

1 A Novel Susceptibility Locus in NFASC Highlights Oligodendrocytes and
2 Myelination in Progressive Supranuclear Palsy Pathology

3 **Pablo García-González**^{1,2,3}, Héctor Rodrigo Lara⁴, Yaroslau Compta^{5,6}, Manuel
4 Fernandez^{5,6}, Sven J. van der Lee^{7,8,9}, Itziar de Rojas^{1,2}, Laura Saiz¹⁰, Celia Painous^{5,6},
5 Ana Camara^{5,6}, Esteban Muñoz^{5,6}, Maria J. Martí^{5,6}, Francesc Valldeoriola^{5,6}, Raquel
6 Puerta^{1,3}, Ignacio Illán-Gala^{11,2}, Javier Pagonabarraga^{12,2}, Oriol Dols-Icardo^{11,2}, Jaime
7 Kulisevsky^{12,2}, Juan Fortea^{11,2}, Alberto Lleó^{11,2}, Claudia Olivé¹, Sterre C.M. de
8 Boer^{13,9,14}, Marc Hulsman^{7,8,9}, Yolande A.L. Pijnenburg^{8,9}, Rafael Díaz Belloso^{15,16},
9 Laura Muñoz-Delgado^{15,16}, Dolores Buiza Rueda^{15,16}, Pilar Gómez-Garre^{15,16}, Iban
10 Aldecoa^{17,18}, Gemma Aragonés¹⁷, Jorge Hernandez Vara^{19,16}, Maite Mendioroz^{20,21,22},
11 Jordi Pérez-Tur^{23,2}, Pieter Jelle Visser^{24,25,26,27}, Anouk den Braber^{24,25,28}, Janne M.
12 Papma^{29,30}, Ángel Martín Montes^{2,31,32}, Eloy Rodríguez-Rodríguez^{33,2}, Josep Blázquez-
13 Folch¹, Andrea Miguel¹, Fernando García-Gutiérrez¹, Amanda Cano^{1,2}, Sergi Valero^{1,2},
14 Marta Marquíe^{1,2}, María Capdevila-Bayo^{1,34}, Maitee Rosende-Roca³⁵, Inés Quintela³⁶,
15 Ángel Carracedo^{37,38}, Lluís Tàrraga^{1,2}, Luis M Real^{39,40}, Jose Luis Royo⁴¹, Maria Elena
16 Erro^{42,43}, Carmen Guerrero⁴⁴, Daniela Corte Torres⁴⁵, Marta Blázquez-Estrada^{46,47},
17 Beatriz San Millán^{48,49}, Susana Teijeira^{48,50}, Dolores Vilas Rolan⁵¹, Isabel Hernández^{1,2},
18 Antonio Sánchez-Soblechero⁵², Beatriz de la Casa-Fages⁵³, Soledad Serrano López^{31,32},
19 Raquel Baviera-Muñoz⁵⁴, Amaya Lavín⁵⁵, Ricardo Taipa^{56,57}, Guillermo Amer⁵⁸, Elena
20 Martínez-Saez^{59,60}, Marta Fernández-Matarrubia^{61,62}, Carmen Lage-Martínez^{63,2,64},
21 Victoria Álvarez^{65,47}, Laura Molina-Porcel^{66,17}, Henne Holstege^{7,8,9}, Pablo Mir^{15,16,67},
22 Olivia Belbin^{68,2}, Mercè Boada^{1,2}, Victoria Fernández¹, **María J. Bullido**^{69,2,31,70},
23 **Alberto Rábano**^{71,2}, **Pascual Sánchez-Juan**^{71,2}, **Agustín Ruiz**^{1,2,72*}

24 * Corresponding author: Agustín Ruiz MD, PhD.

25 *María J. Bullido, Alberto Rábano, Pascual Sánchez-Juan and Agustín Ruiz contributed equally to this work.*

26 ¹Ace Alzheimer Center Barcelona-Universitat Internacional de Catalunya, Barcelona, Spain,

27 ²CIBERNED, Network Center for Biomedical Research in Neurodegenerative Diseases, National

28 Institute of Health Carlos III, Madrid, Spain, ³Universitat de Barcelona (UB), Barcelona, Spain, ⁴Banco

29 de Cerebros de la Región de Murcia, Hospital Clínico Universitario Virgen de la Arrixaca, El Palmar,

30 Spain, ⁵Parkinson's Disease & Movement Disorders Unit, Neurology Service, Hospital Clínic I

31 Universitari de Barcelona; IDIBAPS, CIBERNED (CB06/05/0018-ISCIII), ERN- RND, Institut Clínic de

32 Neurociències (Maria de Maeztu Excellence Centre), Universitat de Barcelona. Barcelona, Catalonia,

33 Spain, ⁶Lab of Parkinson Disease and Other Neurodegenerative Movement Disorders: Clinical and

34 Experimental Research; Department of Neurology, Hospital Clínic de Barcelona, Institut

35 D'Investigacions Biomèdiques August Pi I Sunyer (IDIBAPS), Institut de Neurociències, Universitat de

36 Barcelona, Barcelona, Catalonia, Spain, ⁷Genomics of Neurodegenerative Diseases and Aging, Human

37 Genetics, Vrije Universiteit Amsterdam, Amsterdam UMC location VUmc, Amsterdam, The

38 Netherlands., ⁸Alzheimer Center Amsterdam, Neurology, Vrije Universiteit Amsterdam, Amsterdam

39 UMC location VUmc, Amsterdam, The Netherlands ., ⁹Amsterdam Neuroscience, Neurodegeneration,

40 Amsterdam, The Netherlands, ¹⁰Alzheimer's Disease Research Unit, CIEN Foundation, Queen Sofia

41 Foundation Alzheimer Center, Madrid, Spain, ¹¹Sant Pau Memory Unit, Department of Neurology,

42 Institut de Recerca de Sant Pau; Hospital de Sant Pau, Barcelona, Spain, ¹²Movement Disorders Unit,

43 Department of Neurology, Institut de Recerca de Sant Pau; Hospital de Sant Pau, Barcelona, Spain,

44 ¹³Alzheimer Center Amsterdam, Neurology, Vrije Universiteit Amsterdam, Amsterdam UMC location

45 VUmc, Amsterdam, The Netherlands., ¹⁴School of Psychology and Brain & Mind Centre, The University

46 of Sydney, Sydney, Australia, ¹⁵Unidad de Trastornos del Movimiento, Servicio de Neurología y

47 Neurofisiología Clínica, Instituto de Biomedicina de Sevilla, IBiS / Hospital Universitario Virgen del

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

49 Enfermedades Neurodegenerativas (CIBERNED), Instituto de Salud Carlos III, Madrid, Spain.,
50 ¹⁷Neurological Tissue Bank, Biobanc-Hospital Clínic-FRCB-IDIBAPS, Barcelona, Spain., ¹⁸Pathology
51 Department, Hospital Clinic de Barcelona-University of Barcelona, Barcelona, Spain, ¹⁹Servicio de
52 Neurología y Grupo de Enfermedades Neurodegenerativas. Vall d'Hebron Institut de Recerca (VHIR),
53 Barcelona, Spain, ²⁰Navarrabiomed, Pamplona, Spain, ²¹Department of Neurology, Hospital Universitario
54 de Navarra, Pamplona, Spain, ²²Instituto de Investigación Sanitaria de Navarra (IDISNA), Pamplona,
55 Spain, ²³Unitat de Genètica Molecular, Institut de Biomedicina de València-CSIC, Valencia, Spain,
56 ²⁴Alzheimer Center Amsterdam, Neurology, Vrije Universiteit Amsterdam, Amsterdam UMC location
57 VUmc, Amsterdam, The Netherlands, ²⁵Amsterdam Neuroscience, Neurodegeneration, 1081 HV
58 Amsterdam, The Netherlands, ²⁶Alzheimer Center Limburg, School for Mental Health and Neuroscience,
59 Maastricht University, 6200 MD Maastricht, The Netherlands, ²⁷Department of Neurobiology, Care
60 Sciences and Society, Division of Neurogeriatrics, Karolinska Institutet, 171 77 Stockholm, Sweden,
61 ²⁸Department of Psychiatry and Neurochemistry, Institute of Neuroscience and Physiology, The
62 Sahlgrenska Academy, University of Gothenburg, 405 30 Gothenburg, Sweden, ²⁹Department of
63 Neurology and Alzheimer Center Erasmus MC, Erasmus MC University Medical Center, Rotterdam, the
64 Netherlands, ³⁰Department of Internal Medicine Erasmus MC, Erasmus MC University Medical Center,
65 Rotterdam, the Netherlands, ³¹Hospital La Paz Institute for Health Research, IdiPAZ, Madrid, Spain,
66 ³²Department of Neurology, La Paz University Hospital, Madrid, Spain, ³³Neurology Service, Marqués de
67 Valdecilla University Hospital (University of Cantabria and IDIVAL), Santander, Spain, ³⁴Department of
68 Pharmacology, Toxicology and Therapeutic Chemistry, Faculty of Pharmacy and Food Science,
69 Universitat de Barcelona, Spain, ³⁵Ace Alzheimer Center Barcelona, Universitat Internacional de
70 Catalunya, Barcelona, Spain, ³⁶Grupo de Medicina Xenómica, Fundación Pública Galega de Medicina
71 Xenómica, Santiago de Compostela, Spain, ³⁷Grupo de Medicina Xenómica, CIBERER, CIMUS.
72 Universidade de Santiago de Compostela, Santiago de Compostela, Spain, ³⁸Fundación Pública Galega de
73 Medicina Xenómica- IDIS, Santiago de Compostela, Spain, ³⁹Instituto de Biomedicina de Sevilla,
74 IBiS/Hospital Universitario Virgen de Valme /CSIC/Universidad de Sevilla. Departamento de Bioquímica
75 Médica, Biología Molecular e Inmunología. Sevilla, Spain, ⁴⁰CIBERINFEC, Network Center for
76 Biomedical Research in Infectious Diseases, National Institute of Health of Carlos III, Madrid, Spain,
77 ⁴¹Departamento de Especialidades Quirúrgicas, Bioquímica e Inmunología, Facultad de Medicina,
78 Universidad de Málaga, Málaga, Spain, ⁴²Servicio de Neurología, Hospital Universitario de Navarra,
79 Pamplona, Spain, ⁴³Grupo de investigación Neuroepigenética, Navarrabiomed, Pamplona, Spain, ⁴⁴Banco
80 de Cerebros (Biobanco), Hospital Universitario Fundación Alcorcón, 28922 Alcorcón, Spain, ⁴⁵Biobank
81 of Principado de Asturias, Instituto de Investigación Sanitaria del Principado de Asturias (ISPA), Oviedo,
82 Spain, ⁴⁶Servicio de Neurología, Hospital Universitario Central de Asturias, Spain, ⁴⁷Instituto de
83 Investigación Sanitaria del Principado de Asturias (ISPA), ⁴⁸Rare Diseases and Pediatric Medicine
84 Research Group, Galicia Sur Health Research Institute (IIS Galicia Sur), SERGAS-UVIGO, Hospital
85 Álvaro Cunqueiro, 36312 Vigo, Spain, ⁴⁹Pathology Department, Galicia Sur Health Research Institute
86 (IIS Galicia Sur). SERGAS-UVIGO. Alvaro Cunqueiro Hospital, Vigo, Spain, ⁵⁰Biobank of Galicia Sur
87 Health Research Institute (Biobank IIS Galicia Sur), SERGAS-UVIGO, Hospital Álvaro Cunqueiro,
88 36312 Vigo, Spain, ⁵¹Movement Disorders Unit, Neurology Service, Hospital Germans Trias i Pujol,
89 Badalona, Barcelona, Spain, ⁵²Department of Neurology, Instituto de Investigación Sanitaria Gregorio
90 Marañón, Hospital General Universitario Gregorio Marañón, Madrid, Spain, ⁵³Movement Disorders Unit,
91 Neurology Department, Instituto de Investigación Sanitaria Gregorio Marañón, Hospital General
92 Universitario Gregorio Marañón, Madrid, Spain, ⁵⁴Molecular, Cellular and Genomics Biomedicine
93 Research Group. Instituto de Investigación Sanitaria La Fe, Valencia, Spain, ⁵⁵Navarrabiomed, Hospital
94 Universitario de Navarra (HUN), Universidad Pública de Navarra (UPNA), IdiSNA, Pamplona, Spain,
95 ⁵⁶Department of Neuropathology, Centro Hospitalar Universitário de Santo António, Porto, Portugal,
96 ⁵⁷UMIB - Unit for Multidisciplinary Research in Biomedicine, ICBAS - School of Medicine and
97 Biomedical Sciences, University of Porto, 4050-313, Porto, Portugal, ⁵⁸Cognitive Neurology Unit, Son
98 Espases University Hospital, Palma, Spain, ⁵⁹Pathology Department, Neuropathology. Vall d'Hebron
99 University Hospital, Barcelona, Spain, ⁶⁰Vall d'Hebron University Hospital Biobank, Vall d'Hebron
100 Hospital Research Institute, Barcelona, Spain, ⁶¹Department of Neurology, Clínica Universidad de
101 Navarra, Pamplona, Spain, ⁶²IdiSNA, Navarra Institute for Health Research, Pamplona, Spain,
102 ⁶³Neurology Service, Marqués de Valdecilla University Hospital-IDIVAL, Santander, Spain, ⁶⁴Atlantic
103 Fellow for Equity in Brain health, Global Brain Health Institute, University of California, San Francisco,
104 United States, ⁶⁵Laboratorio de Genética. Hospital Universitario Central de Asturias, Oviedo, Spain,

105 ⁶⁶Alzheimer's disease and other cognitive disorders unit. Neurology Service, Hospital Clínic, Fundació de
106 Recerca Clínic Barcelona-Institut d'Investigacions Biomediques August Pi i Sunyer (FRCB-IDIBAPS),
107 University of Barcelona, Barcelona, Spain, ⁶⁷Departamento de Medicina, Facultad de Medicina,
108 Universidad de Sevilla, Spain, ⁶⁸Department of Neurology, II B Sant Pau, Hospital de la Santa Creu i Sant
109 Pau, Universitat Autònoma de Barcelona, Barcelona, Spain, ⁶⁹Centro de Biología Molecular Severo
110 Ochoa (UAM-CSIC), ⁷⁰Universidad Autónoma de Madrid, ⁷¹Alzheimer's Centre Reina Sofia-CIEN
111 Foundation-ISCIII, 28031 Madrid, Spain, ⁷²Glenn Biggs Institute for Alzheimer's & Neurodegenerative
112 Diseases, University of Texas Health Science Center at San Antonio, San Antonio, TX, USA

113 **Abstract**

114 **We conducted the largest PSP GWAS of the Iberian population to date (522 cases from 22**
115 **Spanish and Portuguese institutions). We independently replicated seven known PSP risk**
116 **variants, and unveiled a novel locus in *NFASC/CNTN2* after meta-analysing our results with**
117 **a newly available Dutch cohort and publicly available summary statistics. These findings**
118 **highlight the importance of neuron-oligodendrocyte interactions in PSP etiopathology.**

119 **Main**

120 Progressive supranuclear palsy (PSP) is a proteinopathy neuropathologically defined by intra-
121 neuronal and intra-glial aggregations of 4-repeat microtubule associated *tau* and clinically
122 expressed as several phenotypes with different combinations of oculomotor, parkinsonian, bulbar,
123 cognitive, and behavioural features¹. The common allele of an ancestral 900 kb inversion, known
124 as the *MAPT* H1 haplotype, is the most important genetic risk factor for sporadic PSP, causing a
125 5.5-fold increase in PSP risk compared to the alternative *MAPT* H2 haplotype². Importantly, the
126 low prevalence of PSP conditions the scarcity of genome-wide association studies (GWAS), with
127 only eight genetic risk loci reported to date³⁻⁵. Elucidating the genetic component of PSP could
128 lead to early and improved diagnosis, identification of pathophysiological biomarkers and novel
129 therapeutic treatments.

