Title

Genome-wide survey of copy number variants finds MAPT duplications in progressive supranuclear palsy

Journal

Movement Disorders, 34(7)

ISSN

0885-3185

Authors

Chen, Zhongbo
Chen, Jason A
Shatunov, Aleksey
et al.

Publication Date

2019-07-01

DOI

10.1002/mds.27702

Peer reviewed

RESEARCH ARTICLE

Genome-Wide Survey of Copy Number Variants Finds MAPT Duplications in Progressive Supranuclear Palsy

Zhongbo Chen, MA, BMBCh, MRCP,[1] Jason A. Chen, PhD,[2] Aleksey Shatunov, PhD,[1] Ashley R. Jones, PhD,[1] Stephanie N. Kravitz, BS,[3] Alden Y. Huang, PhD,[2] Lauren Lawrence, MD,[3] Jennifer K. Lowe, PhD,[3] Cathryn M. Lewis, PhD,[4] Christine A. M. Payan, MD,[5] Wolfgang Lieb, PhD,[6] Andre Franke, PhD,[7] Panagiotis Deloukas, PhD,[8] Philippe Amouyel, PhD,[9] Christophe Tzourio, MD, PhD,[10] Jean-François Dartigues, MD, PhD,[10] NNIPPS and BBBIPPS Study Groups, Albert Ludolph, PhD,[11] Gilbert Bensimon, MD, PhD,[5] P. Nigel Leigh, PhD, FRCP, FAAN, FMedSci,[12] Jeff M. Bronstein, MD, PhD,[13] Giovanni Coppola, MD,[2,3,14] Daniel H. Geschwind, MD, PhD,[3,4,15*] and Ammar Al-Chalabi, PhD, FRCP, DipStat,[1*]

[1]*Department of Basic and Clinical Neuroscience, Maurice Wohl Clinical Neuroscience Institute, King's College London, London, UK*
[2]*Interdepartmental Program in Bioinformatics, David Geffen School of Medicine, University of California, Los Angeles, California, USA*
[3]*Program in Neurogenetics, Department of Neurology, David Geffen School of Medicine, University of California, Los Angeles, California, USA*
[4]*Social, Genetic and Developmental Psychiatry Centre, and Department of Medical and Molecular Genetics, King's College London, UK*
[5]*Département de Pharmacologie Clinique, Hôpital de la Pitié-Salpétrière, Assistance Publique Hôpitaux de Paris, Paris; Pharmacologie, Universités Paris-Sorbonne, UPMC Paris 06, Paris, France*
[6]*Institute of Epidemiology and Biobank Popgen, Christian Albrechts Universitat zu Kiel, Kiel, Germany*
[7]*Institute of Clinical Molecular Biology, Christian Albrechts Universitat zu Kiel, Kiel, Germany*
[8]*William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, Charterhouse Square, London, UK*
[9]*Univ. Lille, Inserm, CHU Lille, Institut Pasteur de Lille, U1167 - RID-AGE - Risk Factor and Molecular Determinants of Aging Diseases, Labex-Distalz, Lille, France*
[10]*University of Bordeaux, INSERM, Bordeaux Population Health Research Centre, UMR-1219, CHU Bordeaux, France*
[11]*Department of Neurology, University of Ulm, Oberer Eselsberg, Ulm, Germany*
[12]*Trafford Centre for Biomedical Research, Brighton and Sussex Medical School, University of Sussex, Falmer, Brighton, UK*
[13]*Program in Movement Disorders, Department of Neurology, David Geffen School of Medicine, University of California, Los Angeles, California, USA*
[14]*Center for Neurobehavioral Genetics, Semel Institute, David Geffen School of Medicine, University of California, Los Angeles, California, USA*
[15]*Department of Human Genetics, University of California, Los Angeles, California, USA*

*__Corresponding author:__ Professor Ammar Al-Chalabi, Maurice Wohl Clinical Neuroscience Institute, King's College London, London, SE5 9NU, UK; E-mail: ammar.al-chalabi@kcl.ac.uk; or Professor Daniel Geschwind, UCLA Neurogenetics Program, 2309 Gonda Bldg., 695 Charles E. Young Dr. South, Los Angeles, CA 90095-1761; E-mail: dhg@mednet.ucla.edu

Funding agencies: The NNIPPS study was supported by the European Union 5th Framework Programme (QLG1-CT-2000-01262); the French Health Ministry, Programme Hospitalier de Recherche Clinique (AOM97073, AOM01125); Sanofi-Aventis affiliates in the UK, France, and Germany provided an unconditional research grant and drug supply throughout the study; the UK National Institute for Health Research (NIHR) Maudsley Biomedical Research Centre at South London and Maudsley NHS Foundation Trust (SLaM) and King's College London, UK. Three academic institutions (Institute of Psychiatry, Psychology and Neuroscience, King's College London; Assistance Publique-Hôpitaux de Paris, and University of Ulm) were sponsors of the NNIPPS study in each country and jointly own the data. The BBBIPPS study was supported by the French Health Ministry, Programme Hospitalier de Recherche Clinique (AOM04035). The Assistance Publique - Hôpitaux de Paris (France) was the sponsor of the study. The protocol and amendments were reviewed and approved by the Comité de Protection des Personnes of Pitié-Salpêtrière Hospital (France). This French control work was supported by the National Foundation for Alzheimer's Disease and Related Disorders (Fondation Plan Alzheimer), Inserm, the Institut Pasteur de Lille, the Centre National de Génotypage and the LABEX (Excellence Laboratory, Program Investment for the Future), DISTALZ (Development of Innovative Strategies for a Transdisciplinary Approach to Alzheimer's Disease). All SNP genotyping was performed at the Neuroscience Genomics Core at UCLA and funded by a generous gift from the Chen Family Foundation (to J.M.B.). UCLA genetics investigators (D.G., G.C., J.L., J.A.C.) were sponsored by the Tau Consortium and the Chen Family Foundation. Z.C. was supported by funding from the National Institute for Health Research (NIHR) Academic Clinical Fellowship scheme.

Relevant conflicts of interests/financial disclosures: Nothing to report.

Zhongbo Chen, Jason A. Chen, and Aleksey Shatunov contributed equally to this article.

Received: 14 October 2018; **Revised:** 1 February 2019; **Accepted:** 26 February 2019

Published online 00 Month 2019 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/mds.27702

ABSTRACT: Background: Progressive supranuclear palsy is a neurodegenerative tauopathy manifesting clinically as a progressive akinetic-rigid syndrome. In this study, we sought to identify genetic variants influencing PSP susceptibility through a genome-wide association analysis of a cohort of well-characterized patients who had participated in the Neuroprotection and Natural History in Parkinson Plus Syndromes and Blood Brain Barrier in Parkinson Plus Syndromes studies.

Methods: We genotyped single-nucleotide polymorphisms in 283 PSP cases from the United Kingdom, Germany, and France and compared these with genotypes from 4472 controls. Copy number variants were identified from genotyping data.

Results: We observed associations on chromosome 17 within or close to the *MAPT* gene and explored the genetic architecture at this locus. We confirmed the previously reported association of rs1768208 in the *MOBP* gene ($P = 3.29 \times 10^{-13}$) and rs1411478 in *STX6* ($P = 3.45 \times 10^{-10}$). The population-attributable risk from the *MAPT*, *MOBP*, and *STX6* single-nucleotide polymorphisms was found to be 0.37, 0.26, and 0.08, respectively. In addition, we found 2 instances of copy number variants spanning the *MAPT* gene in patients with PSP. These copy number variants include tau but few other genes within the chromosome 17 haplotype region, providing additional support for the direct pathogenicity of *MAPT* in PSP.

