UC San Diego
UC San Diego Previously Published Works

Title

Genome-wide association study of corticobasal degeneration identifies risk variants shared with progressive supranuclear palsy

Permalink

https://escholarship.org/uc/item/10b9b78t

Journal

Nature Communications, 6(1)

ISSN

2041-1723

Authors

Kouri, Naomi

Ross, Owen A

Dombroski, Beth

et al.

Publication Date

2015

DOI

10.1038/ncomms8247

Copyright Information

Peer reviewed

ARTICLE

Genome-wide association study of corticobasal degeneration identifies risk variants shared with progressive supranuclear palsy

Naomi Kouri[1,*], Owen A. Ross[1,*], Beth Dombroski[2], Curtis S. Younkin[1,3], Daniel J. Serie[3], Alexandra Soto-Ortolaza[1], Matthew Baker[1], Ni Cole A. Finch[1], Hyejin Yoon[1], Jungsu Kim[1], Shinsuke Fujioka[1], Catriona A. McLean[4], Bernardino Ghetti[5], Salvatore Spina[5], Laura B. Cantwell[2], Martin R. Farlow[6], Jordan Grafman[7,8], Edward D. Huey[9], Mi Ryung Han[2], Sherry Beecher[2], Evan T. Geller[2], Hans A. Kretzschmar[10,‡], Sigrun Roeber[10], Marla Gearing[11], Jorge L. Juncos[12], Jean Paul G. Vonsattel[13], Vivianna M. Van Deerlin[2], Murray Grossman[14], Howard I. Hurtig[14], Rachel G. Gross[14], Steven E. Arnold[15], John Q. Trojanowski[2], Virginia M. Lee[2], Gregor K. Wenning[16], Charles L. White[17], Günter U. Höglinger[18,19,20], Ulrich Müller[21], Bernie Devlin[22], Lawrence I. Golbe[23], Julia Crook[1,3], Joseph E. Parisi[24], Bradley F. Boeve[25], Keith A. Josephs[25], Zbigniew K. Wszolek[26], Ryan J. Uitti[26], Neill R. Graff-Radford[26], Irene Litvan[27], Steven G. Younkin[1], Li-San Wang[2], Nilüfer Ertekin-Taner[1,26], Rosa Rademakers[1], Hakon Hakonarson[28], Gerard D. Schellenberg[2,*] & Dennis W. Dickson[1,*]

Corticobasal degeneration (CBD) is a neurodegenerative disorder affecting movement and cognition, definitively diagnosed only at autopsy. Here, we conduct a genome-wide association study (GWAS) in CBD cases ($n = 152$) and 3,311 controls, and 67 CBD cases and 439 controls in a replication stage. Associations with meta-analysis were 17q21 at *MAPT* ($P = 1.42 \times 10^{-12}$), 8p12 at *lnc-KIF13B-1*, a long non-coding RNA (rs643472; $P = 3.41 \times 10^{-8}$), and 2p22 at *SOS1* (rs963731; $P = 1.76 \times 10^{-7}$). Testing for association of CBD with top progressive supranuclear palsy (PSP) GWAS single-nucleotide polymorphisms (SNPs) identified associations at *MOBP* (3p22; rs1768208; $P = 2.07 \times 10^{-7}$) and *MAPT* H1c (17q21; rs242557; $P = 7.91 \times 10^{-6}$). We previously reported SNP/transcript level associations with rs8070723/*MAPT*, rs242557/*MAPT*, and rs1768208/*MOBP* and herein identified association with rs963731/*SOS1*. We identify new CBD susceptibility loci and show that CBD and PSP share a genetic risk factor other than *MAPT* at 3p22 *MOBP* (myelin-associated oligodendrocyte basic protein).