130 The PSP/DEGESCO cohort consists of clinical (N=327) and histopathological (N=195) PSP
131 subjects from Spain and Portugal (Supplementary Table 1), genotyped using the Affymetrix 815K
132 Spanish Biobank Array (Thermo Fisher). We leveraged our previous cohort,
133 GR@ACE/DEGESCO⁶, to select our controls (N=1950). Since cases and controls were
134 genotyped by the same platform and sequencing centre, we performed joint calling, quality
135 control (QC) and TOPMed imputation⁷ of our genomic data, following standard procedures
136 (Supplementary Information, Extended data Fig. 1). Clinical cases fulfilling the Movement
137 Disorder Society (MDS) criteria for probable PSP⁸ were initially selected for our study, allowing
138 the inclusion of atypical PSP phenotypes. However, we observed that atypical cases had a
139 decreased *MAPT* H1 burden compared to those with histopathological confirmation ($\chi^2=6.29$;
140 $DF=1$; $p=0.01$), and that other known PSP genetic risk factors also showed weaker associations
141 with disease status in atypical cases compared to those with Richardson's syndrome (PSP-RS) or

142 histopathological confirmation (Supplementary Information), indicating the presence of
143 phenotypic heterogeneity in our non-Richardson clinical cases (nR-PSP). As increasing the
144 sample size in detriment of phenotypic accuracy can undermine the detection of small effect
145 variants in GWAS studies⁹, we decided to keep only PSP-RS and histopathologically confirmed
146 cases for further analysis.

147 We ran PC-adjusted case-control genome wide associations (~9M common genetic variants) for
148 our three subcohorts of Spanish and Portuguese ancestry (Extended Data Table 2), and meta-
149 analyzed these results using a common effects inverse variance-weighted fixed-effect approach.
150 We detected a strong genome-wide significant (GWS) signal (Fig. 1) for the well-known PSP-
151 risk *MAPT* H1 haplotype (rs8070723 A: $OR[95\%CI]=4.02[3.12-5.19]$; $p=9.89 \cdot 10^{-27}$), and
152 replicated six additional PSP susceptibility *loci* in our study, showing consistent effect directions
153 with previous reports (Extended Data Table 3): *MOBP* (rs1768208 G: $OR[95\%CI]=0.76[0.64-$
154 $0.89]$; $p=2.49 \cdot 10^{-04}$), *EIF2AK3* (rs7571971 C: $OR[95\%CI]=0.80[0.68-0.95]$; $p=9.06 \cdot 10^{-03}$), *STX6*
155 (rs1411478 G: $OR[95\%CI]=0.84[0.71-0.98]$; $p=0.03$), *SLCO1A2* (rs11568563 A:
156 $OR[95\%CI]=0.54[0.42-0.69]$; $p=1.17 \cdot 10^{-06}$), and *DUSP10* (rs6687758 A:
157 $OR[95\%CI]=0.80[0.66-0.97]$; $p=0.02$). After adjusting by the *MAPT* H1 haplotypic background,
158 we also reproduced an additional independent signal within the *MAPT* locus, which partially tags
159 the PSP risk H1c subhaplotype, (rs242557 G: $OR[95\%CI]=0.72[0.61-0.86]$; $p=2.49 \cdot 10^{-04}$).
160 However, the effect of the GWS *RUNX2* locus, reported by Chen *et al.*⁵, was not significant in
161 our analysis (rs6458446 A: $OR[95\%CI]=0.90[0.75-1.09]$; $p=0.27$). Additionally, a recent
162 preprint study¹⁰ reported a GWS signal in the *APOE* locus where *APOE* $\epsilon 4$ was the protective
163 allele and *APOE* $\epsilon 2$ was the risk allele. In our study, we replicated the described *APOE* $\epsilon 4$
164 protective effect (rs429358 C: $OR[95\%CI]=0.72[0.53-0.96]$; $p=0.02$), but *APOE* $\epsilon 2$ was not
165 significant (rs7412 T: $OR[95\%CI]=1.08[0.78-1.49]$; $p=0.64$). Whether a protective mechanism
166 for *APOE* $\epsilon 4$, the most important genetic risk factor for sporadic AD, exists in PSP, or this effect
167 is caused by a selection bias favouring inclusion of cases without amyloid pathology needs to be
168 clearly established in future studies. Finally, the *TNXB/C4A* locus, described by another recent
169 preprint article¹¹, was not replicated (rs369580 G: $OR[95\%CI]=1.25[0.86-1.82]$; $p=0.24$),
170 although its effect direction was concordant in our cohort.

171 Aiming to find novel PSP associations, we conducted a meta-analysis of our Spanish and
172 Portuguese cohort with the results of the Sanchez-Contreras *et al.* study⁴, the largest publicly
173 available summary statistics for PSP genetics, and an additional dataset including 59 clinical PSP-
174 RS subjects and 1513 controls provided by the Amsterdam Dementia Cohort, totally comprising
175 3099 PSP and 11482 controls (Extended Data Table 4). An important limitation of this approach
176 was that, due to the design of the previous studies^{3,4}, only 36 SNPs were available for meta-
177 analysis. We found a novel GWS association of the *NFASC* locus (rs12744678 C: $OR[95\%$

178 $CIJ=0.83[0.78-0.89]$; $p=4.15 \cdot 10^{-08}$), which had null heterogeneity across cohorts ($I^2=0\%$), and
179 showed concordant effect directions in all groups (Fig. 2, Extended Data Table 5). The *NFASC*
180 gene encodes the neurofascin proteins and is highly expressed by neurons and oligodendrocytes.
181 These proteins are involved in the development of axon initial segments and nodes of Ranvier in
182 the central (CNS) and peripheral nervous system (PNS) by forming paranodal junctions, a very
183 specialized cell-cell adhesion complex that is essential for correct myelination and impulse
184 conduction^{12,13}. Moreover, mutations in *NFASC* have been linked to neurodevelopmental
185 disorders with central and peripheral motor dysfunction [61], and anti-neurofascin antibodies are
186 found in a fraction of patients with CNS and PNS demyelinating diseases, such as multiple
187 sclerosis¹⁴. Interestingly, although we did not find any significant QTLs for this variant, we found
188 that rs4951151, the top variant on the locus (Fig. 2), is an eQTL for the neighbouring gene *CNTN2*
189 in cerebellar tissue (on GTEx <https://gtexportal.org/>), and a pQTL for *CNTN2* in CSF and plasma
190 (1:205001174-T-C, on ONTIME <https://ontime.wustl.edu/>)¹⁵. Importantly, *CNTN2* is almost
191 exclusively expressed by oligodendrocytes and is highly coexpressed with *MOBP* and *SLCO1A2*,
192 two known PSP susceptibility loci (Extended Data Fig. 3). These three genes belong to the same
193 expression cluster, which is implicated in myelination and enriched in this specific cell type
194 (Extended Data Fig. 3). *CNTN2* encodes contactin-2 a transmembrane protein also implicated in
195 the organization of nodes of Ranvier at the juxtaparanodal junctions. A study observed that the
196 absence of this protein can impair oligodendrocyte function and axonal conduction, although it
197 did not affect myelination under pathological conditions¹⁶.

198 The finding of the novel *NFASC/CNTN2* locus, along with other replicated loci, such as *MOBP*,
199 *SLCO1A2* and *DUSP10*, highlight the importance of myelination and oligodendrocyte-neuron
200 interactions in PSP etiopathology: *MOBP* encodes the myelin-associated oligodendrocyte basic
201 protein, a vital protein for myelin sheath structural integrity and maintenance in the central
202 nervous system, *SLCO1A2* is highly expressed in the brain, almost exclusively by
203 oligodendrocytes, and is highly correlated to the oligodendrocytes myelination expression
204 cluster¹⁷, and *DUSP10* is highly expressed in oligodendrocytes and inhibitory neurons, encoding
205 a phosphatase involved in MAP kinases inactivation which is able to induce oligodendrocyte
206 differentiation (Extended Data Fig. 4)¹⁸. Gene ontology enrichment, including the PSP genes we
207 replicated and the novel *NFASC* locus, reveals biological features related to myelination and axon
208 ensheathment (Extended Data Table 6). Moreover, coexpression network analysis revealed that
209 *MARK1*, encoding the microtubule affinity regulating kinase 1, a kinase capable of
210 phosphorylating tau promoting microtubule destabilisation¹⁹, is highly coexpressed with *MOBP*,
211 *SLCO1A2* and *NFASC* (Extended Data Fig. 5). Furthermore, a recent, preprint GWAS has also
212 found that several PSP risk variants have oligodendrocyte-specific effects on gene expression¹¹.
213 Oligodendrocytes have a prominent microtubule structure and express *tau*, which plays a role in

214 early contact initiation in axons, maintaining microtubule stability and myelination²⁰. Moreover,
215 tau pathology, which is predominantly neuronal in AD, is commonly found in white matter
216 oligodendrocytes in PSP and other tauopathies²⁰, and deregulation of myelination processes has
217 been linked to neurodegenerative diseases, with a special emphasis on PSP²¹. Overall, genetic
218 findings suggest that oligodendrocyte dysfunction and impaired myelination may play a
219 significant role in the etiopathogenesis of PSP. Additionally, there is increasing evidence
220 suggesting astrocytic tau pathology occurs in the earliest stages of corticobasal degeneration and
221 PSP^{22,23}. This points to a paradigm shift from neuronal to glial cell involvement in early 4-repeat
222 tau pathology.

223 Of note, a small overlap may exist between our samples and some of those present in the GWAS
224 by Höglinger *et al*³, as 23 cases in the former study were recruited by Spanish centres. However,
225 this potential overlap comprises a very small fraction of the samples included in both datasets,
226 thus we do not expect it to substantially bias our results. In fact, PSP GWASs suffer from a
227 considerable “Russian doll” effect⁹, as new reports consist of expansions on previously available
228 cohorts, causing a large overlap between studies. Our study used exclusively Spanish and
229 Portuguese samples for validation of PSP susceptibility loci, and thus provides a highly
230 independent cohort for replication of the previously reported variants (Extended Data Table 3).
231 Importantly, there is a striking lack of diversity in PSP genetic studies to date. Expanding future
232 studies to include non-European populations is crucial for a better understanding of the disease
233 etiology, enhancing genetic discovery and benefiting a broader range of the human population²⁴.

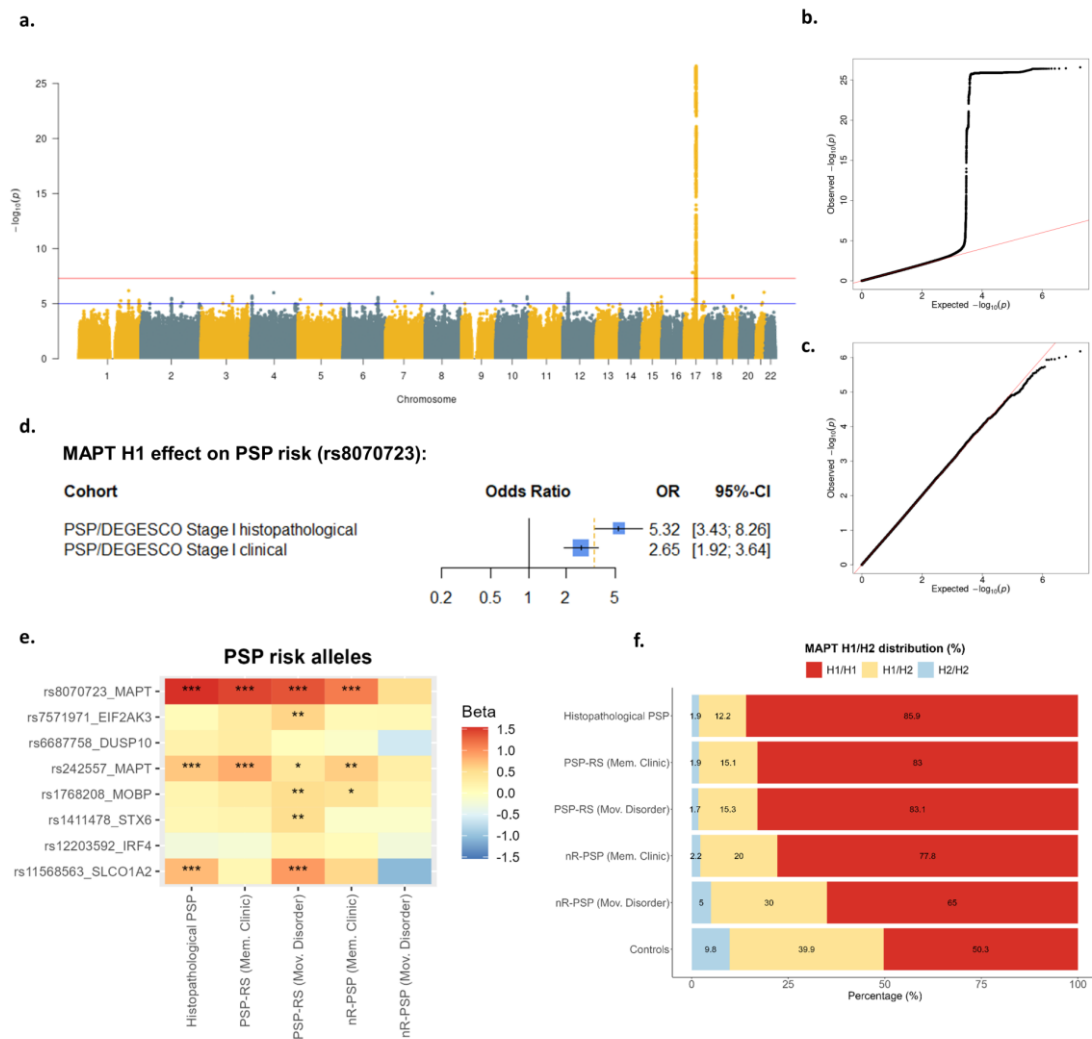
234 **References**

- 235 1. Boxer, A. L. *et al.* Advances in progressive supranuclear palsy: new diagnostic criteria,
236 biomarkers, and therapeutic approaches. *The Lancet Neurology* **16**, 552–563 (2017).
- 237 2. Baker, M. *et al.* Association of an Extended Haplotype in the Tau Gene with Progressive
238 Supranuclear Palsy. *Human Molecular Genetics* **8**, 711–715 (1999).
- 239 3. PSP Genetics Study Group *et al.* Identification of common variants influencing risk of the
240 tauopathy progressive supranuclear palsy. *Nat Genet* **43**, 699–705 (2011).
- 241 4. Sanchez-Contreras, M. Y. *et al.* Replication of progressive supranuclear palsy genome-wide
242 association study identifies SLCO1A2 and DUSP10 as new susceptibility loci. *Mol*
243 *Neurodegeneration* **13**, 37 (2018).