Conclusions: Clinicians should also be aware of *MAPT* duplication as a possible genetic cause of PSP, especially in patients presenting with young age at onset. © 2019 International Parkinson and Movement Disorder Society

Key Words: progressive supranuclear palsy; genome-wide association study; copy number variation

Progressive supranuclear palsy (PSP) is an atypical parkinsonian disorder characterized by vertical supranuclear gaze palsy, postural instability, progressive axial rigidity, and mild cognitive impairment,[1] with a community prevalence of up to 6.5 per 100,000 persons.[2] Variation in the gene encoding tau, *MAPT*, has been established as a genetic risk factor for PSP.[3] Genome-wide association studies (GWASs) have also identified variants in *MOBP*, *STX6*, and other genes as significant risk factors.[4-6] Although further studies have found trends toward association,[7] they have been underpowered, and these associations have not been definitively reproduced in an independent sample.

We aimed to identify common genetic variants influencing the risk of PSP through genome-wide analysis of samples from the Neuroprotection and Natural History in Parkinson Plus Syndromes (NNIPPS)[8] and the Blood Brain Barrier in Parkinson Plus Syndromes (BBBIPPS) studies. NNIPPS and BBBIPPS recruited a well-characterized cohort of patients with PSP. This cohort was also included in our previously reported joint-analysis GWAS of PSP.[5] Here, we performed genome-wide association on this cohort alone to confirm genetic loci implicated in PSP and validate this clinical PSP cohort, comparing with previous GWASs. Then, we used array genotyping data to identify rare copy number variants (CNVs) among PSP patients. Recurrent rare CNVs present in patients with PSP but not in controls were identified and validated.

Materials and Methods

Study Participants

The NNIPPS (ClinicalTrials.gov trial registration: NCT00211224) and BBBIPPS (French National Health Department Registry Number: DGS N°DGS2006/0524) studies enrolled patients aged between 30 and 80 with an akinetic-rigid syndrome, using validated diagnostic criteria to distinguish PSP and multiple system atrophy (MSA) that have been reported in detail previously.[8] In the previously reported study, 112 patients who had died during the trial had postmortem examination. There was good predictive validity of diagnostic criteria against neuropathological findings for PSP (0.95 [95% confidence interval, 0.88–0.98] and 0.84 [95% confidence interval, 0.77–0.87] sensitivity and specificity, respectively). All cases were of European ancestry, recruited from the UK, Germany, and France. DNA was extracted after written informed consent. BBBIPPS used the same NNIPPS diagnostic criteria and was conducted in the same study centers in France. These patients were previously included as a subset of a GWAS in PSP that we had published.[5]

In the United Kingdom, control genotype data from the Wellcome Trust Case Control Consortium 1958 Birth Cohort (WTCCC_1958) was used. In France and Germany, genotyped neurologically normal controls were population-matched. German control genotypes were obtained from the PopGen biorepository.[9] French controls were selected as part of the population-based Three-City (3C) Study in France.[10] Power calculations were performed using Genetic Power Calculator.[11]

Genotyping and Quality Control

Single-nucleotide polymorphisms (SNPs) were genotyped using platforms listed in Supplementary Table 1 and in the UCLA Neuroscience Genomics Core (http://www.semel.ucla.edu/ungc). Alleles were called using GenomeStudio (Illumina Inc., CA).

Quality control filters[12] were applied using PLINK v1.07 (pngu.mgh.harvard.edu/purcell/plink) and PLINK2 (www.cog-genomics.org/plink2).[13] We excluded SNPs with minor allele frequency (MAF) < 0.03, those with nonrandom missingness $P < 10^{-5}$ and abnormal autosomal heterozygosity. SNPs showing departure from Hardy-Weinberg equilibrium ($P < 10^{-6}$) in controls and cryptic relatedness of $\hat{\pi} \geq 0.125$ were also excluded.

To correct for population structure, principal components (PCs) of ancestry, derived from a sample of SNPs in linkage equilibrium (LD) were included in the model as previously described.[14] The number of PC axes included in the analysis was estimated using Tracy-Widom statistics.

Imputation

We used *Mach and Minimac*[15] to conduct the imputation as described in the "Enhancing Neuroimaging Genetics through Meta-Analysis" consortium protocol (http://enigma.ini.usc.edu/wpcontent/uploads/2012/07/ENIGMA2_1KGP_cookbook_v3.pdf) to predict alleles not genotyped based on LD. We imputed 9,065,536 SNPs. Association analysis was done in *mach2dat*, removing genotypes with $r^2 < 0.3$. Imputed SNPs were used only to match SNPs reported in other studies as part of the meta-analysis.

Statistical Association Analyses

Association analysis was performed using a logistic regression model, conditional analyses, and haplotype association, implemented in PLINK. The threshold for significance was $P < 5 \times 10^{-8}$ and for follow-up was $P < 10^{-5}$. Penetrance and population-attributable risk were estimated by standard methods.[16]

The allele-by-allele case-only and case-control epistatic interactions were assessed in PLINK v1.07 to test the interaction between pairwise SNPs associated with PSP, taking into account the allelic dosage of each SNP. We also calculated the proportion of genetic variance accounted for at each locus: genetic variance = 2ln (OR) × MAF × (1 - MAF).

Combined Analysis

We identified a previous GWAS of PSP[6] encompassing 1051 cases and 3560 controls and 1 targeted genotype association analysis[17] using 127 cases and 199 controls. From each of the studies, we extracted information regarding the associated SNPs compared with SNPs identified in our GWAS. For a SNP not genotyped, we identified the matching imputed SNP. The mean r^2 of the imputed SNPs used in the meta-analysis was 0.95.

We performed a meta-analysis using PLINK to combine logistic regression analysis results of each study using both a fixed- and random-effects model.

CNV Analysis

CNVs were identified from genotyping data of patients with PSP and controls. Because the raw data were required for CNV calling, only a subset of control patients (n = 1084) from the WTCCC_1958 was suitable for CNV analysis. Initially, sample reclustering was performed on each sample using Illumina GenomeStudio for each array batch. Samples with low call rates (<98%) were excluded. To compare CNVs between PSP cases and controls from the WTCCC_1958, only SNPs in common between the 2 array platforms were used for subsequent analysis (689,077 SNPs in total). PennCNV[19] was used to call CNVs, using custom PFB and GC model files and with genomic wave adjustment. Adjacent CNV calls were merged if their separation spanned <20% of their combined length. CNVs overlapping (>50%) immunoglobulin, telomeric, and centromeric regions; called on <10 array SNPs; spanning <50,000 base pairs; and having a confidence score <10 were filtered. Subjects with Log R ratio (LRR) standard deviation >0.285, BAF Drift <0.01, waviness factor <0.05, total called CNVs >100, and maximum combined CNV size <10,000,000 base pairs were removed from further analysis. CNVs highlighted by downstream analyses were validated by manual examination of the signal intensity.