[1] Department of Neuroscience, Mayo Clinic College of Medicine, Jacksonville, Florida 32224, USA. [2] Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. [3] Division of Biomedical Statistics and Informatics, Department of Health Sciences Research, Mayo Clinic College of Medicine, Jacksonville, Florida 32224, USA. [4] Victorian Brain Bank Network, Mental Health Research Institute, Parksville, Victoria 3052, Australia. [5] Department of Pathology and Laboratory Medicine, Indiana University School of Medicine, Indianapolis, Indiana 46202, USA. [6] Department of Neurology, Indiana University School of Medicine, Indianapolis, Indiana 46202, USA. [7] Cognitive Neuroscience Laboratory, Brain Injury Research, Rehabilitation Institute of Chicago, Chicago, Illinois 60611, USA. [8] Department of Physical Medicine and Rehabilitation, Northwestern University, Illinois 60208, USA. [9] Departments of Psychiatry and Neurology, Columbia University, New York, New York 10027, USA. [10] Institut for Neuropathology and Prion Research and Brain Net Germany, Ludwig-Maximilians-Universität, Munich 80539, Germany. [11] Department of Pathology and Laboratory Medicine, Emory University School of Medicine, Atlanta, Georgia 30307, USA. [12] Department of Neurology, Emory University, Atlanta, Georgia 30307, USA. [13] Department of Pathology and the Taub Institute for Research on Alzheimer's disease and the Aging Brain, Columbia University, New York, New York 10027, USA. [14] Department of Neurology, University of Pennsylvania Health System, Philadelphia, Pennsylvania 19104, USA. [15] Department of Psychiatry, Center for Neurobiology and Behavior, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104, USA. [16] Department of Neurology, Innsbruck Medical University, Innsbruck 6020, Austria. [17] Department of Pathology, University of Texas Southwestern Medical Center, Dallas, Texas 75235, USA. [18] Department of Neurology, Technical University Munich, 81377 Munich, Germany. [19] Department of Translational Neurodegeneration, German Center for Neurodegenerative Diseases (DZNE), 81677 Munich, Germany. [20] Department of Neurology, Philipps University, 35033 Marburg, Germany. [21] Institut for Humangenetik, Justus-Liebig-Universität, Giessen 35390, Germany. [22] Department of Human Genetics, University of Pittsburgh, Pittsburg, Pennsylvania 15260, USA. [23] Department of Neurology, Rutgers Robert Wood Johnson Medical School, New Brunswick, New Jersey 08901, USA. [24] Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, Minnesota 55905, USA. [25] Department of Neurology, Mayo Clinic, Rochester, Minnesota 55905, USA. [26] Department of Neurology, Mayo Clinic, Jacksonville, Florida 32224, USA. [27] Department of Neurosciences, University of California, San Diego, La Jolla, California 92093, USA. [28] Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania 19104, USA. * These authors contributed equally to this work. ‡ Deceased. Correspondence and requests for materials should be addressed to D.W.D. (email: dickson.dennis@mayo.edu) or to N.K. (email: kouri.naomi@mayo.edu).

CBD is a late-onset neurodegenerative disorder that can only be diagnosed upon neuropathologic examination, which has impeded the ability to conduct genome-wide association studies (GWASs). Similar to progressive supra-nuclear palsy (PSP), Alzheimer's disease (AD), and chronic traumatic encephalopathy, CBD is a tauopathy, having abnormal aggregates of microtubule-associated protein tau in the brain[1]. Neuropathologic diagnostic criteria for CBD are based on tau immunohistochemistry, requiring tau inclusions in neurons and glia, with tau astrocytic plaques, and extensive thread-like pathology in both grey matter and white matter[1].

First described in 1968 by Rebeiz and colleagues, three CBD cases were identified to have a distinct movement disorder and neuropathologic profile[2]. These patients presented with what is now termed corticobasal syndrome (CBS) where patients can exhibit levodopa-unresponsive parkinsonism, asymmetric akinesia and rigidity, accompanied by ideomotor apraxia, dystonia, and myoclonus. Patients who present with the archetypal CBS, however, do not necessarily have CBD pathology upon neuropathologic examination. Furthermore, autopsy-confirmed CBD cases can often present with several different clinical syndromes[3], which are frequently associated with other underlying neurodegenerative disorders such as PSP, AD, and frontotemporal dementia[4]. This phenotypic variability results in <50% of patients with CBS having CBD at autopsy[5–7].