- 244 5. Chen, J. A. *et al.* Joint genome-wide association study of progressive supranuclear palsy
245 identifies novel susceptibility loci and genetic correlation to neurodegenerative diseases. *Mol*
246 *Neurodegeneration* **13**, 41 (2018).
- 247 6. Rojas, I. de *et al.* Common variants in Alzheimer’s disease and risk stratification by
248 polygenic risk scores. *Nature Communications* **12**, 3417 (2021).
- 249 7. Taliun, D. *et al.* Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program.
250 *Nature* **590**, 290–299 (2021).
- 251 8. Höglinger, G. U. *et al.* Clinical diagnosis of progressive supranuclear palsy: The movement
252 disorder society criteria: MDS Clinical Diagnostic Criteria for PSP. *Mov Disord.* **32**, 853–
253 864 (2017).
- 254 9. Escott-Price, V. & Hardy, J. Genome-wide association studies for Alzheimer’s disease:
255 bigger is not always better. *Brain Communications* **4**, fcae125 (2022).
- 256 10. Wang, H. *et al.* Whole-Genome Sequencing Analysis Reveals New Susceptibility Loci
257 and Structural Variants Associated with Progressive Supranuclear Palsy. Preprint at
258 <http://medrxiv.org/lookup/doi/10.1101/2023.12.28.23300612> (2023).
- 259 11. Farrell, K. *et al.* Genetic, Transcriptomic, Histological, and Biochemical Analysis of
260 Progressive Supranuclear Palsy Implicates Glial Activation and Novel Risk Genes. Preprint
261 at <http://biorxiv.org/lookup/doi/10.1101/2023.11.09.565552> (2023).
- 262 12. Sherman, D. L. *et al.* Neurofascins Are Required to Establish Axonal Domains for
263 Saltatory Conduction. *Neuron* **48**, 737–742 (2005).
- 264 13. Zhang, A. *et al.* Neurofascin 140 Is an Embryonic Neuronal Neurofascin Isoform That
265 Promotes the Assembly of the Node of Ranvier. *J. Neurosci.* **35**, 2246–2254 (2015).
- 266 14. Kira, J., Yamasaki, R. & Ogata, H. Anti-neurofascin autoantibody and demyelination.
267 *Neurochemistry International* **130**, 104360 (2019).
- 268 15. Yang, C. *et al.* Genomic atlas of the proteome from brain, CSF and plasma prioritizes
269 proteins implicated in neurological disorders. *Nat Neurosci* **24**, 1302–1312 (2021).
- 270 16. Zoupi, L. *et al.* The function of contactin-2/TAG-1 in oligodendrocytes in health and
271 demyelinating pathology. *Glia* **66**, 576–591 (2018).

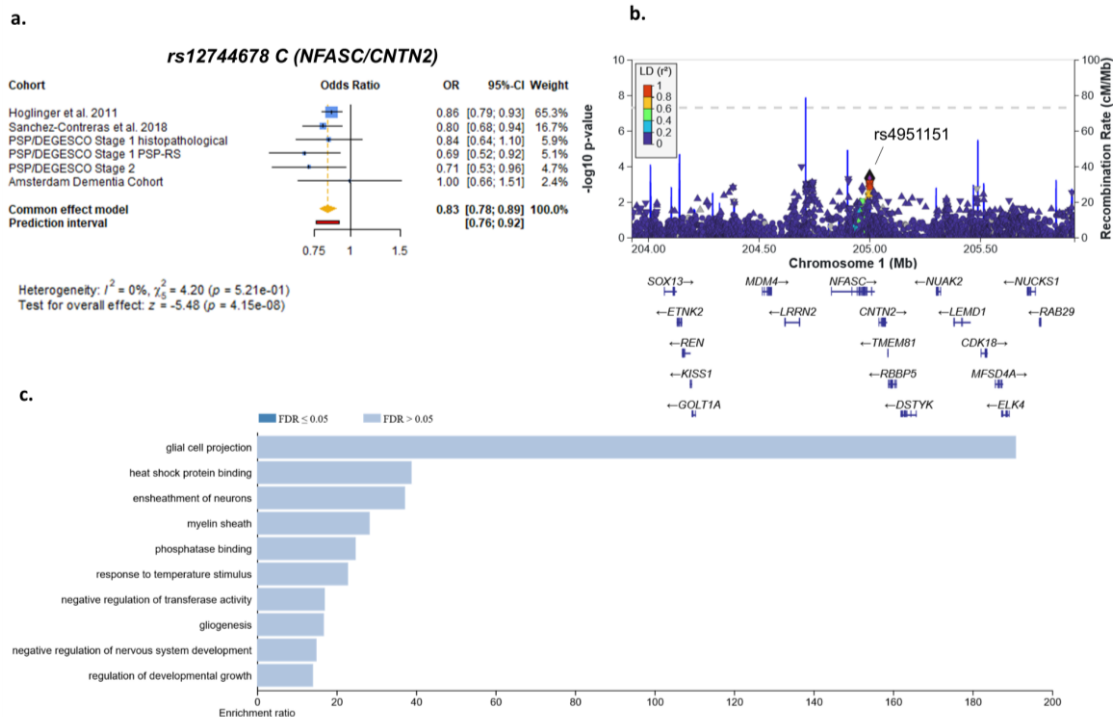
- 272 17. Sjöstedt, E. *et al.* An atlas of the protein-coding genes in the human, pig, and mouse
273 brain. *Science* **367**, eaay5947 (2020).
- 274 18. Gobert, R. P. *et al.* Convergent functional genomics of oligodendrocyte differentiation
275 identifies multiple autoinhibitory signaling circuits. *Mol Cell Biol* **29**, 1538–1553 (2009).
- 276 19. Gu, G. J. *et al.* Role of Individual MARK Isoforms in Phosphorylation of Tau at Ser262
277 in Alzheimer’s Disease. *Neuromol Med* **15**, 458–469 (2013).
- 278 20. Kahlson, M. A. & Colodner, K. J. Glial Tau Pathology in Tauopathies: Functional
279 Consequences. *J Exp Neurosci* **9s2**, JEN.S25515 (2015).
- 280 21. Allen, M. *et al.* Conserved brain myelination networks are altered in Alzheimer’s and
281 other neurodegenerative diseases. *Alzheimer’s & Dementia* **14**, 352–366 (2018).
- 282 22. Briel, N., Pratsch, K., Roeber, S., Arzberger, T. & Herms, J. Contribution of the
283 astrocytic tau pathology to synapse loss in progressive supranuclear palsy and corticobasal
284 degeneration. *Brain Pathology* **31**, e12914 (2021).
- 285 23. Ling, H. *et al.* Astrogliopathy predominates the earliest stage of corticobasal
286 degeneration pathology. *Brain* **139**, 3237–3252 (2016).
- 287 24. Graham, S. E. *et al.* The power of genetic diversity in genome-wide association studies
288 of lipids. *Nature* **600**, 675–679 (2021).

289 **Figures**



290

291 **Fig. 1. PSP GWAS in the Spanish and Portuguese population and evaluation of the clinical**
 292 **subgroups. a.** Manhattan plot displaying PC-adjusted PSP genome-wide associations of common
 293 variants. **b-c.** QQ-plots displaying expected vs. observed p-values in our PSP GWAS including
 294 and excluding chr17, respectively. **d.** Forest plot displaying effects of rs8070723, a variant tagging
 295 the *MAPT* H1 haplotype in our histopathological and clinical Stage I samples. **e.** Effect sizes and
 296 significance of the risk allele of known reported PSP risk variants, as reported by Sanchez-
 297 Contreras *et al*⁴ **f.** Distribution of *MAPT* H1/H2 haplotypes in the different groups. OR, odds
 298 ratio; PSP-RS, Richardson-Syndrome; nR-PSP, non-Richardson PSP.



299

300 **Fig. 2. NFASC/CNTN2 is a novel locus for progressive supranuclear palsy.** **a.** Forest plot
 301 displaying effect sizes and meta-analysis of the rs12744678 C allele among the different cohorts.
 302 **b.** Locus plot of the *NFASC/CNTN2* in cohorts where genome-wide association results were
 303 available (PSP/DEGESCO and ADC). **c.** Gene ontology enrichment ratio calculated using
 304 WebGestalt including PSP replicated loci (*MAPT*, *MOBP*, *EIF2AK3*, *STX6*, *DUSP10*, *SLCO1A2*)
 305 and the novel *NFASC* locus.

306 **Online methods**

307 **PSP/DEGESCO cohort**

308 Stage I samples consisted of 176 histopathologically confirmed PSP samples recruited between
309 2002 and 2021 from Spanish and Portuguese brain biobanks, and 312 samples fulfilling a clinical
310 diagnosis of probable PSP provided by research groups linked to DEGESCO (Supplementary
311 Table 1). Neuropathological diagnosis of PSP was assessed based on the distribution of
312 neurofibrillary tangles in the brain, according to National Institute of Neurological Disorders and
313 Stroke (NINDS) criteria²⁵, while clinical diagnosis was based on the Movement Disorder Society
314 criteria⁸. Briefly, neurologists evaluated four functional domains: ocular motor dysfunction,
315 postural instability, akinesia and cognitive dysfunction. Considering these four domains a
316 probable PSP diagnosis can be assigned to not only patients with PSP-RS, but also to those
317 individuals with atypical clinical presentations, such as PSP with progressive gait freezing, with
318 predominant parkinsonism or with predominant frontal presentation. Controls in this study were
319 obtained from the GR@ACE/DEGESCO cohort (Supplementary Information)⁶. Finally, we
320 recruited an additional batch of PSP samples for further validation of our results (PSP/DEGESCO
321 Stage II), which included 39 histopathological and 154 clinical samples (Supplementary Table 1).
322 Written informed consent was obtained from all participants in this study. This study has been
323 approved by the competent research ethical committee (MED-FACE-2020-01, Universidad
324 Internacional de Catalunya, Sant Cugat del Vallés, Spain).

325 **Amsterdam Dementia Cohort**

326 Patient records of 397 patients diagnosed with PSP and/or corticobasal degeneration within the
327 Amsterdam Dementia Cohort (ADC) between 2000 and 2023 were retrieved²⁶. A clinician
328 retrospectively evaluated each patient file, according the Movement Disorder Society criteria⁸,
329 for ocular motor dysfunction, postural instability, akinesia and cognitive dysfunction. For this
330 study, a total of 59 patients had genotyping available, were unrelated and of European ancestry,
331 and met criteria for PSP-RS. Controls were patients without cognitive complaints from the same
332 cohort (N=1105) and cognitively normal individuals from various other studies: one twin from
333 each pair of twins from the EPAD study²⁷ (N=97), partners of children of participants from the
334 100-plus Study²⁸ (N=83) and controls included in the context of the Parelsnoer Institute²⁹
335 (N=228). The local Medical Ethics Committees have approved the protocols for the ADC and the
336 other studies and all participants of the ADC provided written informed consent.

337 **PSP/DEGESCO: DNA genotyping, QC & imputation**

338 DNA from PSP cases was genotyped by the Spanish National Center for Genotyping (CEGEN)
339 using the Axiom 815K Spanish Biobank Array (Thermo Fisher). We performed genotype calling

340 using Affymetrix power tools (APT) software v1.15.0, and curated calls following the Axiom
341 data analysis workflow. To improve calling accuracy we combined our cases with 5000 controls
342 genotyped in the same center and using the same platform from the GR@ACE/DEGESCO
343 cohort^{6,30}. Called variants passing quality control (QC) criteria defined by the manufacturer
344 (Affymetrix) were kept for downstream analysis (N=737,174). After genotype calling, we
345 removed samples with low genotype call rates (<97%), high heterozygosity (+3SD over mean
346 heterozygosity), discordant genetic-reported sex, and non-European ancestry, based on the 1000G
347 European population cluster. After sample QC, we removed variants with high missingness
348 (>5%), low frequency (MAF<0.01), differential missingness between cases and controls ($p<10^{-5}$)
349 or failing the Hardy Weinberg equilibrium test ($p<10^{-6}$) in the control population. Samples with
350 genotype call rates below 0.97 after variant QC were also removed. Additionally, we excluded
351 duplicated and related individuals within our study by applying a GENESIS³¹ kinship filter of
352 0.046875. For PSP/DEGESCO Stage II samples, we merged the controls and cases passing QC
353 from Stage I during the genotype calling stage and followed the same procedure described above.
354 Additionally, at the IBD filtering step we removed Stage II cases overlapping or related to cases
355 from Stage I. A complete overview of the QC process is detailed at the Supplementary
356 Information section.

357 **Amsterdam Dementia Cohort: genotyping, QC and imputation**

358 We genotyped individuals using the Illumina Global Screening Array and applied established
359 quality control methods³². We used high-quality genotypes in all individuals (individual call rate
360 >99%, variant call rate >99%), individuals with sex mismatches were excluded and departure
361 from Hardy-Weinberg equilibrium was considered significant at $p<1\times 10^{-6}$. Genotypes were then
362 lifted over to GRCh38 and prepared for imputation using provided scripts (HRC-1000G-check-
363 bim.pl) specifying TOPMed as reference panel. This script compares variant ID, strand, and allele
364 frequencies to the TOPMed reference panel (version r2, N=194,512 haplotypes from N=97,256
365 individuals)⁷. Finally, all variants were submitted to the Michigan Imputation server
366 (<https://imputation.biodatacatalyst.nih.gov/>). The server uses EAGLE (v2.4) to phase data
367 and Minimac4 to perform genotype imputation to the reference panel. After quality control and
368 genotype imputation of the genetic data, we kept only individuals of European ancestry (based on
369 1000Genomes clustering)³³, and excluded individuals with a family relation (identity-by-descent
370 ≥ 0.2)³⁴, leaving 59 clinically diagnosed PSP-RS cases and 1513 controls for analysis.

371 **Statistical analysis**

372 For GWAS analysis, we included approximately four sex-matched controls per PSP case
373 available (Extended Data Table 2), as the increase in statistical power above this control-to-case
374 ratio is generally negligible³⁵. We removed rare SNPs (MAF<0.01) or those with low imputation

375 quality ($R^2 < 0.3$), leaving us with ~9M variants for association testing. To test the association of
376 SNP allele dosages with PSP risk we fitted logistic regression models using PLINK v2.00a3.7LM
377 software. We adjusted the models by the first four PCs to account for population microstructure.
378 As we used population-based controls in our analysis, we did not adjust the models by age to
379 avoid jeopardizing small effects, since the large age differences between cases and controls would
380 lead to adjustment for ‘caseness’, biasing the analysis towards the null³⁶. In addition, to ensure
381 that the association of tested variants was not age-driven, we ran PC-adjusted linear regressions
382 for age in the full subset (N=1,450) of Stage I controls (Supplementary Table 2). Additionally, to
383 check for H1/H2 independent associations around the *MAPT* gene, we ran models adjusted by the
384 rs1122380 genotype, which is in full linkage disequilibrium (LD) with rs1800547 (CEU:
385 $R^2 = 1.00$; $D' = 1.00$), the SNP previously used to report *MAPT* H1/H2 haplotype effects in our
386 population³⁷. We performed fixed effect inverse variance-weighted meta-analysis of the discovery
387 and replication sets using the METAL software³⁸. Locus association plots were made using
388 LocusZoom (<http://locuszoom.org/>). All additional analyses, processing and visual representation
389 of the data were performed using R software v4.1.1. We strand corrected and flipped effect alleles
390 to align our results with the previous ones when needed and used the meta R package³⁹ to run
391 fixed effect inverse variance-weighted meta-analysis on specific SNPs.

392 **In-silico functional analysis**

393 Enrichment analysis was performed using WebGestalt⁴⁰ (<https://www.webgestalt.org/>).
394 Expression of target genes was assessed using GTEx Portal and Human Protein Atlas for RNA
395 and protein levels, respectively. The Online Neurodegenerative Trait Integrative Multi-Omics
396 Explorer (ONTIME) web resource was accessed to check association of variants with quantitative
397 endophenotypes¹⁵. Finally, gene co-expression networks were generated using GeneFriends
398 (<https://www.genefriends.org/>)⁴¹. Details on these methodologies are provided in the
399 Supplementary Information.

400 **Data availability**

401 The data that support the findings of this study are not openly available due to reasons of
402 sensitivity and are available from the corresponding author upon reasonable request. Summary
403 statistics detailing association results will be made publicly available upon acceptance of the
404 manuscript.