Rare CNVs (frequency <1%) were considered for downstream analysis. CNV burden and association testing was performed using PLINK v1.07 and the Bedtools package.[20] Empirical *P* values were calculated using the maxT test with 50,000 permutations, with statistical significance defined at corrected $P < 0.05$.

Quantitative Polymerase Chain Reaction Validation of CNVs

Detected CNVs of interest were validated using quantitative polymerase chain reaction (qPCR) as previously outlined.[21,22] Specifically, primers were selected to amplify a genomic region within the CNV of interest and in a genomic region outside the CNV of interest that did not contain a polymorphic CNV. For the CNV region, primers within the selected region were designed with PrimerZ[23] (available at genepipe.ncgm.sinica.edu.tw). For the duplication at *MAPT*, 2 primer pairs were designed. The first had forward primer sequence CTACC TGATCCCCCTTCCTC and reverse primer sequence TCTCTGTTCCCCATCACTCC and amplified a 106-bp sequence between chr17:43,985,893 and 43,985,998 (GRCh37/hg19). The second had forward primer sequence CCACGTTCTCCTCCACATTT and reverse primer sequence CCTGCTCCAAACCCTGATAA and amplified a 102-bp sequence between chr17:43,986,157 and 43,986,258 (GRCh37/hg19). For the control region, a primer within the housekeeping gene RNase P (gene symbol *RPPH1*) was used.

The qPCR of 5 ng of genomic DNA of patients suspected of harboring the CNV of interest and 5 ng of control sample (using both the CNV primer and control primer) was performed using a SensiFAST SYBR No-ROX kit (Bioline, Taunton, MA). The control sample was a mixture of genomic DNA from 10 healthy people collected from a separate study, thereby minimizing the possibility of an undetected copy number variant within the region of interest and increasing the robustness of the assay. Note that this step was not used to assess the frequency of the CNV of interest in a control population; it is assumed that copy number variation at the amplicon site is rare, and that in total, the mixture of samples will not deviate from normal copy number. Primer efficiency was checked using the standard curve method and shown to be near 100%; therefore, in subsequent calculations the primer efficiency (E) was assumed to be maximal (E = 2). A quantitative estimate of copy number was then defined as $2^{-\Delta\Delta Ct}$, where $\Delta\Delta C_t$ is the relative difference in number of cycles to threshold of the CNV primer in the respective sample, relative to the control primer and control genomic DNA sample. The qPCR validation was performed in quadruplicate; the copy number was determined as the mean copy number determined across replicates, rounded to the nearest integer.

Results

In total, there were 158 patients from France, 50 from Germany, and 75 from the United Kingdom, with a median age of 68.3 years (range, 40–81 years). This gave >80% power to detect a variant with MAF of 0.4, conferring a relative risk of 1.3 under an additive model at $P < 0.05$ before correction for multiple testing.

The total genotyping rate was 0.998. After stringent quality control and exclusion of SNPs not overlapping on a genotyping platform, 284,674 SNPs remained. Six principal components were used to correct for population substructure.

Genome-Wide Association Analysis

Multiple SNPs on chromosome 17 surpassed genome-wide significance (Fig. 1a), and statistics behaved as expected (λ_{GC} = 1.022; Fig. 1b). A plot of the genetic architecture surrounding the associated region revealed strong linkage disequilibrium, with the most significant SNP in the association analysis (rs12185268, $P = 2.01 \times 10^{-16}$) in the *MAPT* gene (Fig. 1c and Supplementary Table 2). Conditional analysis in both the genotyped and imputed data confirmed that there was no other independent association in the single-study analysis (Supplementary Fig. 1).

Meta-analysis showed that additional SNPs in *STX6* and *MOBP* reached genome-wide significance (Table 1). SNP rs1768208 in the *MOBP* gene had an odds ratio (OR) of 1.52 ($P = 2.65 \times 10^{-5}$) in the NNIPPS data set. This gave a combined OR of 1.42, $P = 3.29 \times 10^{-13}$, making it the second most significant SNP in the meta-analysis (Fig. 2a). We also confirmed that SNP rs1411478 (*STX6*) was associated with PSP ($P = 3.45 \times 10^{-10}$) after the meta-analysis. SNP rs242557 (*MAPT*), identified in 3 studies, had a consistent effect size, with an OR of 1.91, $P = 1.58 \times 10^{-22}$ (Fig. 2b).

Gene-by-Gene Interactions

No other SNP found to be significantly associated in the meta-analysis showed epistatic interaction with the most significant SNP, rs12185268, *MAPT* (Supplementary Table 3).

Population-attributable risk and penetrance of the SNPs reaching genome-wide significance are shown in Table 2. There was haplotypic association of *MAPT* SNPs rs242557 and rs16940742 (omnibus $P = 9.43 \times 10^{-15}$) and rs242557 and rs2435200 (omnibus $P = 7.61 \times 10^{-16}$).

CNV Association

A total of 4866 CNVs in 281 PSP patients and 1084 controls passed quality control. Of these, 2769 CNVs including 1205 deletions and 1564 duplications were present at less than 1% frequency (considered rare). We did not detect an increased burden of rare CNVs in PSP patients after adjusting for multiple comparisons, even when stratifying by CNV size (Supplementary Table 4). A trend toward enrichment of 200- to 500-kb duplications was detected in PSP (mean of 0.34 segments in PSP patients versus mean of 0.26 segments in controls, unadjusted $P = 0.02$). To identify rare CNVs that might cause PSP, recurrent genic CNVs (found in 2 or more PSP cases but not in WTCCC controls) were identified. We found 4 recurrent duplications, located within 2q37.1 (spanning *ALPP, ECEL1P2*, and *ALPPL2*), 4q31.21 (spanning *RNF150* and *ZNF330*), 16p12.2 (spanning *METTL9, IGSF6*, and *OTOA*), and 17q21.31 (including a portion of the *MAPT* gene); see Supplementary Figure 2. Each recurrent CNV was detected in 2 PSP patients.

We focused on the 2 people with duplication in 17q21.31 (Fig. 3). We identified a 40-year-old (at time of inclusion) European French male patient with autopsy-confirmed PSP and an unusually early age of onset of 37, who carried a 460-kb duplication, copy number = 3, at chr17:43,728,377-44,189,068 (hg19). At study enrollment, he had experienced an akinetic-rigid syndrome associated with genitourinary incontinence, bulbar/pseudobulbar signs, and behavioral disturbances. The initial Hoehn & Yahr score was 2.5. MRI revealed enlargement of the aqueduct of Sylvius and lateralization of the marginal rim of the putamen (left greater than right), suggestive of PSP after rating with the previously validated NNIPPS MRI rating scale.[24] The patient progressed with

