhage in rats. *Transl*  
809 *Stroke Res.* 2017;8:474–83.
- 810 64. Smit, AFA, Hubley, R & Green, P. RepeatMasker Open-4.0. 2013-2015  
811 <<http://www.repeatmasker.org>>.
- 812 65. Chen JA, Chen Z, Won H, Huang AY, Lowe JK, Wojta K, et al. Joint genome-wide association  
813 study of progressive supranuclear palsy identifies novel susceptibility loci and genetic  
814 correlation to neurodegenerative diseases. *Mol Neurodegener.* 2018 Aug 8;13(1):41.
- 815 66. Rizzu P, Van Swieten JC, Joosse M, Hasegawa M, Stevens M, Tibben A, et al. High prevalence  
816 of mutations in the microtubule-associated protein tau in a population study of frontotemporal  
817 dementia in the Netherlands. *Am J Hum Genet.* 1999;64(2):414–21.
- 818 67. Rovelet-Lecrux A, Lecourtois M, Thomas-Anterion C, Le Ber I, Brice A, Frebourg T, et al.  
819 Partial deletion of the MAPT gene: A novel mechanism of FTDP-17. *Hum Mutat.*  
820 2009;30(4):E591–602.
- 821 68. Farrer LA, Cupples LA, Haines JL, Hyman B, Kukull WA, Mayeux R, et al. Effects of age, sex,  
822 and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease: a  
823 meta-analysis. *Jama.* 1997;278(16):1349–56.
- 824 69. Selkoe DJ, Podlisny MB. Deciphering the genetic basis of Alzheimer’s disease. *Annu Rev*  
825 *Genomics Hum Genet.* 2002;3(1):67–99.
- 826 70. Bertram L, McQueen MB, Mullin K, Blacker D, Tanzi RE. Systematic meta-analyses of  
827 Alzheimer disease genetic association studies: the AlzGene database. *Nat Genet.*  
828 2007;39(1):17–23.
- 829 71. Shi Y, Yamada K, Liddel SA, Smith ST, Zhao L, Luo W, et al. ApoE4 markedly exacerbates  
830 tau-mediated neurodegeneration in a mouse model of tauopathy. *Nature.*  
831 2017;549(7673):523–7.
- 832 72. Rasmussen KL, Tybjærg-Hansen A, Nordestgaard BG, Frikke-Schmidt R. Associations of  
833 Alzheimer disease–protective APOE variants with age-related macular degeneration. *JAMA*  
834 *Ophthalmol.* 2023;141(1):13–21.
- 835 73. Klaver CC, Kliffen M, van Duijn CM, Hofman A, Cruts M, Grobbee DE, et al. Genetic  
836 association of apolipoprotein E with age-related macular degeneration. *Am J Hum Genet.*  
837 1998;63(1):200–6.
- 838 74. Lee S, Abecasis GR, Boehnke M, Lin X. Rare-variant association analysis: study designs and  
839 statistical tests. *Am J Hum Genet.* 2014;95(1):5–23.
- 840 75. Cassandri M, Smirnov A, Novelli F, Pitolli C, Agostini M, Malewicz M, et al. Zinc-finger

- 841 proteins in health and disease. *Cell Death Discov.* 2017 Nov 13;3(1):1–12.
- 842 76. Fedotova AA, Bonchuk AN, Mogila VA, Georgiev PG. C2H2 Zinc Finger Proteins: The Largest  
843 but Poorly Explored Family of Higher Eukaryotic Transcription Factors. *Acta Naturae.*  
844 2017;9(2):47–58.
- 845 77. Bu S, Lv Y, Liu Y, Qiao S, Wang H. Zinc Finger Proteins in Neuro-Related Diseases  
846 Progression. *Front Neurosci.* 2021 Nov 18;15:760567.
- 847 78. Al-Naama N, Mackeh R, Kino T. C2H2-Type Zinc Finger Proteins in Brain Development,  
848 Neurodevelopmental, and Other Neuropsychiatric Disorders: Systematic Literature-Based  
849 Analysis. *Front Neurol.* 2020;11:32.
- 850 79. Shin JH, Ko HS, Kang H, Lee Y, Lee YI, Pletinkova O, et al. PARIS (ZNF746) Repression of  
851 PGC-1 $\alpha$  Contributes to Neurodegeneration in Parkinson's Disease. *Cell.* 2011 Mar  
852 4;144(5):689–702.
- 853 80. Li R, Strohmeyer R, Liang Z, Lue LF, Rogers J. CCAAT/enhancer binding protein delta  
854 (C/EBPdelta) expression and elevation in Alzheimer's disease. *Neurobiol Aging.* 2004  
855 Sep;25(8):991–9.
- 856 81. Ko CY, Chang LH, Lee YC, Sterneck E, Cheng CP, Chen SH, et al. CCAAT/enhancer binding  
857 protein delta (CEBPD) elevating PTX3 expression inhibits macrophage-mediated phagocytosis  
858 of dying neuron cells. *Neurobiol Aging.* 2012 Feb;33(2):422.e11-25.
- 859 82. Nicolas E, Poitelon Y, Chouery E, Salem N, Levy N, Mégarbané A, et al. CAMOS, a  
860 nonprogressive, autosomal recessive, congenital cerebellar ataxia, is caused by a mutant  
861 zinc-finger protein, ZNF592. *Eur J Hum Genet EJHG.* 2010 Oct;18(10):1107–13.
- 862 83. Vodopiutz J, Seidl R, Prayer D, Khan MI, Mayr JA, Streubel B, et al. WDR73 Mutations Cause  
863 Infantile Neurodegeneration and Variable Glomerular Kidney Disease. *Hum Mutat.* 2015  
864 Nov;36(11):1021–8.
- 865 84. Parikshak NN, Luo R, Zhang A, Won H, Lowe JK, Chandran V, et al. Integrative functional  
866 genomic analyses implicate specific molecular pathways and circuits in autism. *Cell.*  
867 2013;155(5):1008–21.
- 868 85. Takahashi M, Weidenheim KM, Dickson DW, Ksiezak-Reding H. Morphological and  
869 biochemical correlations of abnormal tau filaments in progressive supranuclear palsy. *J*  
870 *Neuropathol Exp Neurol.* 2002 Jan;61(1):33–45.
- 871 86. Roemer SF, Grinberg LT, Crary JF, Seeley WW, McKee AC, Kovacs GG, et al. Rainwater  
872 Charitable Foundation criteria for the neuropathologic diagnosis of progressive supranuclear  
873 palsy. *Acta Neuropathol (Berl).* 2022 Oct;144(4):603–14.

874