405 **Code availability**

406 Custom code was used to calculate polygenic risk scores and is publicly available
407 (https://github.com/Pablo-GarGon/PRS_Generator). Other analyses were performed using

408 publicly available software, tools and algorithms, as detailed in the Methods and Supplementary
409 Information sections.

410 **References**

- 411 25. Hauw, J.-J. *et al.* Preliminary NINDS neuropathologic criteria for Steele-Richardson-
412 Olszewski syndrome (progressive supranuclear palsy). *Neurology* **44**, 2015–2015 (1994).
- 413 26. Van Der Flier, W. M. & Scheltens, P. Amsterdam Dementia Cohort: Performing
414 Research to Optimize Care. *JAD* **62**, 1091–1111 (2018).
- 415 27. Konijnenberg, E. *et al.* The EMIF-AD PreclinAD study: study design and baseline
416 cohort overview. *Alz Res Therapy* **10**, 75 (2018).
- 417 28. Holstege, H. *et al.* The 100-plus Study of cognitively healthy centenarians: rationale,
418 design and cohort description. *Eur J Epidemiol* **33**, 1229–1249 (2018).
- 419 29. Jansen, I. E. *et al.* Genome-wide meta-analysis for Alzheimer’s disease cerebrospinal
420 fluid biomarkers. *Acta Neuropathol* **144**, 821–842 (2022).
- 421 30. Moreno-Grau, S. *et al.* Genome-wide association analysis of dementia and its clinical
422 endophenotypes reveal novel loci associated with Alzheimer’s disease and three causality
423 networks: The GR@ACE project. *Alzheimer’s & Dementia* **15**, 1333–1347 (2019).
- 424 31. Gogarten, S. M. *et al.* Genetic association testing using the GENESIS R/Bioconductor
425 package. *Bioinformatics* **35**, 5346–5348 (2019).
- 426 32. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat Genet* **48**,
427 1284–1287 (2016).
- 428 33. The 1000 Genomes Project Consortium *et al.* A global reference for human genetic
429 variation. *Nature* **526**, 68–74 (2015).
- 430 34. Anderson, C. A. *et al.* Data quality control in genetic case-control association studies.
431 *Nature Protocols* **5**, 1564–1573 (2011).
- 432 35. Taylor, J. M. Choosing the number of controls in a matched case-control study, some
433 sample size, power and efficiency considerations. *Stat Med* **5**, 29–36 (1986).
- 434 36. Baker, E. *et al.* What does heritability of Alzheimer’s disease represent? *PLoS ONE* **18**,
435 e0281440 (2023).

- 436 37. Sánchez-Juan, P. *et al.* The MAPT H1 Haplotype Is a Risk Factor for Alzheimer's
437 Disease in APOE ϵ 4 Non-carriers. *Front. Aging Neurosci.* **11**, 327 (2019).
- 438 38. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of
439 genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
- 440 39. Balduzzi, S., Rücker, G. & Schwarzer, G. How to perform a meta-analysis with R: a
441 practical tutorial. *Evid Based Mental Health* **22**, 153–160 (2019).
- 442 40. Wang, J., Vasaikar, S., Shi, Z., Greer, M. & Zhang, B. WebGestalt 2017: a more
443 comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit.
444 *Nucleic Acids Research* **45**, W130–W137 (2017).
- 445 41. Raina, P. *et al.* GeneFriends: gene co-expression databases and tools for humans and
446 model organisms. *Nucleic Acids Research* **51**, D145–D158 (2023).

447 Acknowledgements

448 The present work has been performed as part of the Doctoral thesis of Pablo García-González at
449 the University of Barcelona (Barcelona, Spain). We would like to thank patients and controls who
450 participated in this project. Some control samples and data from patients included in this study
451 were provided in part by the National DNA Bank Carlos III (www.bancoadn.org, University of
452 Salamanca, Spain) and Hospital Universitario Virgen de Valme (Sevilla, Spain); they were
453 processed following standard operating procedures with the appropriate approval of the Ethical
454 and Scientific Committee. We are grateful to the Biomedic Diagnostic Center (CBD) of the
455 Hospital Clinic de Barcelona for access to, and assistance with, FLUOSTAR Omega equipment
456 and RT-QuIC experiments (Drs. Raquel Ruiz and Laura Naranco). Hospital Clinic de Barcelona
457 receives support from the CERCA programme from Generalitat de Catalunya (Catalan
458 government). Several authors of this manuscript are members of the European Reference Network
459 for Rare Neurological Diseases - Project ID No 739510, as well as CIBERNED (CB06/05/0018-
460 ISCIII). We acknowledge the Biobank of the University of Navarra for its collaboration. We
461 thank all study participants and all personnel involved in data collection for the contributing
462 studies. Research of Alzheimer Center Amsterdam is part of the neurodegeneration research
463 program of Amsterdam Neuroscience. Alzheimer Center Amsterdam is supported by Stichting
464 Alzheimer Nederland and Stichting Steun Alzheimercentrum Amsterdam. The chair of Wiesje
465 van der Flier is supported by the Pasma stichting. The clinical database structure was developed
466 with funding from Stichting Dioraphte. The work in this manuscript was carried out on the
467 Snellius supercomputer, which is embedded in the Dutch national e-infrastructure with the
468 support of SURF Cooperative. Computing hours were granted in 2016, 2017, 2018 and 2019 to
469 H. Holstege by the Dutch Research Council (project name: ‘100plus’; project numbers 15318 and
470 17232). Samples and data from patients included in this study were provided by the Vall d’Hebron
471 University Hospital Biobank (PT20/00107), integrated in the Spanish National Biobanks
472 Network, and they were processed following standard operating procedures with the appropriate
473 approval of the Ethical and Scientific Committees. We are indebted to the Biobank-Hospital
474 Clinic-FRCB-IDIBAPS for samples and data procurement. This work was funded by CIBERNED
475 (Centro Investigación Biomédica en Red Enfermedades Neurodegenerativas) 10th internal call
476 for cooperative projects (Register number: 2019/09). Agustín Ruiz, Pascual Sanchez-Juan,
477 Alberto Rábano, and Maria J. Bullido received funding from this convocatory. Pablo García-
478 González received support by CIBERNED employment plan CNV-304-PRF-866. CIBERNED is
479 integrated into ISCIII (Instituto de Salud Carlos III). Mercé Boada, Marta Marquíé and Agustín
480 Ruiz are also supported by national grants PI13/02434, PI16/01861, PI17/01474, PI19/01240,
481 PI19/01301, PI19/00335 and PI22/011403, and CIBERNED grant 2019/08. The Genome
482 Research @ Fundació ACE project (GR@ACE) is supported by Grifols SA, Fundació bancaria
483 “La Caixa”, Fundació ACE, and CIBERNED. Acció Estratègica en Salut is integrated into the
484 Spanish National R + D + I Plan and funded by ISCIII—Subdirección General de Evaluación and
485 the Fondo Europeo de Desarrollo Regional (FEDER—“Una manera de hacer Europa”). Agustín
486 Ruiz is supported by ISCIII national grant PMP22/00022, funded by the European Union
487 (NextGenerationEU). The genotyping service to generate GR@ACE and PSP/DEGESCO GWAS
488 data was carried out at CEGEN-PRB3-ISCIII; it is supported by grant PT17/0019, of the PE I+D+i
489 2013-2016, funded by ISCIII and ERDF”. Itziar de Rojas received support by a national grant
490 from ISCIII FI20/00215. Amanda Cano acknowledges the support of the Spanish Ministry of
491 Science, Innovation and Universities under the grant Juan de la Cierva (FJC2018-036012-I), the
492 support of ISCIII under the grant Sara Borrell (CD22/00125), Spanish Ministry of Science and
493 Innovation, Proyectos de Generación de Conocimiento grant PID2021-122473OA-I00, and

494 Fundación ADEY under the program “Proyectos de Investigación en Salud 2023”. Fondo de
495 Investigación en Salud (Instituto Carlos III) (PI17/00096) in the frame of the European Regional
496 Development Fund (ERDF) and to Fundació la Marató de TV3 (PI043296 and 202009-10). This
497 work was supported by the Spanish Ministry of Science and Innovation (RTC2019-007150-1),
498 the Instituto de Salud Carlos III (ISCIII) and co-funded by the European Union (PI14/01823,
499 PI16/01575, PI18/01898, PI19/01576, PI21/01875), the Consejería de Economía, Innovación,
500 Ciencia y Empleo de la Junta de Andalucía (CVI-02526, CTS-7685, PY20_00896), and the
501 Consejería de Salud y Bienestar Social de la Junta de Andalucía (PI-0471-2013, PE-0210-2018,
502 PI-0459-2018, PE-0186-2019). Pilar Gómez-Garre was supported by the “Nicolás Monardes”
503 program (C-0048-2017) from Andalusian Regional Ministry of Health. Laura Muñoz-Delgado
504 was supported by the “Río Hortega” program [CM21/00051] from the Instituto de Salud Carlos
505 III (ISCIII-FEDER). The funders had no role in study design, data collection and analysis,
506 decision to publish, or preparation of the manuscript. Sven J. van der Lee was funded for this
507 study by NWO (#733050512, PROMO-GENODE: a PROspective study of MONoGENic causes
508 Of Dementia) a substantial donation by Edwin Bouw Fonds and Dioraphte. SvdL further received
509 funding for the GeneMINDS consortium, which is powered by Health~Holland, Top Sector Life
510 Sciences & Health. Sven J. van der Lee is a recipient of ABOARD, which is a public-private
511 partnership receiving funding from ZonMW (#73305095007) and Health~Holland, Topsector
512 Life Sciences & Health (PPP-allowance; #LSHM20106). More than 30 partners participate in
513 ABOARD. ABOARD also receives funding from Edwin Bouw Fonds and Gieskes-Strijbisfonds.
514 ABOARD also receives funding from de Hersenstichting, Edwin Bouw Fonds and Gieskes-
515 Strijbisfonds. Array genotyping was performed in the context of EADB (European Alzheimer
516 DNA biobank) funded by the JPCo-fuND FP-829-029 (ZonMW projectnumber 733051061). This
517 work has been supported by the Queen Sofia Foundation. Pascual Sánchez-Juan is supported by
518 grants from ISCIII (PMP22/00022 and PI20/01011) and TED2021-131676B-100. Luis Miguel
519 Real received support from “The absence of seroconversion after exposition to hepatitis C virus
520 is not related to KIR-HLA genotype combinations” (GEHEP-012 study). Consejería de Salud de
521 la Junta de Andalucía (grant number PI-0001/2017) and CIBERINFEC -Consorcio Centro de
522 Investigación Biomédica en Red de Enfermedades Infecciosas- Instituto de Salud Carlos III,
523 Ministerio de Ciencia e Innovación y Unión Europea (Spain) (NextGeneration EU) (grant number
524 CB21/13/00118). Jose Luis Royo received support from Project 23.04.2021 from Santángela
525 Foundation (Sevilla, Spain).

526 **Contributions**

527 P.Gar-Gon. wrote the manuscript. Y.C., V.F., M.Bu., A.Ra., P.S-J., and A.Ru. jointly supervised
528 the work. H.R.L., M.Bu., A.Ra., P.S-J., and A.Ru. designed and conceptualized the study. P.Gar-
529 Gon., S.vdL., and M.H performed bioinformatics and statistical analyses. P.Gar-Gon., M.F.,
530 S.vdL., I.dR., L.S., R.P., C.O., Y.P., I.Q., A.C., and H.H. were involved in data generation.
531 H.R.L., Y.C., M.F., S.vdL., L.S., C.P., A.Cam., E.M., M.J.M., F.V., I. I-G., J.P., I. D-I., J, K, J.F.,
532 A.Ll., S.CM.dB., Y.P., R.D-B., L.M-D., D.B-R., P.Gom-Gar., I.A., G.Ar., J.H.V., M.Me., J.P-T-
533 , P.J.V., A.d.B., J.M.P., A.M.M., L.T., L.M.R., JL. R., M.E.E., C.G., D.C.T., M.B.E., B.S.M.,
534 S.T.B., D.V.R., I.H., A.S-S., B.C-F., S.R.L., R.B.M., A.La., R.T., G.Am., E.M.S., M.F-M., C.L-
535 M., V.A., L.M.P., H.H., P.M., O.B., and M.Bo. were involved in sample contribution. All authors
536 critically revised the manuscript for important intellectual content and approved this final version.

537 **Competing interests**

538 Vigil Neuroscience and Prevail therapeutics are in a public-private partnership research program
539 of Sven J. van der Lee. All funding is paid to his institution.

Extended Data

Extended Data Table 1. Summary of samples remaining after QC steps in the PSP/DEGESCO cohort.

<i>PSP/DEGESCO Stage I</i>		PSP		Controls		
QC step:	N	Fail	Fail (%)	N	Fail	Fail (%)
DNA QC	488	88	18.03%	NA	NA	NA
Calling	400	8	2.00%	5000	26	0.52%
Heterozygosity	392	3	0.77%	4974	4	0.08%
Sexcheck	389	10	2.57%	4970	0	0.00%
PCA	379	6	1.58%	4970	15	0.30%
IBD	373	20	5.36%	4955	300	6.05%
FINAL	353			4655		
<i>PSP/DEGESCO Stage II</i>		PSP		Controls		
QC step:	N	Fail	Fail (%)	N	Fail	Fail (%)
DNA QC	193	0	0.00%	NA	NA	NA
Calling	193	12	6.22%	5000	12	0.24%
Heterozygosity	181	0	0.00%	4988	0	0.00%
Sexcheck	181	1	0.55%	4988	0	0.00%
PCA	180	0	0.00%	4988	10	0.20%
IBD	180	11	6.11%	4978	302	6.07%
FINAL	169			4676		
N, sample size						

Extended Data Table 2. Demographic features of samples included in the Spanish and Portuguese PSP GWAS used for replication of previous known PSP *loci*.

	CA			CO		
	N	%Males	Age* (SD)	N	%Males	Age* (SD)
PSP/DEGESCO Stage I (histopathological)	156	56.4	76.5 (9.5)	650	56.5	67.8 (10.1)
PSP/DEGESCO Stage I (PSP-RS)	112	48.2	67.2 (7.8)	800	46.8	68.6 (10.3)
PSP/DEGESCO Stage II (hist. + PSP-RS)	119	54.6	69.2 (8.2)	500	54.6	68.1 (10.5)
TOTAL	465			1950		

N, sample size; SD, standard deviation.

*Controls: Mean age at last visit; histopathological PSP: Mean age at death; clinical PSP: Mean age at motor symptoms onset.