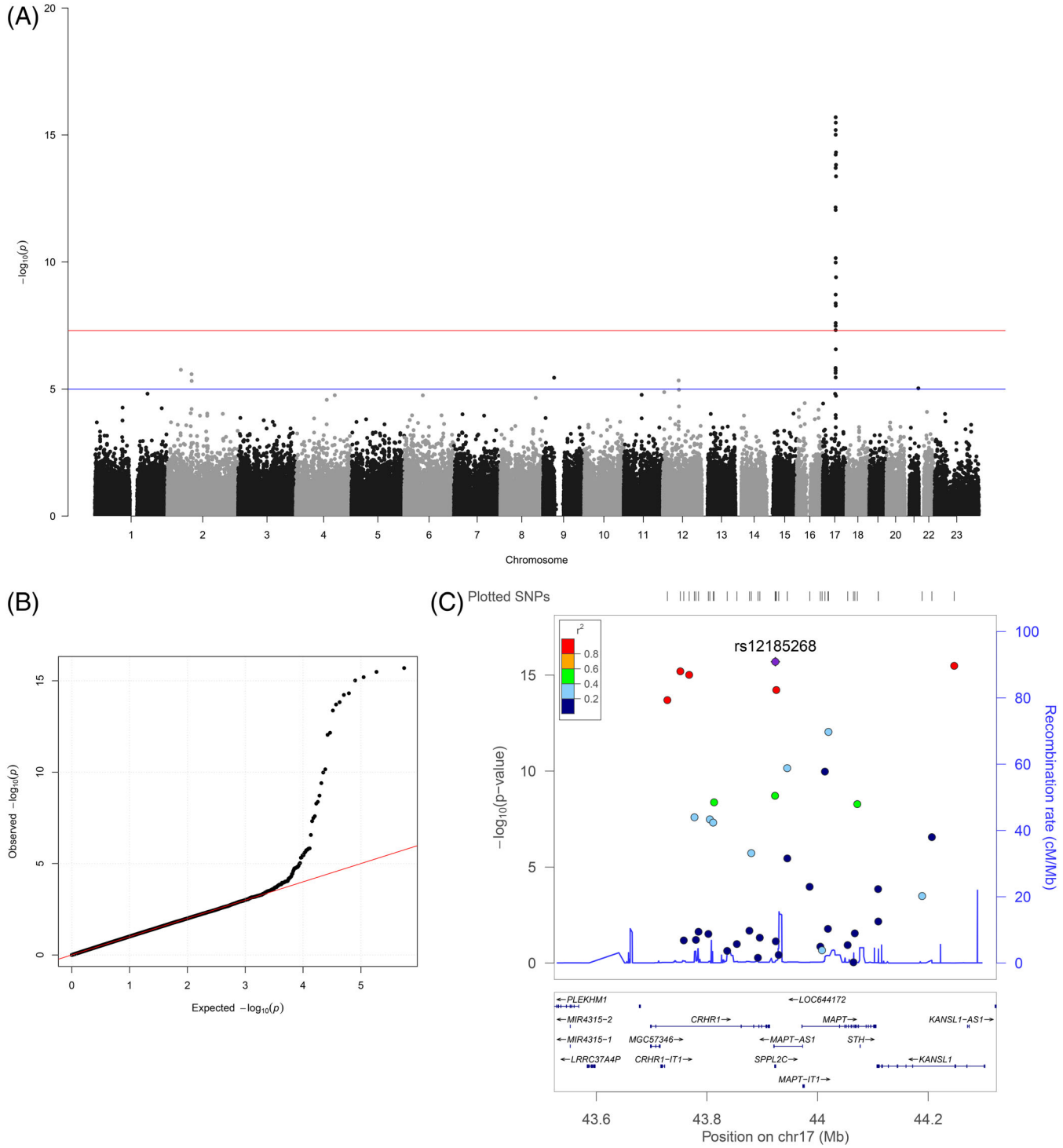

FIG. 1. (A) Manhattan plot depicting *P* values for GWAS. The genome-wide level of significance ($P < 5 \times 10^{-8}$) and significance level of interest ($P < 1 \times 10^{-5}$) are shown. (B) Q-Q plot showing observed and expected *P* values for the association analysis. Each cross on the graph represents an individual SNP. (C) Genetic architecture of the associated region on chromosome 17. Region plot generated using LocusZoom (http://locuszoom.org/). SNPs are colored according to r^2, a measure of linkage disequilibrium, with rs12185268 represented by a diamond. [Color figure can be viewed at wileyonlinelibrary.com]

aggravation of bulbar and pseudobulbar signs, including dysphagia. The patient died at age 42 from acute respiratory failure and aspiration pneumonia. A similar duplication (copy number = 3) was found in a 61-year-old (at time of inclusion) European French woman with autopsy-confirmed PSP diagnosed at age 57, affecting 503 kb at chr17:43,685,926-44,189,068. Her presentation was characterized by an akinetic-rigid syndrome with cerebellar, bulbar and pseudobulbar, pyramidal, and cognitive and behavioral dysfunction, alongside a resting tremor. The initial Hoehn & Yahr score was 4. Axial T2 MRI demonstrated hypointensity of the red nuclei,

TABLE 1. Summary of meta-analysis

					NNIPPS			Hoglinger			Meta-analysis					
Chr	SNP	Gene	A1	A2	MAF	OR	*P*	MAF	OR	*P*	*P*	P (R)	OR	OR (R)	Q	I
17	rs242557	*MAPT*	A	G	0.48	1.93	9.03×10^{-13}	0.50	1.96	4.20×10^{-70}	1.05×10^{-20}	1.05×10^{-20}	1.94	1.94	0.91	0
3	rs1768208	*MOBP*	T	C	0.27	1.52	2.65×10^{-05}	0.35	1.39	1.00 × 10-16	3.29×10^{-13}	3.29×10^{-13}	1.42	1.42	0.43	0
1	rs1411478	*STX6*	A	G	0.41	1.19	0.05	0.46	1.27	2.30×10^{-10}	3.45×10^{-10}	3.45×10^{-10}	1.25	1.25	0.55	0
1	rs6687758	*intergenic*	G	A	0.24	1.31	0.01	0.23	1.25	2.80×10^{-07}	1.22×10^{-08}	1.22×10^{-08}	1.26	1.26	0.69	0
12	rs11568563	*SLCO1A2*	C	A	0.08	1.45	0.03	0.08	1.45	1.90×10^{-07}	1.71×10^{-08}	1.71×10^{-08}	1.45	1.45	0.99	0
10	rs2142991	*BMS1*	C	T	0.16	0.82	0.13	0.14	0.77	4.90×10^{-07}	1.65×10^{-07}	1.65×10^{-07}	0.78	0.78	0.66	0
17	rs11650531	*WNT3*	C	T	0.31	1.54	1.85×10^{-05}	0.33	1.19	0.000068	8.11×10^{-08}	0.02	1.24	1.33	0.02	81.52
2	rs7571971	*EIF2AK3*	T	C	0.29	1.01	0.94	0.31	1.33	3.20×10^{-13}	2.26×10^{-08}	0.22	1.28	1.18	0.03	78.94
4	rs6852535	*IL2/IL21*	A	G	0.30	1.01	0.89	0.29	0.81	1.30×10^{-07}	1.17×10^{-06}	0.29	0.84	0.89	0.04	76.51
6	rs12203592	*IRF4*	T	C	0.16	1.01	0.96	0.16	0.67	6.10×10^{-14}	1.16×10^{-07}	0.30	0.73	0.81	0.00	87.37
2	rs6547705	*CD8B*	G	A	0.20	1.03	0.79	0.18	0.78	5.20×10^{-08}	5.90×10^{-07}	0.35	0.81	0.88	0.03	79.01
6	rs2493013	*EXOC2*	A	C	0.20	0.93	0.51	0.23	1.23	6.00×10^{-07}	8.78×10^{-06}	0.53	1.19	1.09	0.02	82.09
Combined NNIPPS, Hoglinger, Cruchaga					NNIPPS and Hoglinger results as above			Cruchaga								
17	rs242557	*MAPT*	G	A				0.47	1.69	0.003	1.58×10^{-22}	$\mathbf{1.58 \times 10^{-22}}$	1.9061	**1.9061**	0.7718	0
Cruchaga, NNIPPS					NNIPPS			Cruchaga								
17	rs1880753	*MAPT*	G	A	0.33	0.59	4.79E-08	0.38	0.54	0.0004	8.62×10^{-11}	$\mathbf{8.62 \times 10^{-11}}$	0.5804	**0.5804**	0.6361	0

A1, the minor allele; A2, major allele; *P,* fixed-effects meta-analysis *P* value; P(R), random-effects meta-analysis *P* value; OR, fixed-effects meta-analysis OR; OR(R), random-effects OR estimate; Q, *P* value for Cochrane's Q-statistic; I, I^2 heterogeneity index.
Where there is study heterogeneity (measured by a high I^2 value) and the *P* values differ between the fixed- and random-effects models, the more conservative *P* value (random effects) should be taken, to account for the heterogeneity.