There are few reported families affected with CBD or PSP[8–10], yet genetic association studies of PSP have repeatedly shown an increased risk for individuals carrying the H1 MAPT haplotype at chromosome 17q21 (refs 11–14). An association of CBD with the H1 haplotype was first reported in a series of 18 cases (16 corticobasal syndrome patients and two autopsy-confirmed CBD cases)[15] and subsequently confirmed in a larger autopsy series of 57 cases[16]. Recently, the first PSP GWAS[17] discovered three non-MAPT susceptibility loci at STX6, EIF2AK3, and MOBP, raising the possibility that additional non-MAPT genetic risk factors or modifiers may also exist for CBD. Recognition of the lack of specificity of the corticobasal syndrome for CBD has promoted increasing interest among clinicians to have autopsy confirmation of patients thought to have CBD. Brain banks focusing on atypical parkinsonian disorders, such as that supported by the Foundation for PSP|CBD and Related Brain Diseases[18], have facilitated accumulation of sufficient number of CBD cases to make identification of genetic contributions more feasible. Here, we report a two-stage GWAS in autopsy-proven CBD cases (n = 152 discovery stage, n = 67 replication stage) compared with controls (n = 3,311 discovery stage, n = 457 replication stage). The results confirm the association at MAPT and identify a novel susceptibility locus at 8p12 associated with CBD at the genome-wide significant level, as well as suggestive associations at 2p22 and 3p22. Importantly, we show that CBD and PSP share a genetic risk factor other than MAPT, at 3p22 MOBP. These findings highlight the genetic similarities and differences between CBD and PSP and suggest that the two disorders may, at least in part, share common disease processes.

Results

Discovery GWAS and replication. Genotype data from 152 autopsy-proven CBD cases and 3,311 controls (Table 1) were previously generated as part of the PSP GWAS[17]. The National Institute of Health Office of Rare Diseases Research criteria were used for making a neuropathologic diagnosis of CBD[1]. Discovery stage cases collected from multiple institutions (Supplementary Table 1) were genotyped by the Center for Applied Genomics at the Children's Hospital of Philadelphia using Human 660 W-Quad Infinium BeadChips, and control samples were genotyped using the Illumina Human HapMap550 Infinium BeadChip. Replication stage samples included 67 autopsy-proven CBD cases (Mayo Clinic Florida Brain Bank, 6 CBD cases that failed quality control and 61 new CBD cases collected since the GWAS genotyping was performed) and 457 control individuals (mean age at blood draw = 74 years) collected from Mayo Clinic Florida, all of which were clinically diagnosed as being free of any neurological disorder. For the replication stage, genotyping was performed on the 67 CBD cases using Taqman genotyping assays or Sanger sequencing and 30 CBD cases from the discovery stage as an internal genotyping control.

These data were subjected to accepted GWAS quality control measures. Individual samples were excluded with a genotype failure rate >2%, cryptic relatedness or duplicate samples based on identity by state. Genetic outliers were excluded on the basis of distance to the nearest-neighbour approach, and population substructure was assessed using multidimensional scaling (MDS) analysis with PLINK[19] (http://pngu.mgh.harvard.edu/~purcell/plink/; Supplementary Fig. 1). Quantile–quantile (QQ) plots and the genomic inflation factor (λ) show that the first principal component as a covariate in logistic regression analyses was sufficient to correct for population substructure based on a reduction of λ from the unadjusted ($\lambda = 1.06$) to adjusted ($\lambda = 1.01$; Supplementary Fig. 2).

For the discovery stage, we analysed association between disease and 533,898 single-nucleotide polymorphisms (SNPs) in 152 CBD cases and 3,311 control individuals by conditional logistic regression under an additive model using the first MDS principle component as a covariate using PLINK[19]. Given the neuropathologic, clinical, and genetic overlaps between CBD and PSP, we selected the top PSP GWAS SNPs at MAPT (the H1c subhaplotype-tagging SNP, rs242557), MOBP, EIF2AK3, and STX6 to test for association with CBD. The discovery stage identified one genome-wide significant association at 17q21.31,

Table 1 | Description of samples.

Cohort	Female			Age onset/blood draw			Age at death			Disease duration		
	Total	%	n*	Mean age	Range (s.d.)	n	Mean age	Range (s.d.)	n	Mean duration†	Range (s.d.)	n
Discovery stage‡												
CBD cases	152	50	150	64	33–80 (8.6)	103	70	44–91 (9.2)	144	6	2–14 (2.4)	103
Replication stage												
CBD cases	67	48	67	64	40–83 (7.9)	67	70	54–96 (7.8)	67	6	2–14 (2.5)	61
Controls	457	54	457	74	24–97 (13.1)	457	—	—	—	—	—	—

*n is the number of samples with available data. Values of n for each type of analysis do not add up to the total samples used because of missing values.
†Duration in years.
‡Controls (n = 3,311) were young healthy subjects recruited from the Children's Hospital of Philadelphia Health Care Network (see Online Methods for details).