Extended Data Table 3. Summary statistics of the replication in Spanish and Portuguese population of PSP GWS susceptibility *loci* reported to date. The direction column shows the effect directions in the PSP/DEGESCO Stage I histopathological, PSP/DEGESCO Stage I PSP-RS, and PSP/DEGESCO Stage II, respectively. Genomic coordinates correspond to the GRCh38 assembly. EA, effect allele; OA, other allele; OR, odds ratio; SE, standard error; P, p-value; CI, confidence interval.

rs	CHR	POS	EA	OA	Gene	OR*** (reported)	PSP/DEGESCO					Direction
							OR	SE	P	CI 2.5%	CI 97.5%	
rs8070723	chr17	46003698	A	G	<i>MAPT</i>	5.51 ¹	4.02	0.13	9.89·10 ⁻²⁷	3.12	5.19	+++
rs242557*	chr17	45942346	G	A	<i>MAPT</i>	0.51 ¹	0.72	0.09	2.49·10 ⁻⁰⁴	0.61	0.86	---
rs1768208	chr3	39481512	G	A	<i>MOBP</i>	0.69 ¹	0.76	0.08	1.02·10 ⁻⁰³	0.64	0.89	---
rs7571971	chr2	88595833	C	T	<i>EIF2AK3</i>	0.77 ¹	0.80	0.09	9.06·10 ⁻⁰³	0.68	0.95	---
rs1411478	chr1	180993146	G	A	<i>STX6</i>	0.79 ¹	0.84	0.08	2.63·10 ⁻⁰²	0.71	0.98	---
rs11568563	chr12	21304500	A	C	<i>SLCO1A2</i>	0.67 ¹	0.54	0.13	1.17·10 ⁻⁰⁶	0.42	0.69	---
rs6687758	chr1	221991606	A	G	<i>DUSP10</i>	0.80 ¹	0.80	0.10	2.03·10 ⁻⁰²	0.66	0.97	---
rs6458446	chr6	45532733	A	G	<i>RUNX2</i>	0.77 ²	0.90	0.10	0.27	0.75	1.09	+-
rs429358	chr19	44908684	C	T	<i>APOE (ε4)</i>	0.55 ³	0.72	0.15	2.46·10 ⁻⁰²	0.53	0.96	---
rs7412	chr19	44908822	T	C	<i>APOE (ε2)</i>	2.39 ³	1.08	0.16	0.64	0.78	1.49	++

rs369580**	chr6	32052461	A	G	<i>TNXB/C4A</i>	1.43 ⁴	1.25	0.19	0.24	0.86	1.82	++NA
-------------------	------	----------	---	---	-----------------	-------------------	------	------	------	------	------	------

* Adjusted by MAPT H1 dose (rs1122380)

** This variant was not correctly imputed in PSP/DEGESCO Stage II.

***Study: 1 - Sanchez-Contreras *et al.* 2018¹. 2 - Chen *et al.* 2018², 3- Wang *et al.* 2023 (medRxiv)³; 4- Farrell *et al.* 2023 (BioRxiv)⁴

Extended Data Table 4. Main features of the cohorts available for meta-analysis.

Study	Variants available	PSP			Control		
		N	Males (%)	Age (SD)	N	Males (%)	Age (SD)
Höglinger <i>et al.</i> 2011 ⁵	4,099	2120	54.0	~66.5 (NA)	6847	NA	NA
Sanchez-Contreras <i>et al.</i> 2018 ¹	37	533	55.0	75.0 (7.6)	1172	52.5	77.0 (8.9)
PSP/DEGESCO Stage I (histopathological)	8,990,045	156	56.4	76.5 (9.5)	650	56.5	67.8 (10.1)
PSP/DEGESCO Stage I (PSP-RS)	9,008,824	112	48.2	67.2 (7.8)	800	46.8	68.6 (10.3)
PSP/DEGESCO Stage II (hist. + PSP-RS)	8,433,148	119	54.6	69.2 (8.2)	500	54.6	68.1 (10.5)
Amsterdam Dementia Cohort (PSP-RS)	8,794,028	59	54.2	67.7 (5.5)	1513	58.5	62.3 (10.5)
Meta-analysis	36	3099			11482		

N, sample size; SD, standard deviation.

*Controls: Mean age at last visit; histopathological PSP: Mean age at death; clinical PSP: Mean age at motor symptoms onset.

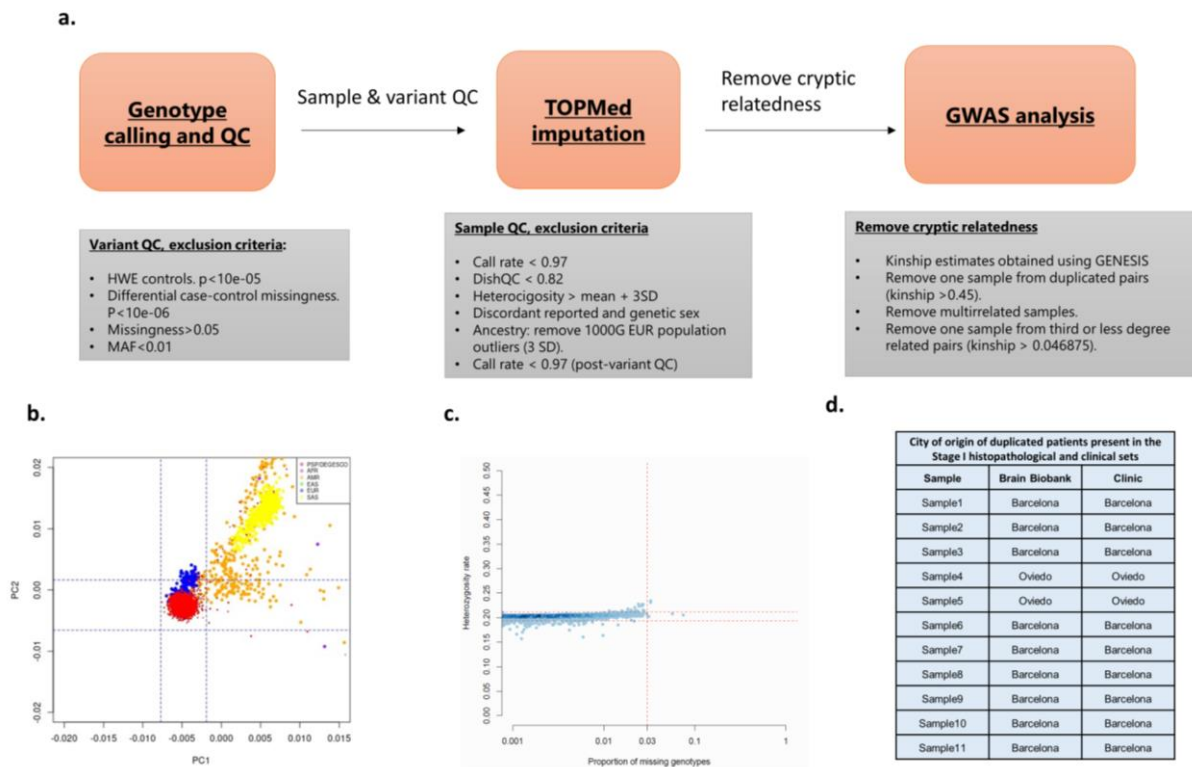
Extended Data Table 5. Summary statistics of the meta-analysis of the 36 SNPs available SNPs including data from the Höglinger *et al.*⁵ and Sanchez-Contreras *et al.*¹ studies, PSP/DEGESCO and ADC. Genomic coordinates correspond to the GRCh38 assembly. EA, effect allele; OA, other allele; OR, odds ratio; P, p-value; EAF, effect allele frequency; SE, standard error; CI, confidence interval; I², heterogeneity estimate.

****SUPPLIED SEPARATELY IN EXCEL FORMAT****

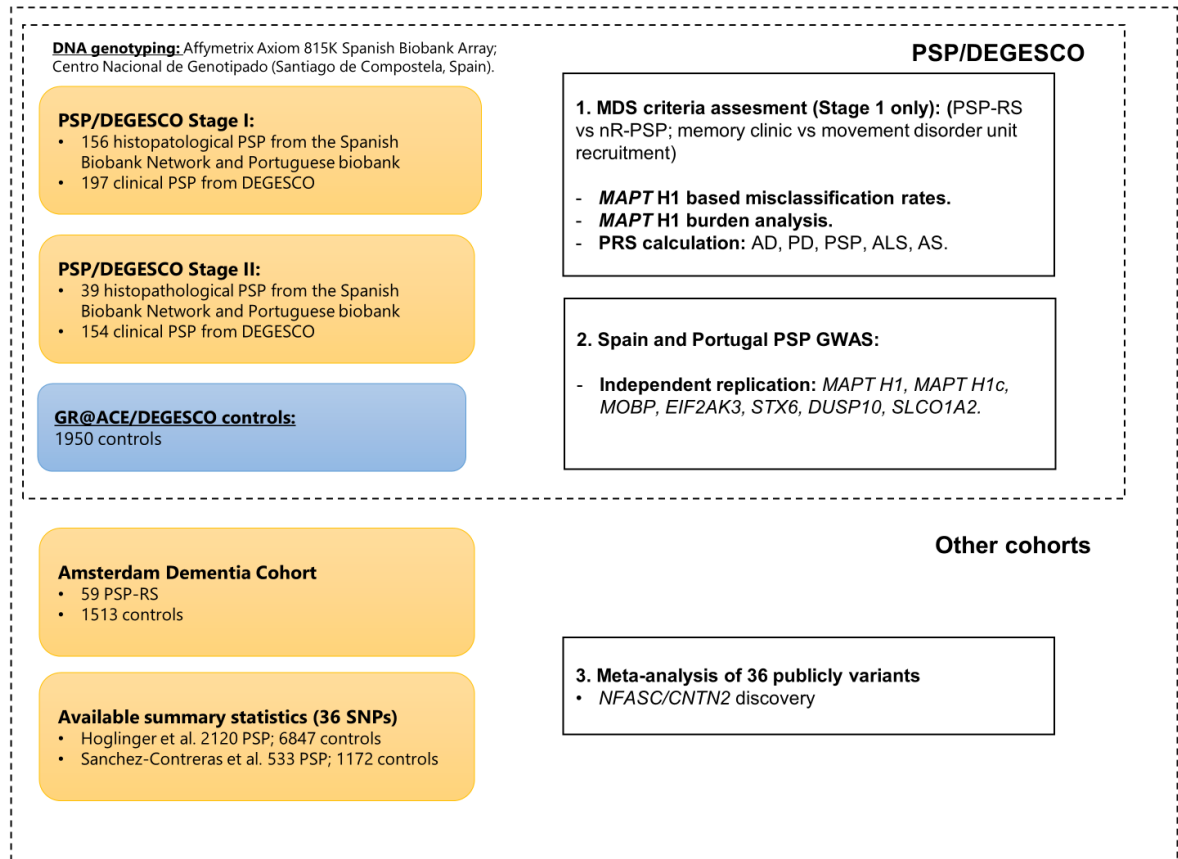
Extended Data Table 6. Top enriched genesets using the WebGestalt Over Representation Analysis including the 6 PSP loci replicated in the PSP/DEGESCO cohort (*MAPT*, *MOBP*, *EIF2AK3*, *STX6*, *DUSP10* and *SLCO1A2*) and the novel *NFASC* locus. Size depicts the number of genes in the geneset. Overlap depicts the number of genes in our input list present in the geneset. Expect depicts the number of genes expected to overlap with each geneset at random given the number of imputed genes.

geneSet	Description	Size	Overlap	Expect	Enrichment Ratio	P	FDR	database
GO:0097386	glial cell projection	23	2	0.01047563	190.9	4.48E-05	0.06	Cellular Component
GO:0031072	heat shock protein binding	113	2	0.05146724	38.9	1.10E-03	0.52	Molecular Function
GO:0007272	ensheathment of neurons	118	2	0.05374455	37.2	1.20E-03	0.52	Biological Process
GO:0043209	myelin sheath	155	2	0.07059666	28.3	2.05E-03	0.67	Cellular Component
GO:0019902	phosphatase binding	177	2	0.08061683	24.8	2.67E-03	0.68	Molecular Function
GO:0009266	response to temperature stimulus	192	2	0.08744876	22.9	3.13E-03	0.68	Biological Process
GO:0051348	negative regulation of transferase activity	258	2	0.11750927	17.0	5.58E-03	0.69	Biological Process
GO:0042063	gliogenesis	261	2	0.11887566	16.8	5.70E-03	0.69	Biological Process
GO:0051961	negative regulation of nervous system development	294	2	0.13390591	14.9	7.19E-03	0.69	Biological Process
GO:0048638	regulation of developmental growth	312	2	0.14210424	14.1	8.06E-03	0.69	Biological Process

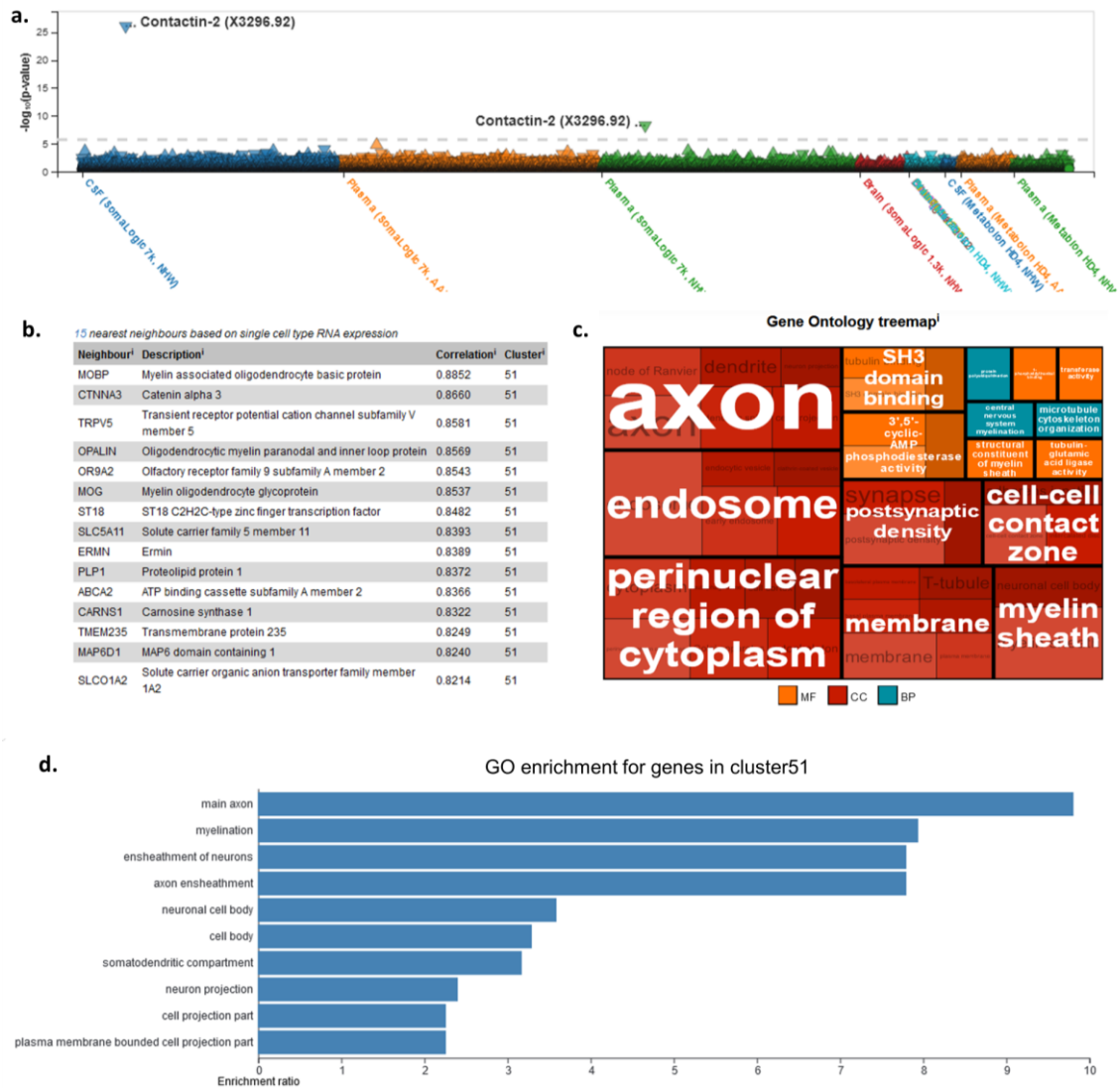
P, p-value; FDR, false discovery rate adjusted p-value.