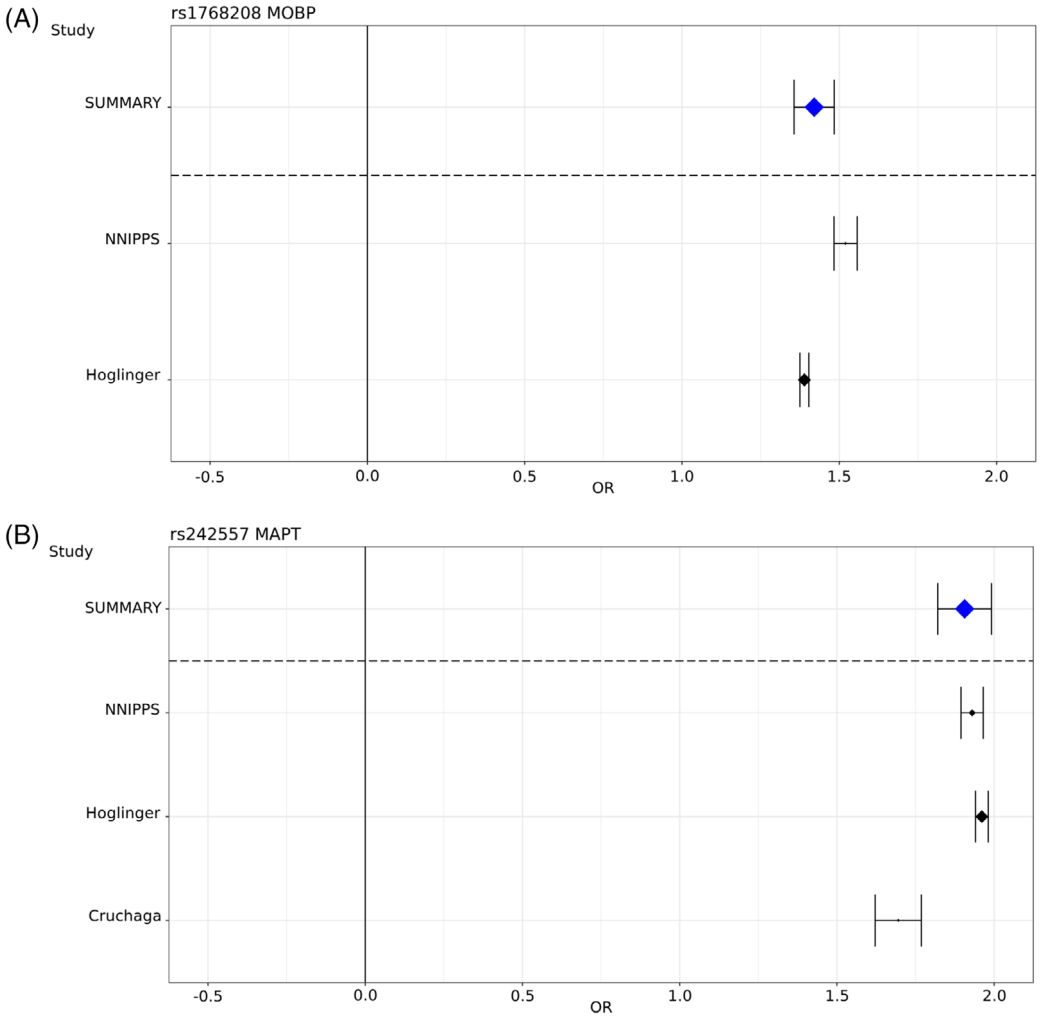

FIG. 2. Forest plots of the meta-analysis of (A) rs1768208 and (B) rs242557. Point estimate denotes the OR of the studies. Error bars show 95% confidence intervals (CIs). [Color figure can be viewed at wileyonlinelibrary.com]

dentate nuclei, and substantia nigra, and lateralization of the marginal rim of the putamen (left greater than right), most consistent with MSA. The patient developed autonomic dysfunction and worsening oculomotor signs and died at age 63 from cardiac arrest. Despite the atypical presentations, both patients were found to have typical neuropathological findings of PSP.

In both patients, duplications that spanned the entirety of the *MAPT* gene, as well as the last 7 exons of *CRHR1* and the first 6 exons of *KANSL1*, were noted. The presence of the discovered *MAPT* duplications were validated by qPCR, using 2 distinct primers within the *MAPT* gene; both primer pairs found a copy number of 3 for the corresponding amplified genomic regions in each patient.

Discussion

We present the results of a GWAS of PSP cases from the NNIPPS and BBBIPPS studies, which comprised

TABLE 2. Population-attributable risks (PAR) and penetrance for alleles A1 and A2 in homozygous and heterozygous states for the top 3 SNPs associated at genome-wide significance from meta-analysis

Chr	SNP	Gene	A1	A2	PAR	Penetrance			% Genetic variance
						A1A1	A1A2	A2A2	
17	rs242557	*MAPT*	A	G	0.37	0.00010	5.01×10^{-05}	3.13×10^{-05}	0.33
3	rs1768208	*MOBP*	T	C	0.26	9.21×10^{-05}	5.99×10^{-05}	3.7×10^{-05}	0.14
1	rs1411478	*STX6*	A	G	0.080	7.27×10^{-05}	4.51×10^{-05}	4.6 × 10-05	0.11

The proportion (%) of genetic variance at each locus is based on the minor allele frequency.