which encompasses *MAPT* (rs393152, odds ratio (OR) = 3.45, $P = 6.71 \times 10^{-9}$) testing using conditional logistic regression analysis (Fig. 1a; Table 2) and 12 SNPs in LD with rs393152 (Supplementary Table 2). QQ plots including and excluding SNPs at 17q21 show that the *P* value distribution is consistent with the null, except for departures in the extreme tail (Supplementary Fig. 3). We also found nominal evidence for association testing using conditional logistic regression analysis at two novel susceptibility loci—rs643472 (OR = 1.88, $P = 7.12 \times 10^{-7}$), an intronic SNP in *lnc-KIF13B-1* (TCONS_00014956, a large intergenic non-coding RNA, or lincRNA) located between

KIF13B and *DUSP4* on chromosome 8p12, and rs963731 (OR = 2.46, $P = 2.04 \times 10^{-6}$), an intronic SNP in *SOS1* on chromosome 2q22 (Supplementary Fig. 4). Testing for association using conditional logistic regression analysis with the top PSP GWAS SNPs identified rs1768208 (OR = 1.65, $P = 3.86 \times 10^{-5}$), an intronic SNP in *MOBP* (myelin-associated oligodendrocyte basic protein), and rs242557 (OR = 1.48, $P = 1.20 \times 10^{-3}$) at the *MAPT* locus to be associated with CBD. The top SNP at *EIF2AK3* (rs7571971) did not show an association with CBD ($P = 0.057$; OR = 1.27), but the OR was of similar effect size as observed for PSP (Supplementary Table 3).
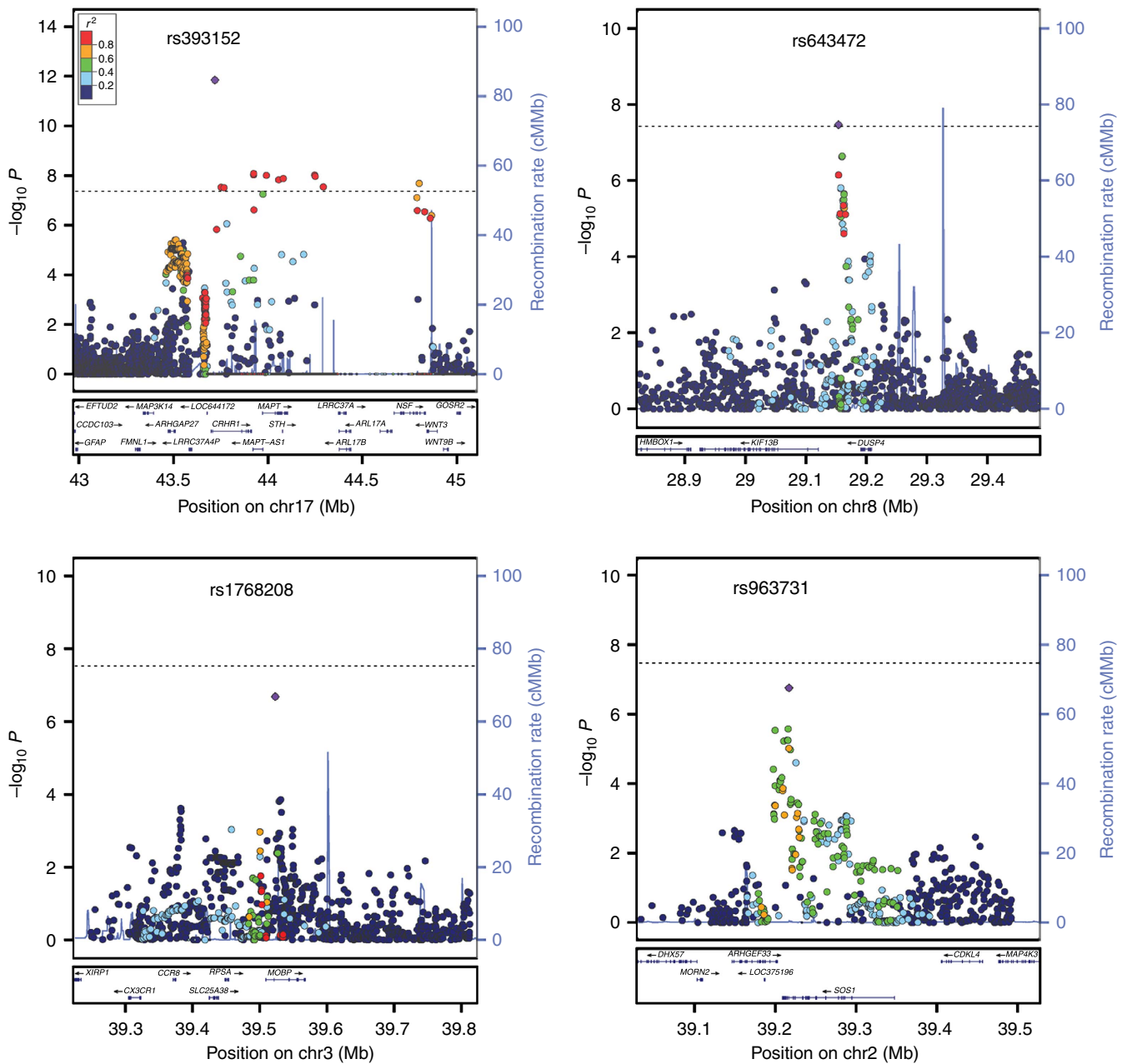
Figure 1 | Regional association plots. (**a–d**) Top loci associated with CBD compared with controls in the Discovery Stage (152 CBD, 3,311 controls) and the Replication Stage (67 CBD, 457 controls). The most significant association is at 17q21, tagging the 1.2 Mb inversion containing *MAPT* (**a**). New susceptibility loci for CBD are located at 8p12 (**b**), 3p22 (**c**) and 2p22 (**d**). Genetic association testing was performed using conditional logistic regression analysis under an additive model and corrected for population substructure using the first principal component from multidimensional scaling analysis (MDS) as a covariate ($\lambda = 1.01$). The *y* axis is $-\log_{10} P$ values and *x* axis is the genomic location of each SNP (circles) genotyped and imputed. Linkage disequilibrium coefficients were derived from hg19 (1,000 Genomes March 2012, European population) and local estimates of recombination rates are from HapMap samples (2008-03_rel22_B36; http://hapmap.ncbi.nlm.nih.gov). The random effects meta-analysis top SNP at each locus (purple diamonds).