Extended Data Fig. 1.a. Overview of the quality control process for PSP cases and controls. **b.** Scatterplot of the first two principal components calculated using the 1000 Genomes dataset⁶, showing the different 1000G populations and the samples in PSP/DEGESCO Stage I. We removed samples in PSP/DEGESCO that were outliers (3 standard deviations away from the mean) with respect to the 1000G European cluster. **c.** Scatterplot representing sample call rates and heterozygosity rates in PSP/DEGESCO Stage I. Contaminated samples are expected to display an increase in both the proportion of uncalled genotypes and heterozygosity rate. **d.** City of origin of duplicated individuals found across PSP/DEGESCO Stage I clinics and brain biobanks.

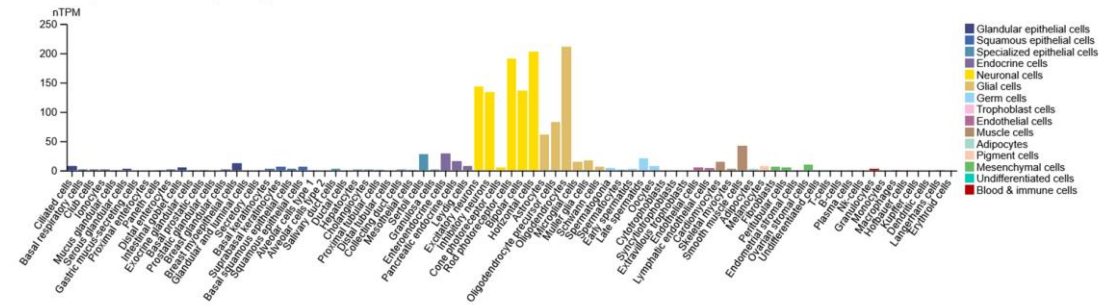


Extended Data Fig. 2. Schematic depiction of cohorts used and steps taken in our study.

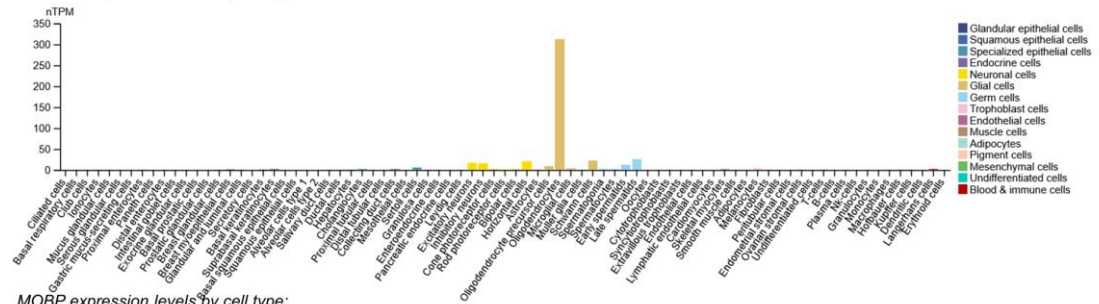


Extended Data Fig. 3. **a.** Association of rs4951151, the top variant detected in the *NFASC/CNTN2* locus with quantitative endophenotypes available on ONTIME <https://ontime.wustl.edu/>⁷. Significant signals were observed for contactin-2 protein levels in CSF and plasma. **b.** Top correlated genes in the Human Protein Atlas single cell expression cluster⁸ for *CNTN2*, referred to as ‘cluster 51’. **c.** Gene ontology treemap of the expression cluster 51, showing the main molecular functions, cellular components and biological processes enriched among these genes. **d.** Gene ontology enrichment for genes in cluster 51 performed using WebGestalt⁹.

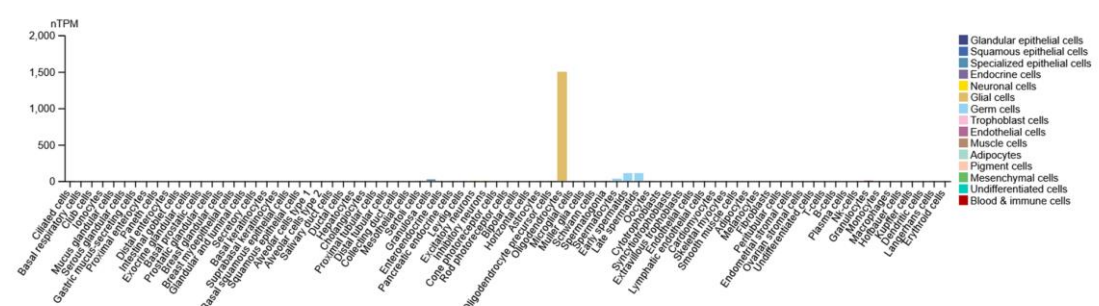
NFASC expression levels by cell type:



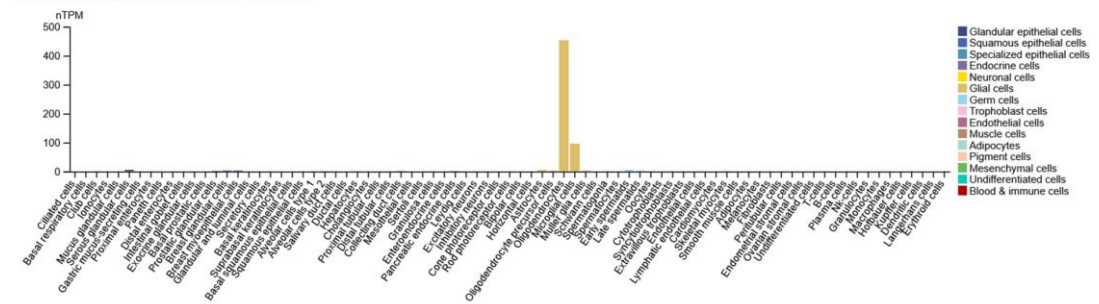
CNTN2 expression levels by cell type:



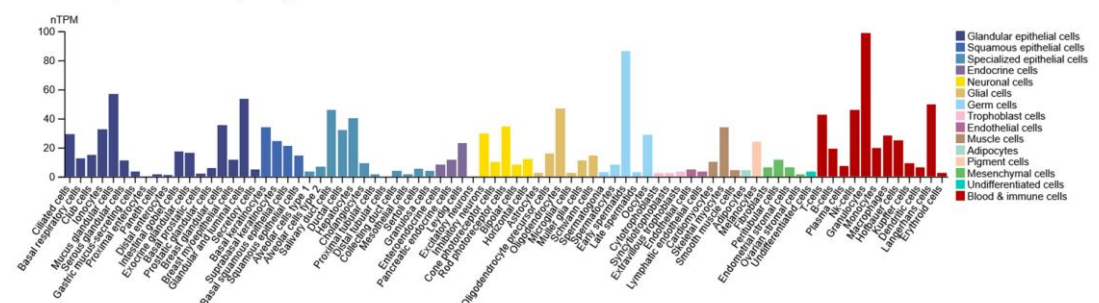
MOBP expression levels by cell type:



SLCO1A2 expression levels by cell type:

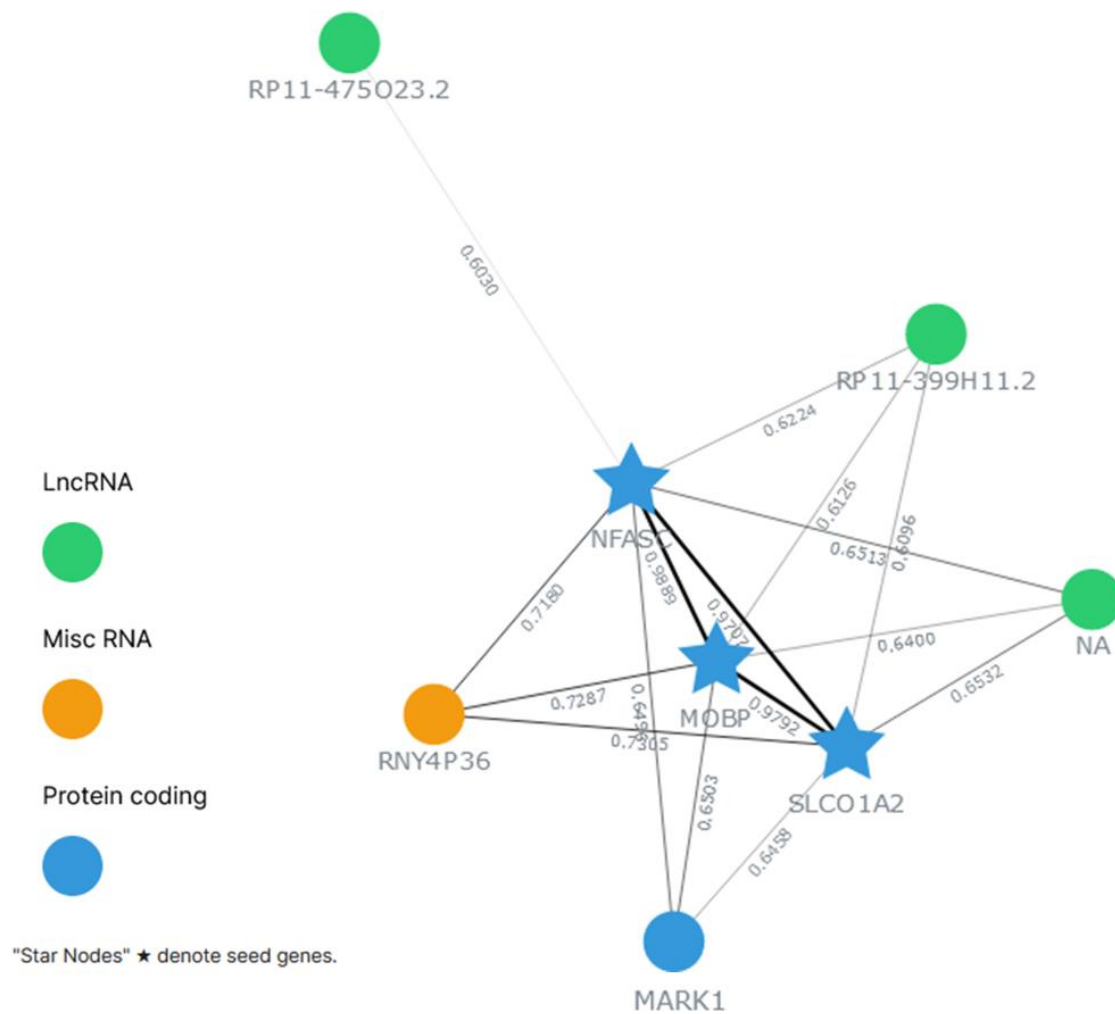


DUSP10 expression levels by cell type:



Extended data Fig. 4. RNA expression levels by cell type for *MOBP*, *SLCO1A2*, *DUSP10*, *NFASC* and *CNTN2*, respectively, based on data from the Human Protein Atlas⁸.

1



2

3

4 **Extended Data Fig. 5.** Gene coexpression network generated using GeneFriends¹⁰ from our input
5 gene list: *MAPT*, *MOBP*, *EIF2AK3*, *STX6*, *DUSP10*, *SLCO1A2*, *NFASC* and *CNTN2*. Edges were
6 labelled with the Pearson correlation coefficient existent between transcripts belonging to the
7 network.

8

9 References

- 10 1. Sanchez-Contreras, M. Y. *et al.* Replication of progressive supranuclear palsy genome-wide
11 association study identifies SLCO1A2 and DUSP10 as new susceptibility loci. *Mol*
12 *Neurodegeneration* **13**, 37 (2018).
- 13 2. Chen, J. A. *et al.* Joint genome-wide association study of progressive supranuclear palsy
14 identifies novel susceptibility loci and genetic correlation to neurodegenerative diseases. *Mol*
15 *Neurodegeneration* **13**, 41 (2018).
- 16 3. Wang, H. *et al.* Whole-Genome Sequencing Analysis Reveals New Susceptibility Loci and
17 Structural Variants Associated with Progressive Supranuclear Palsy. Preprint at
18 <http://medrxiv.org/lookup/doi/10.1101/2023.12.28.23300612> (2023).
- 19 4. Farrell, K. *et al.* Genetic, Transcriptomic, Histological, and Biochemical Analysis of
20 Progressive Supranuclear Palsy Implicates Glial Activation and Novel Risk Genes. Preprint
21 at <http://biorxiv.org/lookup/doi/10.1101/2023.11.09.565552> (2023).
- 22 5. PSP Genetics Study Group *et al.* Identification of common variants influencing risk of the
23 tauopathy progressive supranuclear palsy. *Nat Genet* **43**, 699–705 (2011).
- 24 6. The 1000 Genomes Project Consortium *et al.* A global reference for human genetic variation.
25 *Nature* **526**, 68–74 (2015).
- 26 7. Yang, C. *et al.* Genomic atlas of the proteome from brain, CSF and plasma prioritizes
27 proteins implicated in neurological disorders. *Nat Neurosci* **24**, 1302–1312 (2021).
- 28 8. Karlsson, M. *et al.* A single-cell type transcriptomics map of human tissues. *Sci. Adv.* **7**,
29 eabh2169 (2021).
- 30 9. Wang, J., Vasaikar, S., Shi, Z., Greer, M. & Zhang, B. WebGestalt 2017: a more
31 comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit.
32 *Nucleic Acids Research* **45**, W130–W137 (2017).
- 33 10. Raina, P. *et al.* GeneFriends: gene co-expression databases and tools for humans and
34 model organisms. *Nucleic Acids Research* **51**, D145–D158 (2023).

35

36 Supplementary Information: A Novel Susceptibility Locus in *NFASC*
37 Highlights Oligodendrocytes and Myelination in Progressive Supranuclear
38 Palsy Pathology

39 1. Supplementary notes: DNA genotyping, QC & imputation in the PSP/DEGESCO cohort

40 DNA from PSP cases was extracted from 25 mg of cerebellar cortex (histopathological PSP) or
41 peripheral blood (clinical PSP). After excluding samples with insufficient concentration (<5
42 ng/ μ L) or degraded DNA, samples were genotyped by the Spanish National Center for
43 Genotyping (CEGEN) using the Axiom 815K Spanish Biobank Array (Thermo Fisher). We
44 performed genotype calling and QC using Affymetrix power tools (APT) software v1.15.0
45 following the Axiom data analysis workflow. We kept samples with dishQC values >0.82, a
46 parameter determined by measuring the resolution of AT and GC channels within monomorphic
47 SNPs, and genotype call rates >0.97. We also checked that the average call rate within plates was
48 >0.985 to detect potentially problematic batches. We jointly performed genotype calling for the
49 remaining PSP samples and 5,000 Spanish controls obtained from the GR@ACE/DEGESCO
50 cohort, which had been previously genotyped with the same platform, using the same array and
51 in the same centre, thus minimizing potential batch effects^{1,2}. Called variants passing quality
52 control (QC) criteria defined by the manufacturer (Affymetrix) were kept for downstream analysis
53 (N=737,174).