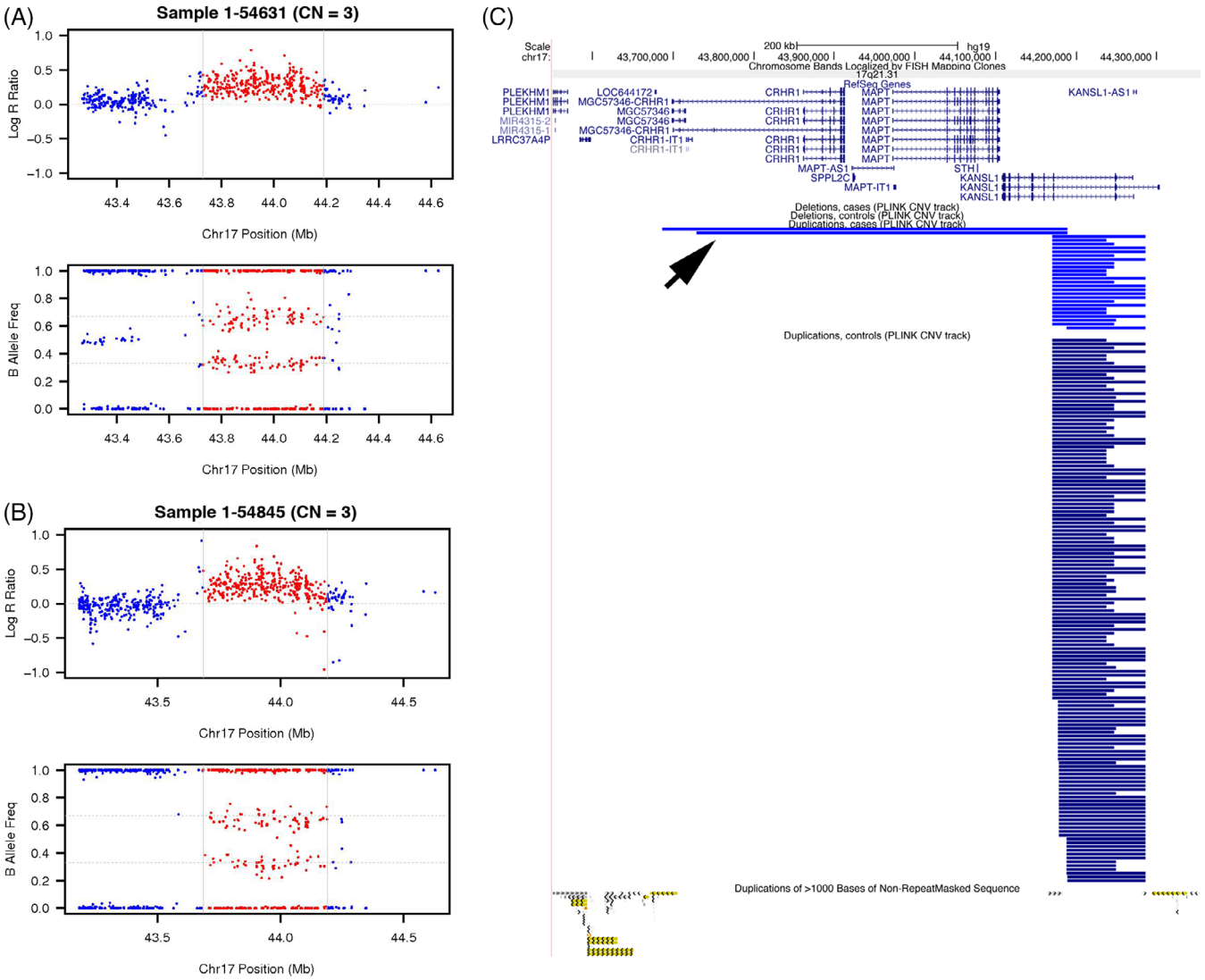

FIG. 3. Recurrent rare CNVs within the 17q21.31 region. A CNV was identified in a 40-year old man with PSP within 17q21.31 based on signal intensity (A). A second CNV with similar break points in chr17q21.31 was found in a 62-year old woman with PSP (B). These CNVs overlapped *MAPT* and neighboring genes, as demonstrated on the UCSC Genome Browser Track (C; arrowhead). Segmental duplications are also shown. [Color figure can be viewed at wileyonlinelibrary.com]

283 well-characterized European patients. Through this, we have provided independent replication that *MOBP* and *STX6* are PSP risk genes. We were also able to identify multiple associations on chromosome 17, reflecting linkage disequilibrium with a known risk haplotype in the *MAPT* gene, supporting the diagnostic validity of the NNIPPS and BBBIPPS studies.

Although we did not detect an overall increase in rare CNV burden in cases, we did identify 2 genomic duplications spanning the entire *MAPT* locus in 2 patients with PSP. Because *MAPT* was the only gene fully contained within these duplications, this may be the first report of *MAPT* gene duplications causing PSP, although we note that additional cases are needed to confirm this association statistically because of the rarity of such events.

Replication of Association at the *MOBP* and *STX6* Loci

Polymorphisms in *MOBP* and *STX6* have been demonstrated as risk variants in PSP through multiple genome-wide association studies.[4-6] Of note, the GWAS described by Chen et al[5] includes the NNIPPS cohort as a subset of samples. *MOBP*, on chromosome 3p22.1, encodes a protein that is produced by oligodendrocytes and expressed in central nervous system myelin, preferentially distributed in the brain stem and cerebellum, areas affected in PSP.[25] In our GWAS, rs1768208 (*MOBP*) conferred an OR of 1.52 for susceptibility to PSP, with confirmation on meta-analysis (OR, 1.42; $P = 3.29 \times 10^{-13}$), amounting to a population-attributable risk of 0.26. This SNP is situated in a cluster of SNPs in

high linkage disequilibrium at the 5′ end of the *MOBP* gene, consistent with a role in gene regulation.[26]

Oligodendroglial inclusions have been found in areas of the central nervous system associated with degeneration in PSP postmortem brains.[27] It has been suggested that oligodendrocyte dysfunction can be linked to *MOBP* via myelination errors, given that oligodendrocyte tau inclusions are characteristic in PSP.

A GWAS of corticobasal degeneration (CBD), encompassing 152 cases and 3,311 controls, showed the minor allele of rs1768208 in *MOBP* to confer an OR of 1.65 ($P = 3.86 \times 10^{-5}$) in the discovery stage and an OR of 1.89 ($P = 1.30 \times 10^{-3}$) in the replication stage.[28] Furthermore, a risk locus in MOBP has been confirmed to be associated at genome-wide significance with amyotrophic lateral sclerosis (ALS).[29] This shows a shared susceptibility locus between the three neurodegenerative disorders of ALS, CBD, and PSP outside the *MAPT* region. The overlap in genetic risk profiles between these disorders provides further credibility to the role of *MOBP* in PSP pathogenesis.

We have also confirmed that rs1411478 in *STX6* is significantly associated with PSP susceptibility ($P = 3.45 \times 10^{-10}$). *STX6* encodes a SNAP (Soluble NSF Attachment Protein) REceptor (SNARE) protein localized at the trans-Golgi network, which may influence transportation of misfolded proteins such as tau.[30] Expression quantitative trait locus analyses have found significantly lower expression of *STX6* in the white matter in carriers of the risk allele.[26] Similar to *MOBP*, the differential expression of *STX6* could imply the involvement of white matter and dysmyelination in PSP pathogenesis. The population-attributable risk for this SNP is 0.08.

The association of SNPs at loci near *MOBP* and *STX6* with PSP highlights the importance of the pathological role of tau in white-matter atrophy and damage, and suggests that this may be an area of fruitful future investigations. In addition, the strength of these results indicates that our PSP patient population is highly similar to pathologically confirmed cases of PSP and furthermore points to a relatively oligogenic genetic architecture in the disease.