54 After genotype calling, we removed samples with low genotype call rates (<97%) and high
55 heterozygosity (+3SD over mean heterozygosity). To identify mislabelled samples, we performed
56 a sex-check and removed those with discordant reported and genetic sex. We ran a principal
57 component (PC) analysis to identify and remove samples from non-European ancestries based on
58 the 1000 Genomes European population cluster (Extended Data Fig. 1). After sample QC, we
59 removed variants with high missingness (>5%), low frequency (MAF<0.01), differential
60 missingness between cases and controls ($p<10^{-5}$) or failing the Hardy Weinberg equilibrium test
61 ($p<10^{-6}$) in the control population. Samples with genotype call rates below 0.97 after variant QC
62 were also removed from the dataset. Genotype imputation was performed for 379 PSP samples
63 and 4,955 controls in the TOPMed Imputation Server (Michigan, USA) using the TOPMed
64 reference panel based on 587,195 genotyped variants passing QC. Prior to GWAS analysis, we
65 used the GENESIS R package³ to detect and remove potential cryptic relatedness within our
66 population. We first used a kinship filter of 0.45 to identify and remove sample pairs
67 corresponding to the same individual (N=11). After removing duplicates, we detected related
68 pairs by applying a kinship filter of 0.046875, which represents the theoretical kinship value
69 between third- and fourth-degree relatedness. We first removed the samples present in multiple
70 related pairs, and for the remaining unique relative pairs we applied the following criteria: 1) if

71 one sample was a PSP and the other one was a control we removed the control; 2) if one sample
72 was a histopathologically confirmed PSP the other sample was a clinical PSP, we kept the
73 histopathological PSP; 3) if both samples belonged to the same clinical group (control, clinical or
74 histopathological PSP), we kept the sample with the highest genotype call rate. For
75 PSP/DEGESCO Stage II samples, we performed the genotype calling including the controls and
76 cases passing QC from Stage I, and followed the same procedure described above. Additionally,
77 at the IBD filtering step we removed Stage II cases overlapping or related to cases from Stage I.
78 An overview of the complete QC process is provided (Extended Data Fig. 1, Extended Data Table
79 1, Supplementary Table 1).

80 2. Supplementary notes: DNA genotyping, QC & imputation in the Amsterdam Dementia 81 Cohort

82 We genotyped all individuals using the Illumina Global Screening Array and applied established
83 quality control methods⁴. We used high-quality genotypes in all individuals (individual call rate
84 >99%, variant call rate >99%), individuals with sex mismatches were excluded and departure
85 from Hardy-Weinberg equilibrium was considered significant at $p < 1 \times 10^{-6}$. Genotypes were then
86 lifted over to GRCh38 and prepared for imputation using provided scripts (HRC-1000G-check-
87 bim.pl) specifying TOPMED as reference panel. This script compares variant ID, strand, and
88 allele frequencies to the TOPMED reference panel (version r2, N=194,512 haplotypes from
89 N=97,256 individuals)⁵. Finally, all variants were submitted to the Michigan Imputation server
90 (<https://imputation.biodatacatalyst.nihbi.nih.gov/>). The server uses EAGLE (v2.4) to phase data
91 and Minimac4 to perform genotype imputation to the reference panel. After quality control and
92 genotype imputation of the genetic data, we kept only individuals of European ancestry (based on
93 1000Genomes clustering)⁶, and excluded individuals with a family relation (identity-by-descent
94 ≥ 0.2)⁷, leaving 59 clinically diagnosed PSP-RS cases and 1513 controls for analysis.

95

96 3. Supplementary notes: Detecting potential clinical heterogeneity using genetic 97 information.

98 We leveraged the *MAPT* H1 haplotype, a well-known genetic risk factor for PSP⁸, to make
99 inferences on misdiagnose rates in the Stage I clinical PSP dataset and its subgroups. We
100 considered the confirmed pathological PSP cases as a reference of a pure PSP group, and the
101 controls as a reference for a non-PSP population. Theoretically, for a mix of PSP and non-PSP
102 samples, the expected value of the *MAPT* H1 allele frequency (AF) should range between its
103 frequency in the pure/histopathological PSP ($H1_{\text{Freq}}=0.9199$), and the non-PSP ($H1_{\text{Freq}}=0.7024$)
104 populations in our study, and should be proportional to the ratio of PSP/non-PSP samples in the
105 mix. Thus, the *MAPT* H1 AF in the mixed sample would be equal to the *MAPT* H1 AF in the

106 histopathological PSP sample minus the difference in AF between the histopathological PSP and
107 control groups, (i.e. the total AF range between the pure PSP and non-PSP groups), multiplied by
108 the proportion of misdiagnosed samples in the mix:

$$109 \quad Fg = Fh - M \cdot (Fh - Fc)$$

110 *Fg*: MAPT H1 AF in the mixed group; *Fh*: MAPT H1 AF in the histopathological PSP group (AF=0.9199);
111 *M*: misdiagnose rate; *Fc*: MAPT H1 AF in the non PSP group (AF=0.7024); In the case of using logistic
112 regression-derived effect sizes instead of frequencies, the log(OR) of these groups can be introduced instead
113 of AF values.

114 Rearranging the terms in this equation, we can estimate the percentage of misclassified samples
115 in each group:

$$116 \quad \text{Misdiagnose (\%)} = 100 \cdot \frac{Fh - Fg}{Fh - Fc}$$

117 ($Fh - Fg$), difference in MAPT H1 AF between the histopathological PSP and mixed group.

118 ($Fh - Fc$), difference in MAPT H1 AF between the histopathological PSP and the general
119 population.

120 The same logic can be applied using linear effect size estimates (log odds ratios) derived from
121 logistic regressions instead of AFs. As population microstructure might affect the MAPT H1 allele
122 frequency in the Iberian population⁹, we also calculated the misclassification rates using the log
123 odds ratios obtained from fitting logistic regressions adjusted by the first four PCs in each
124 subgroup. For better comparability between all groups, we included the controls from the
125 discovery and validation setup (N=1,450) in all models. Additionally, because a proportion of the
126 misclassified patients are expected to have Parkinson's Disease (PD)¹⁰ and the MAPT H1 locus
127 is known to also be a risk factor for PD, we used a OR=1.00 (neutral) and a OR=1.32 (Parkinson's
128 Disease MAPT H1)¹¹ to calculate the effect size-based misdiagnosis estimations.

129 4. Supplementary notes: PRS calculation

130 We calculated the PRS of AD, PD, amyotrophic lateral sclerosis (ALS), PSP and all stroke (AS)
131 using reported genome-wide significant variants¹²⁻¹⁶. We constructed the PRSs selecting only
132 those variants with MAF>0.01 and imputation quality R2>0.3 in our dataset (Supplementary
133 Tables 3-7). A few considerations were taken in the AD PRS calculation: a) No APOE variants
134 were included, b) variants with MAF<0.01 were kept since these were previously validated in the
135 original publication and they had very good imputation quality in our dataset (R2>0.9), c) because
136 some controls of our study were included in the first stage of the AD GWAS, we used the
137 independent effects reported exclusively in the second stage. To calculate the different PRSs, we
138 designed a custom software (https://github.com/Pablo-GarGon/PRS_Generator) that added the

139 dosages of the risk alleles multiplied by their reported effect sizes (betas). Finally, we also
140 calculated the PRSs excluding the chromosome 17 to avoid pleiotropic effects caused by
141 including variants linked to the *MAPT* locus.

142 5. Supplementary notes: A Decreased burden of known PSP genetic risk factors was
143 detected in atypical PSP clinically diagnosed cases

144 Because, to our very best knowledge, this the first case-control PSP GWAS including clinical
145 cases diagnosed using the recently proposed MDS criteria for probable PSP¹⁷, and our *MAPT* H1
146 association results suggested the existence of phenotypic heterogeneity in our clinical PSP cohort
147 (Fig. 1), we decided to estimate genetically-based misclassification rates in the available Stage I
148 clinical PSP subjects (Supplementary Table 1). We classified these samples into four subgroups
149 based on two criteria: a) Samples with or without confirmed Richardson's Syndrome and b)
150 samples recruited by movement disorder units or memory clinics. The first classification allowed
151 us to examine the specificity of the MDS criteria for assigning a probable PSP diagnosis to non-
152 Richardson PSP (nR-PSP) patients, while the second one is useful to study potential ascertainment
153 bias caused by the recruitment in different neurology set-ups (movement disorders units and
154 memory clinics). As expected, the highest *MAPT* H1 frequency and effect size was found in the
155 histopathological PSP group, and the misclassification estimates were considerably larger for nR-
156 PSP than for PSP-RS samples (Fig. 1, Supplementary Table 1). Moreover, *MAPT* H1 frequencies
157 were significantly reduced in the nR-PSP patients compared to histopathological PSP ($\chi^2=6.29$;
158 $DF=1$; $p=0.01$), while no significant differences were observed for the PSP-RS patients in this
159 comparison ($\chi^2=0.16$; $DF=1$; $p=0.69$), supporting the presence of misclassified patients in the nR-
160 PSP group. Interestingly, PSP-RS samples had similar misclassification rates despite recruitment
161 centre, while nR-PSP samples recruited in memory clinics displayed lower misdiagnose rates than
162 those recruited by movement disorder units, although these differences were not significant
163 ($\chi^2=1.38$; $DF=1$; $p=0.24$).

164 To further complement these results, we generated the AD, PD, ALS, PSP and AS PRSs using
165 known disease susceptibility variants and compared the PRS values between our PSP subgroups
166 and controls (Supplementary Fig. 1, Supplementary Table 10). The AD PRS was significantly
167 increased in both nR-PSP subgroups: memory-clinic ($OR[95\%CI]=2.42[1.10-5.35]$; $p=0.03$) and
168 movement disorder unit-recruited ($OR[95\%CI]=2.64[1.13-6.22]$; $p=0.03$). Meanwhile, the PD
169 PRS was associated with disease status in both groups recruited by movement disorder units: PSP-
170 RS ($OR[95\%CI]=1.90[1.22-2.96]$; $p=0.004$) and nR-PSP ($OR[95\%CI]=1.72[1.00-2.93]$;
171 $p=0.05$). Finally, the PSP PRS was associated with disease status in all subgroups except
172 movement disorder recruited nR-PSP ($OR[95\%CI]=1.20[0.96-1.52]$; $p=0.12$), and no significant
173 associations were found between disease status and the ALS and AS PRSs. Because *MAPT* is a

174 pleiotropic risk factor for multiple neurodegenerative diseases, we repeated this analysis
175 excluding chromosome 17. After removing variants in the *MAPT* locus, an increase could still be
176 observed for the AD PRS in the nR-PSP subgroups and for the PD PRS in movement disorder
177 unit-recruited subgroups, but the statistical significance was lost (Supplementary Fig. 1,
178 Supplementary Table 7). Interestingly, the association with the PSP PRS was lost in both nR-PSP
179 subgroups, and even showed a non-significant opposite effect direction in movement disorder
180 recruited nR-PSP patients ($OR[95\%CI]=0.69[0.32-1.48]$; $p=0.34$). Furthermore, none of the
181 individual known PSP risk variants were associated with disease status in this subgroup either
182 (Fig. 1). Our results indicate that the movement disorder recruited nR-PSP patients lacked an
183 enrichment not only for *MAPT* H1 haplotypes but also for other known PSP risk factors.
184 Subgroup-specific association of each individual risk variant constituting the different PRSs is
185 available (Supplementary Tables 10-14).

186 Our data suggests two possible sources of contamination in our clinical dataset. On the one hand,
187 nR-PSP patients who do not evolve to PSP-RS, the hallmark PSP phenotype, are hard to
188 conclusively diagnose pre-mortem¹⁸. Thus, a weaker enrichment in known genetic PSP risk
189 factors and, conversely, an enrichment in risk factors for other neurodegenerative diseases can be
190 somewhat expected in the nR-PSP group. On the other hand, the observed enrichment in PD
191 genetic risk variants in the context of movement disorder units suggests the presence of a fraction
192 of misclassified parkinsonian PSP-like phenocopies in this clinical setting¹⁹. PSP-P patients
193 comprise up to a third of the total number of PSP patients and are particularly challenging to
194 accurately diagnose based solely on clinical symptoms²⁰. As these patients are characterized by
195 an initial phase with PD-compatible features, a higher chance of recruitment by movement
196 disorder units rather than by memory clinics might explain our results²¹. However, a smaller risk
197 effect of the *MAPT* H1 haplotype has also been described in PSP-P subjects compared to PSP-
198 RS²², which may also explain the decrease in *MAPT* H1 frequency we observed in the nR-PSP
199 groups. Overall, our findings suggest that there might be room for improvement of the specificity
200 of the MDS criteria by increasing the diagnostic accuracy of atypical PSP phenotypes, specially
201 PSP-P. Various biomarkers are under development to further help discriminate PSP from PSP-
202 like patients: Real time quaking induced conversion (RT-QuIC) techniques to discard α -synuclein
203 parkinsonian pathology, or for confirmation of 4R tauopathy are underway as CSF biomarkers,
204 and MRI parameters like the magnetic resonance parkinsonism index are showing promising
205 results in predicting PSP-RS conversion in early PSP-P subjects²¹.

206

207 6. Supplementary notes: Sensitivity analysis of PSP sentinels

208 Based on our previous results, we decided to run a sensitivity analysis for known PSP risk
209 variants, based on GWS hits by Sanchez-Contreras *et al*¹⁵, stratifying our Stage I cases in three
210 groups of decreasing PSP certainty: histopathological PSP, PSP-RS and nR-PSP. To improve
211 comparability of the calculated estimates, we included the complete set of Stage I controls
212 (N=1450) in all three strata. For all tested variants, we found consistent effect directions between
213 the histopathological and PSP-RS groups (Supplementary Fig. 2). The effects in the nR-PSP
214 group were smaller or even had opposite directions for the tested variants, except for rs242557,
215 the variant tagging the *MAPT* H1c haplotype suggesting it may similarly impact genetic
216 susceptibility to PSP-like syndromes. Finally, we decided to repeat the meta-analysis of the Stage
217 I discovery and validation datasets, this time keeping only the PSP-RS samples in the validation
218 set. After removing nR-PSP samples, the effect sizes were stronger in all PSP susceptibility
219 variants (Supplementary Table 9), showcasing the benefits of removing these samples with lower
220 diagnostic accuracy from our GWAS analysis.

221 7. Supplementary notes: Enrichment analysis

222 Enrichment analysis was performed using WebGestalt (<https://www.webgestalt.org/>) over-
223 representation analysis²³. Briefly, this method evaluates the fraction of genes which constitute the
224 different gene sets that are present in the list of genes introduced, testing the significance using a
225 2x2 contingency table²⁴. We introduced the list of replicated genes (*MAPT*, *MOBP*, *EIF2AK3*,
226 *STX6*, *DUSP10*, *SLCO1A2*) and our novel *NFASC* gene and ran the over representation analysis
227 algorithm, enriching specifically for non-redundant gene ontology terms and using the “Affy
228 Axiom Biobank1” option as the reference gene set. Although none of the top enriched terms were
229 significant after multiple test correction (Extended Data Table 6), the term GO:0097386,
230 representing glial cell projection was borderline significant (FDR adjusted p-value = 0.06) and
231 several of the top ranked terms were related to axon ensheathment and myelination, pinpointing
232 these processes as those more represented by our list of genes.

233 8. Supplementary notes: Gene co-expression network analysis.

234 Gene co-expression networks were generated using GeneFriends software
235 (<https://www.genefriends.org/>)²⁵, including all interesting confirmed loci and candidates (*MAPT*,
236 *MOBP*, *EIF2AK3*, *STX6*, *DUSP10*, *SLCO1A2*, *NFASC* and *CNTN2*) using the Sequence Read
237 Archive (SRA) human tissue co-expression map as a reference, which comprises 20 tissues from
238 46080 RNA-seq samples. Briefly, this method first estimates correlations between each pair of
239 genes, then uses co-expression associations to construct the co-expression network, and finally
240 modules, i.e. groups of co-expressed genes are identified using clustering techniques, allowing to
241 group those genes with similar expression patterns across the different samples. The Pearson
242 correlation threshold was adjusted to 0.6 to allow us to identify the top genes co-expressed with

243 our main cluster, consisting of *MOBP*, *SLCO1A2* and *NFASC*. We found that *MARK1*, as well as
244 three additional lncRNAs and one pseudogene, were strongly co-expressed ($R > 0.6$) with this gene
245 cluster (Extended Data Fig. 5).

246 **Supplementary Table Legends**

247 **Supplementary Table 1.** QC summary of contributed samples by collaborating center and stage.