Exploring *MAPT* and the Surrounding Haplotype

Previous studies have consistently identified an extended haplotype on 17q.21, which contains *MAPT*, as the main genetic locus associated with sporadic cases of PSP. The 6 major tau protein isoforms expressed are generated from alternative splicing by polymorphisms in *MAPT*, with PSP being a primary 4-repeat tauopathy.[3] One study suggests that the population-attributable risk for the *MAPT* H1 haplotype alone is ~68% in PSP.[18]

The H1c subhaplotype, tagged by rs242557, is associated with exon 10 splicing and increased tau expression.[3,17] We found a consistent effect size across the studies, with a significant OR of 1.94 for the minor allele. This gives a population-attributable risk of 0.37. The strongest haplotype association of rs242557 was with rs2435200 on *MAPT* (omnibus $P = 7.61 \times 10^{-16}$). SNP rs1880753 was associated with PSP with an OR of 0.6 ($P = 8.62 \times 10^{-11}$), supporting a previous study of Iberian patients.[17] Our haplotype association confirms the role of the *MAPT* H1 haplotype in PSP pathogenesis.

Copy Number Variation in PSP

We found 2 instances of duplication of *MAPT* in patients with PSP that spanned minimal additional genes. Taken together with a copious body of evidence connecting *MAPT* with PSP genetically and neuropathologically, this finding suggests that duplication of *MAPT* is a significant although relatively uncommon cause of PSP. Duplications spanning the *MAPT* gene have been previously reported in 12 patients, including in children with mild learning difficulties, developmental delay, and variable dysmorphic features[31]; a patient with a presumed familial frontotemporal dementia phenotype[32]; 3 siblings with early-onset Alzheimer's disease[33]; and 5 additional patients with Alzheimer's-like cognitive impairment.[34] In most cases, the extent of the causal CNV is not clearly defined. In the cases with Alzheimer's disease, the CNV appeared to span the *MAPT* gene[33] and possibly include the *CRHR1* and *KANSL1* genes,[34] as in our case, suggesting that the clinical phenotype of *MAPT* duplications may have variable expressivity.

The localization of the observed duplications also narrows the search space and provides additional mechanistic insight compared with GWAS SNPs alone. Although the CNVs span several genes in addition to *MAPT*, which includes *CRHR1*, *STH* (contained within *MAPT* intron 9), and *KANSL1*, the region is much smaller than the approximated 1-Mb *MAPT* haplotype on which previously identified duplications resulted in developmental anomalies. We favor a causal role for *MAPT* over other genes that are spanned by the CNV because of the clear pathogenic connection; however, from a genetic standpoint, the pathogenicity of the neighboring genes cannot be excluded. Furthermore, the putative effect on gene expression is much clearer, whereas the 17q21.31 haplotype does not seem to affect overall expression and probably alters splicing of the *MAPT* exon 3.[35] Overall, the association of *MAPT* duplication with PSP for the first time provides direct genetic evidence that specific overexpression of tau can lead to PSP. This is consistent with data from models showing that human wild-type tau overexpression is sufficient to cause pathological tau deposition and neurodegeneration, even in the absence of mutations.[36]

We identified 2 patients with duplications of the *MAPT* gene who each experienced early-onset, atypical

presentations of a Parkinson's-plus disorder. The clinical presentations in both cases featured autonomic involvement, and imaging characteristics in one case was more supportive of a diagnosis of MSA.[24] However, both cases had typical PSP postmortem neuropathology, supporting the role of the pathogenicity of the *MAPT* duplications. In addition to the 2 patients harboring duplications of the entire tau locus, we found several additional genic regions with rare, recurrent CNVs in PSP patients, including duplications on 2q37.1 (spanning *ALPP, ECEL1P2*, and *ALPPL2*), 4q31.21 (spanning *RNF150* and *ZNF330*), and 16p12.2 (spanning *METTL9, IGSF6*, and *OTOA*). However, given our sample size and the rarity of these events, the pathogenicity of these variants is difficult to establish without prior evidence of involvement in PSP. Our finding of rare CNVs within *MAPT* in PSP strongly suggests that CNVs can cause PSP, and future work building on our results may yet identify additional PSP genetic contributors within these regions.

Genes Flanking *MAPT*

We identified SNPs associated with PSP beyond the genome-wide significance level, situated in the *MAPT* genes and 12 other SNPs on the gene-rich 17q21. Previous studies have attempted to decipher whether the association is because of linkage disequilibrium with *MAPT* or independent pathogenic variants.[37,38]

Even at the telomeric end of this region with possible loss of linkage disequilibrium with *MAPT* beyond the wingless-type Mouse Mammary Tumor Virus integration-site family (*WNT3*) gene, associations are dependent on the *MAPT* haplotype.[37] Therefore, significant associations with *WNT3* variants rs415430 ($P = 4.83 \times 10^{-15}$) and rs2074404 ($P = 3.98 \times 10^{-10}$) and the *WNT3* rs11650531- rs415430 haplotype (omnibus $P = 1.21 \times 10^{-11}$) are not because of a direct and independent pathogenic role but related to the *MAPT* locus.[37,38] Conditioning for the most significantly associated SNP in *MAPT* in our study shows that the association is tagged by a single variant in *MAPT* with no evidence for independent association of the other SNPs in the region. The discovery of 2 CNVs spanning *MAPT* reinforces the idea that the *MAPT* gene is the central mediator of PSP risk in the 17q21 inversion region.

Limitations

The GWAS of 283 PSP patients and 4472 controls is limited in terms of sample size. However, given that PSP is a rare disorder, the well-characterized patient cohort prevents selection of falsely positive cases for genetic analysis. Not all control individuals in the study were sex-matched, although they were matched for geographical origin and ancestry. For a disease as rare as PSP, sex matching does not increase statistical power, whereas it would for a more common disorder.

Conclusion

We have identified genetic loci associated with PSP in an analysis of a multicenter European study of well-characterized patients. This study confirms in an independent sample *MOBP* and *STX6* as additional PSP susceptibility loci, both of which are implicated in myelin metabolism. This potential link should be investigated further and may correlate with neuroimaging findings. The 17q2.31 locus was identified as harboring a PSP risk variant, and specific CNV duplication of *MAPT* in 2 cases of early-onset atypical PSP provides support for *MAPT* as the disease gene, rather than one of the many neighboring genes in this region of high linkage disequilibrium. Clinicians should be aware of the uncommon *MAPT* duplication as a possibility and PSP as the diagnosis, especially in patients presenting with a young-onset, atypical Parkinson's-plus syndrome.

Acknowledgments: We thank the patients and their families for their commitment and altruism, as well as the French and UK PSP Associations, the UK Parkinson's Disease Research Group, and the Richard Rainwater family and the Tau Consortium for their help and support. We are grateful to the many colleagues who were not formally part of the NNIPPS Study Group but whose support contributed to the success of the study. We thank the participants and investigators of the Three-City Study for the recruitment of the French controls. We thank the Chen Family Foundation for supporting the SNP genotyping. Z.C. was supported by funding from the National Institute for Health Research (NIHR) Academic Clinical Fellowship scheme. J.A.C. was supported by grant F31 NS084556 from the National Institute of Neurological Disorders and Stroke (NINDS), D.H.G., J.L., and G.C. are supported by the Tau Consortium, and J.B. is supported by the Chen Family Foundation. A.A.C was supported by the UK NIHR Maudsley Biomedical Research Centre at South London and Maudsley NHS Foundation Trust (SLaM) and King's College London, UK.