248 **Supplementary Table 2.** Age linear regressions results for the complete set of controls used in
249 PSP/DEGESCO Stage I (N=1450). Genomic coordinates correspond to the GRCh38 assembly.

250 **Supplementary Table 3.** Reported summary statistics of the variants used to construct the
251 Alzheimer's Disease (AD) PRS. We used exclusively effects from stage 2 of the Bellenguez *et al.*
252 study¹². Genomic coordinates correspond to the GRCh38 assembly.

253 **Supplementary Table 4.** List of the variants used to construct the amyotrophic lateral sclerosis
254 (ALS) PRS. These summary statistics correspond to the European ancestry GWAS from the
255 original study. Genomic coordinates correspond to the GRCh38 assembly.

256 **Supplementary Table 5.** List of the variants used to construct the all stroke (AS) PRS. Genomic
257 coordinates correspond to the GRCh38 assembly.

258 **Supplementary Table 6.** List of the variants used to construct the Parkinson's Disease (PD) PRS.
259 Genomic coordinates correspond to the GRCh38 assembly.

260 **Supplementary Table 7.** List of the variants used to construct the progressive supranuclear palsy
261 (PSP) PRS. Genomic coordinates correspond to the GRCh38 assembly.

262 **Supplementary Table 8.** Estimation of the percentage of misdiagnosed clinical PSP samples
263 classified by those with/without Richardson's Syndrome and by recruitment centre (Movement
264 Disorder Unit/Memory Clinic) in PSP/DEGESCO Stage I. ORs were calculated fitting logistic
265 regressions adjusted by the first four PCs and using the same set of controls in all groups
266 (N=1450). Percentage of misdiagnosed samples was calculated based on MAPT H1 haplotype
267 frequencies and log odds ratios. Age data represents age at death, age at onset of motor symptoms
268 and age at last visit for histopathological PSP, clinical PSP and controls, respectively.

269 **Supplementary Table 9.** Sensitivity analysis results for the replicated variants. We showcase the
270 results obtained in the previous study and the PSP/DEGESCO meta-analysis results including and
271 excluding nR-PSP samples. Models were adjusted by the first four PCs. EA, effect allele; OA,
272 other allele; OR, odds ratio; P, p-value.

273 **Supplementary Table 10.** Association in case control logistic regression of neurodegenerative
274 disease PRSs with PSP in the different clinical subgroups. PRSs were calculated using all reported
275 variants, and removing variants in chr17 to account for MAPT pleiotropy. Associations are
276 reported in the following format: OR [95% CI]; P. OR, odds ratio; PRS, polygenic risk score; AD,
277 Alzheimer's Disease; PD, Parkinson's Disease; ALS, amyotrophic lateral sclerosis; PSP,
278 progressive supranuclear palsy; AS, all stroke.

279 **Supplementary Table 11.** Association in case control logistic regression of the individual
280 variants used to construct the Alzheimer's Disease (AD) PRS with the different PSP subgroups in
281 PSP/DEGESCO Stage I.

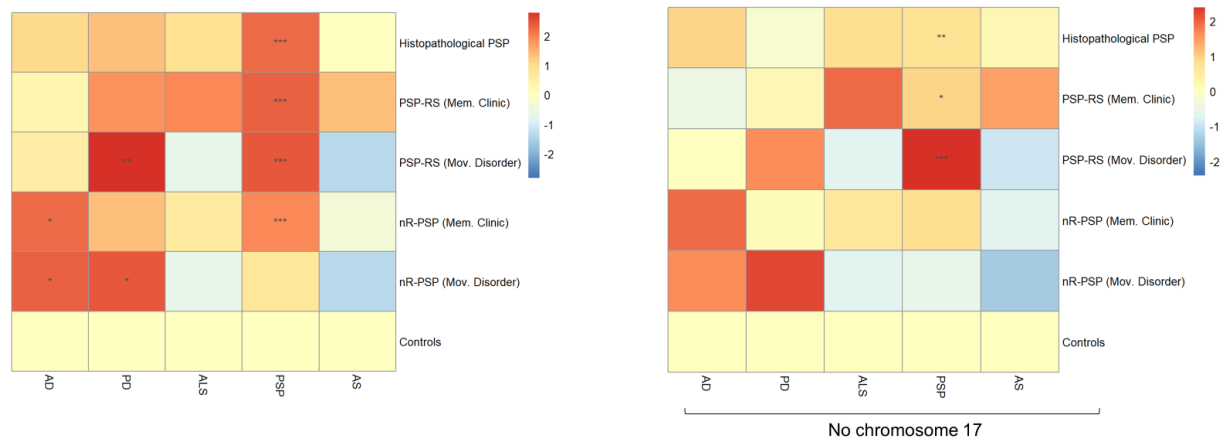
282 **Supplementary Table 12.** Association in case control logistic regression of the individual
283 variants used to construct the amyotrophic lateral sclerosis (ALS) PRS with the diferent PSP
284 subgroups in PSP/DEGESCO Stage I.

285 **Supplementary Table 13.** Association in case control logistic regression of the individual
286 variants used to construct the all stroke (AS) PRS with the diferent PSP subgroups in
287 PSP/DEGESCO Stage I.

288 **Supplementary Table 14.** Association in case control logistic regression of the individual
289 variants used to construct the Parkinson's disease (PD) PRS with the diferent PSP subgroups in
290 PSP/DEGESCO Stage I.

291 **Supplementary Table 15.** Association in case control logistic regression of the individual
292 variants used to construct the progressive supranuclear palsy (PSP) PRS with the diferent PSP
293 subgroups in PSP/DEGESCO Stage I.

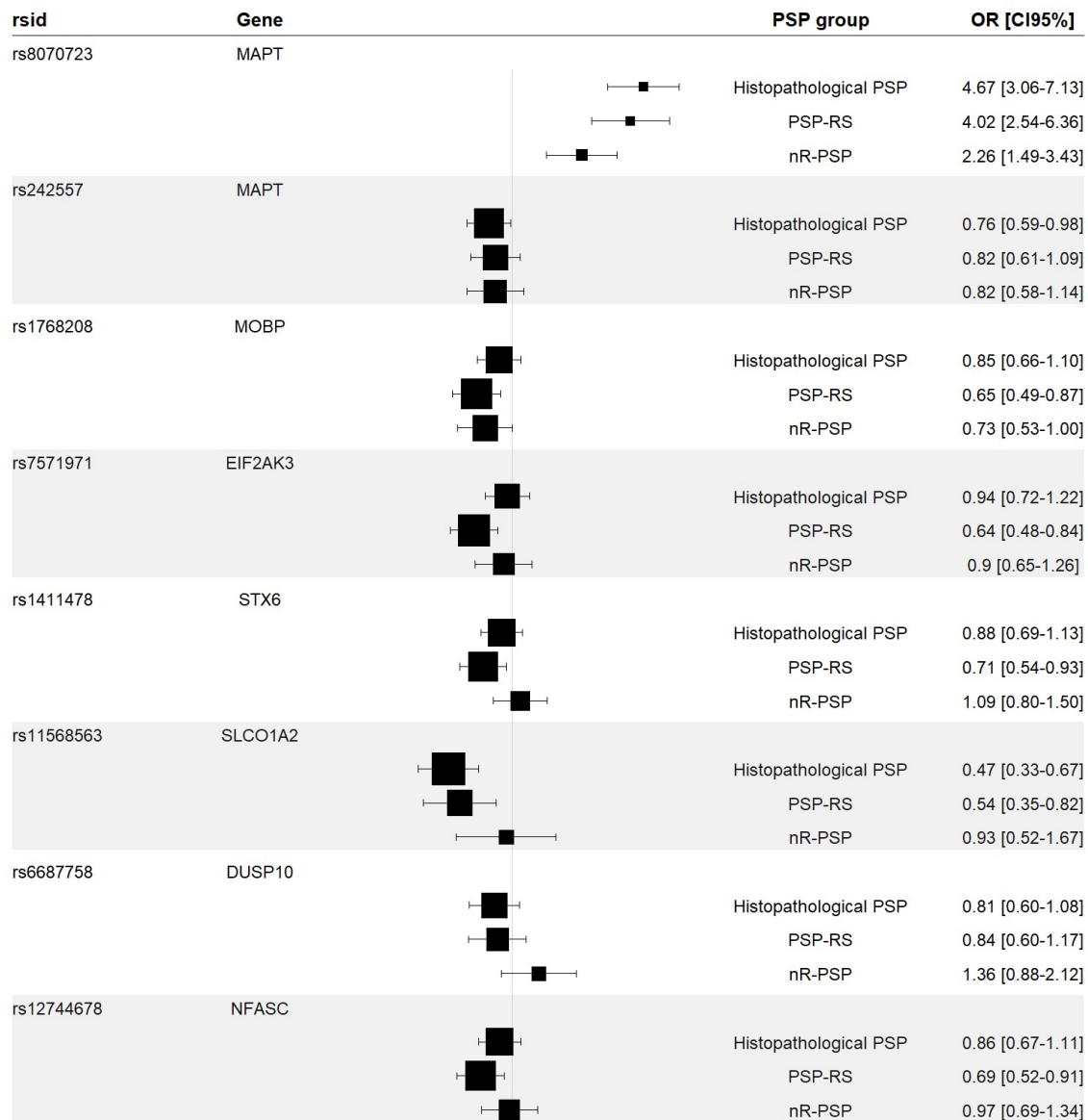
294 **Supplementary Figures**



295

296 **Supplementary Fig. 1.** Mean AD, PD, ALS, PSP and AS PRS in each PSP subgroup including
 297 and excluding variants in chr17. Values were zero-centred with respect to the control group to
 298 better reflect the increase/decrease in mean PRS in the PSP groups with respect to the controls.
 299 OR, odds ratio; PRS, polygenic risk score; AD, Alzheimer's Disease; PD, Parkinson's Disease;
 300 ALS, amyotrophic lateral sclerosis; PSP, progressive supranuclear palsy; AS, all stroke; ***
 301 p<0.001; ** p<0.01; * p<0.05.

302



303

304

305 **Supplementary Fig. 2. Forest plot displaying effect sizes of SNPs in the histopathological**
 306 **and clinical groups (PSP-RS or nR-PSP) of PSP/DEGESCO Stage I.** Models were adjusted
 307 by PCs1-4 and included the same set of controls (N=1450) to improve comparability. OR, odds
 308 ratio; CI95%, 95% confidence interval.

309

310

311

312

313

314

315

316

317 **Supplementary References**

- 318 1. de Rojas, I. *et al.* Common variants in Alzheimer's disease and risk stratification by
319 polygenic risk scores. *Nature Communications* (2021) doi:10.1038/s41467-021-22491-8.
- 320 2. Moreno-Grau, S. *et al.* Genome-wide association analysis of dementia and its clinical
321 endophenotypes reveal novel loci associated with Alzheimer's disease and three causality
322 networks: The GR@ACE project. *Alzheimer's and Dementia* (2019)
323 doi:10.1016/j.jalz.2019.06.4950.
- 324 3. Gogarten, S. M. *et al.* Genetic association testing using the GENESIS R/Bioconductor
325 package. *Bioinformatics* (2019) doi:10.1093/bioinformatics/btz567.
- 326 4. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat Genet* **48**,
327 1284–1287 (2016).
- 328 5. Taliun, D. *et al.* Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program.
329 *Nature* **590**, 290–299 (2021).
- 330 6. The 1000 Genomes Project Consortium *et al.* A global reference for human genetic variation.
331 *Nature* **526**, 68–74 (2015).
- 332 7. Anderson, C. A. *et al.* Europe PMC Funders Group Data quality control in genetic case-
333 control association studies. *Nature Protocols* (2011) doi:10.1038/nprot.2010.116.Data.
- 334 8. Baker, M. *et al.* Association of an Extended Haplotype in the Tau Gene with Progressive
335 Supranuclear Palsy. *Human Molecular Genetics* **8**, 711–715 (1999).
- 336 9. Gayán, J. *et al.* Genetic Structure of the Spanish Population. *BMC Genomics* **11**, 326 (2010).
- 337 10. Höglinger, G. U. *et al.* Identification of common variants influencing risk of the
338 tauopathy progressive supranuclear palsy. in *Nature Genetics* (2011). doi:10.1038/ng.859.
- 339 11. Simón-Sánchez, J. *et al.* Genome-wide association study reveals genetic risk underlying
340 Parkinson's disease. *Nat Genet* **41**, 1308–1312 (2009).
- 341 12. Bellenguez, C. *et al.* New insights into the genetic etiology of Alzheimer's disease and
342 related dementias. *Nat Genet* **54**, 412–436 (2022).

- 343 13. Nalls, M. A. *et al.* Identification of novel risk loci, causal insights, and heritable risk for
344 Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet Neurol* **18**,
345 1091–1102 (2019).
- 346 14. van Rheenen, W. *et al.* Common and rare variant association analyses in amyotrophic
347 lateral sclerosis identify 15 risk loci with distinct genetic architectures and neuron-specific
348 biology. *Nat Genet* **53**, 1636–1648 (2021).
- 349 15. Sanchez-Contreras, M. Y. *et al.* Replication of progressive supranuclear palsy genome-
350 wide association study identifies SLCO1A2 and DUSP10 as new susceptibility loci. *Mol*
351 *Neurodegeneration* **13**, 37 (2018).
- 352 16. Mishra, A. *et al.* Stroke genetics informs drug discovery and risk prediction across
353 ancestries. *Nature* **611**, 115–123 (2022).
- 354 17. Höglinger, G. U. *et al.* Clinical diagnosis of progressive supranuclear palsy: The
355 movement disorder society criteria: MDS Clinical Diagnostic Criteria for PSP. *Mov Disord.*
356 **32**, 853–864 (2017).
- 357 18. Respondek, G. *et al.* Accuracy of the national institute for neurological disorders and
358 stroke/society for progressive supranuclear palsy and neuroprotection and natural history in
359 Parkinson plus syndromes criteria for the diagnosis of progressive supranuclear palsy.
360 *Movement Disorders* **28**, 504–509 (2013).
- 361 19. Boxer, A. L. *et al.* Advances in progressive supranuclear palsy: new diagnostic criteria,
362 biomarkers, and therapeutic approaches. *The Lancet Neurology* **16**, 552–563 (2017).
- 363 20. Alster, P., Madetko, N., Koziorowski, D. & Friedman, A. Progressive Supranuclear
364 Palsy—Parkinsonism Predominant (PSP-P)—A Clinical Challenge at the Boundaries of PSP
365 and Parkinson's Disease (PD). *Front. Neurol.* **11**, 180 (2020).
- 366 21. Coughlin, D. G. & Litvan, I. Progressive supranuclear palsy: Advances in diagnosis and
367 management. *Parkinsonism Relat Disord* **73**, 105–116 (2020).
- 368 22. Williams, D. R. *et al.* Characteristics of two distinct clinical phenotypes in
369 pathologically proven progressive supranuclear palsy: Richardson's syndrome and PSP-
370 parkinsonism. *Brain* **128**, 1247–1258 (2005).

- 371 23. Wang, J., Vasaikar, S., Shi, Z., Greer, M. & Zhang, B. WebGestalt 2017: a more
372 comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit.
373 *Nucleic Acids Research* **45**, W130–W137 (2017).
- 374 24. Khatri, P., Sirota, M. & Butte, A. J. Ten Years of Pathway Analysis: Current
375 Approaches and Outstanding Challenges. *PLoS Comput Biol* **8**, e1002375 (2012).
- 376 25. Raina, P. *et al.* GeneFriends: gene co-expression databases and tools for humans and
377 model organisms. *Nucleic Acids Research* **51**, D145–D158 (2023).

378

379

380