References

1. Litvan I, Agid Y, Calne D, et al. Clinical research criteria for the diagnosis of progressive supranuclear palsy (Steele-Richardson-Olszewski syndrome): report of the NINDS-SPSP international workshop. Neurology 1996;47(1):1–9.
2. Nath U, Ben-Shlomo Y, Thomson RG, et al. The prevalence of progressive supranuclear palsy (Steele-Richardson-Olszewski syndrome) in the UK. Brain 2001;124(Pt 7):1438–1449.
3. Baker M, Litvan I, Houlden H, et al. Association of an extended haplotype in the tau gene with progressive supranuclear palsy. Human molecular genetics 1999;8(4):711–715.
4. Sanchez-Contreras MY, Kouri N, Cook CN, et al. Replication of progressive supranuclear palsy genome-wide association study identifies SLCO1A2 and DUSP10 as new susceptibility loci. Mol Neurodegener 2018;13(1):37.
5. Chen JA, Chen Z, Won H, et al. Joint genome-wide association study of progressive supranuclear palsy identifies novel susceptibility loci and genetic correlation to neurodegenerative diseases. Mol Neurodegener 2018;13(1):41.
6. Hoglinger GU, Melhem NM, Dickson DW, et al. Identification of common variants influencing risk of the tauopathy progressive supranuclear palsy. Nat Genet 2011;43(7):699–705.
7. Chen JA, Wang Q, Davis-Turak J, et al. A multiancestral genome-wide exome array study of Alzheimer disease, frontotemporal dementia, and progressive supranuclear palsy. JAMA Neurol 2015;72(4):414–422.

8. Bensimon G, Ludolph A, Agid Y, et al. Riluzole treatment, survival and diagnostic criteria in Parkinson plus disorders: the NNIPPS study. Brain 2009;132(Pt 1):156–171.
9. Krawczak M, Nikolaus S, von Eberstein H, Croucher PJ, El Mokhtari NE, Schreiber S. PopGen: population-based recruitment of patients and controls for the analysis of complex genotype-phenotype relationships. Community Genet 2006;9(1):55–61.
10. Ancelin ML, Carriere I, Barberger-Gateau P, et al. Lipid lowering agents, cognitive decline, and dementia: the three-city study. J Alzheimers Dis 2012;30(3):629–637.
11. Purcell S, Cherny SS, Sham PC. Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. Bioinformatics 2003;19(1):149–150.
12. Weale ME. Quality control for genome-wide association studies. Methods Mol Biol 2010;628:341–372.
13. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007;81(3):559–575.
14. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 2006;38.
15. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nat Genet 2012;44(8):955–959.
16. Witte JS, Visscher PM, Wray NR. The contribution of genetic variants to disease depends on the ruler. Nat Rev Genet 2014;15(11):765–776.
17. Cruchaga C, Vidal-Taboada JM, Ezquerra M, et al. 5′-Upstream variants of CRHR1 and MAPT genes associated with age at onset in progressive supranuclear palsy and cortical basal degeneration. Neurobiol Dis 2009;33(2):164–170.
18. Melquist S, Craig DW, Huentelman MJ, et al. Identification of a novel risk locus for progressive supranuclear palsy by a pooled genomewide scan of 500,288 single-nucleotide polymorphisms. Am J Hum Genet 2007;80(4):769–778.
19. Wang K, Li M, Hadley D, et al. PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. Genome Res 2007;17(11):1665–1674.
20. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 2010;26(6):841–842.
21. Bucan M, Abrahams BS, Wang K, et al. Genome-Wide Analyses of Exonic Copy Number Variants in a Family-Based Study Point to Novel Autism Susceptibility Genes. PLoS Genet 2009;5(6):e1000536.
22. Luo R, Sanders Stephan J, Tian Y, et al. Genome-wide Transcriptome Profiling Reveals the Functional Impact of Rare De Novo and Recurrent CNVs in Autism Spectrum Disorders. Am J Hum Genet 2012;91(1):38–55.
23. Tsai M-F, Lin Y-J, Cheng Y-C, et al. PrimerZ: streamlined primer design for promoters, exons and human SNPs. Nucleic Acids Res 2007;35(suppl 2):W63–W65.
24. Rolland Y, Vérin M, Payan CA, et al. A new MRI rating scale for progressive supranuclear palsy and multiple system atrophy: validity and reliability. J Neurol Neurosurg Psychiatry 2011;82(9):1025–1032.
25. Yamamoto Y, Mizuno R, Nishimura T, et al. Cloning and expression of myelin-associated oligodendrocytic basic protein. A novel basic protein constituting the central nervous system myelin. J Biol Chem 1994;269(50):31725–31730.
26. Ferrari R, Ryten M, Simone R, et al. Assessment of common variability and expression quantitative trait loci for genome-wide associations for progressive supranuclear palsy. Neurobiol Aging 2014;35(6):1514 e1511–1514 e1512.
27. Ahmed Z, Asi YT, Lees AJ, Revesz T, Holton JL. Identification and quantification of oligodendrocyte precursor cells in multiple system atrophy, progressive supranuclear palsy and Parkinson's disease. Brain Pathol 2013;23(3):263–273.
28. Kouri N, Ross OA, Dombroski B, et al. Genome-wide association study of corticobasal degeneration identifies risk variants shared with progressive supranuclear palsy. Nat Commun 2015;6:7247.
29. van Rheenen W, Shatunov A, Dekker AM, et al. Genome-wide association analyses identify new risk variants and the genetic architecture of amyotrophic lateral sclerosis. Nat Genet 2016;48:1043.
30. Jahn R, Scheller RH. SNAREs—engines for membrane fusion. Nat Rev Mol Cell Biol 2006;7(9):631–643.
31. Gregor A, Krumbiegel M, Kraus C, Reis A, Zweier C. De novo triplication of the MAPT gene from the recurrent 17q21.31 microdeletion region in a patient with moderate intellectual disability and various minor anomalies. Am J Med Genet A 2012;158a(7):1765–1770.
32. Rovelet-Lecrux A, Hannequin D, Guillin O, et al. Frontotemporal dementia phenotype associated with MAPT gene duplication. J Alzheimers Dis 2010;21(3):897–902.
33. Hooli BV, Kovacs-Vajna ZM, Mullin K, et al. Rare autosomal copy number variations in early-onset familial Alzheimer's disease. Mol Psychiatry 2014;19(6):676–681.
34. Le Guennec K, Quenez O, Nicolas G, et al. 17q21.31 duplication causes prominent tau-related dementia with increased MAPT expression. Mol Psychiatry 2017;22(8):1119–1125.
35. Trabzuni D, Wray S, Vandrovcova J, et al. MAPT expression and splicing is differentially regulated by brain region: relation to genotype and implication for tauopathies. Hum Mol Ggenet 2012;21(18):4094–4103.
36. Jackson GR, Wiedau-Pazos M, Sang TK, et al. Human wild-type tau interacts with wingless pathway components and produces neurofibrillary pathology in Drosophila. Neuron 2002;34(4):509–519.
37. Pittman AM, Myers AJ, Duckworth J, et al. The structure of the tau haplotype in controls and in progressive supranuclear palsy. Hum Mol Genet 2004;13(12):1267–1274.
38. Hoglinger GU, Melhem NM, Dickson DW, et al. Identification of common variants influencing risk of the tauopathy progressive supranuclear palsy. Nat Genet 2011;43(7):699–705.

Supporting Data

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