Table 2 | Results from discovery stage, replication stage and meta-analysis.

SNP	Chr	Gene(s)	Stage*	MAF		OR (95% CI)	P
				Cases	Controls		
rs393152[†]	17q21	MAPT (H1/H2)	Discovery	0.079	0.24	3.45 (2.29–5.34)	6.71×10^{-9}
			Replication	0.068	0.24	4.37 (2.16–8.83)	4.12×10^{-5}
			Meta-analysis			3.70	1.42×10^{-12}
rs643472	8p12	lnc-KIF13B-1	Discovery	0.35	0.23	1.88 (1.46–2.40)	7.12×10^{-7}
			Replication	0.32	0.22	1.67 (1.11–2.52)	0.014
			Meta-analysis			1.82	3.41×10^{-8}
rs963731	2p22	SOS1	Discovery	0.12	0.054	2.46 (1.70–3.57)	2.04×10^{-6}
			Replication	0.09	0.049	2.21 (1.09–4.50)	0.029
			Meta-analysis			2.41	1.76×10^{-7}
rs1768208	3p22	MOBP	Discovery	0.39	0.29	1.65 (1.30–2.09)	3.86×10^{-5}
			Replication	0.39	0.25	1.89 (1.28–2.77)	1.30×10^{-3}
			Meta-analysis			1.71	2.07×10^{-7}
rs242557	17q21	MAPT (H1c)	Discovery	0.45	0.35	1.48 (1.17–1.88)	1.50×10^{-3}
			Replication	0.48	0.36	1.77 (1.25–2.52)	1.50×10^{-3}
			Meta-analysis			1.57	7.91×10^{-6}
rs1860743	7q36	PRKAG2	Discovery	0.20	0.11	2.05 (1.52–2.78)	3.46×10^{-6}
			Replication	0.09	0.13	0.71 (0.39–1.31)	0.27
			Meta-analysis			1.25	0.67
rs875125	21q22	TSPEAR	Discovery	0.17	0.10	2.02 (1.48–2.75)	7.92×10^{-6}
			Replication	0.05	0.08	0.65 (0.29–1.46)	0.29
			Meta-analysis			1.22	0.73

*Discovery stage (152 CBD, 3,311 controls); Replication stage (67 CBD, 457 controls). The discovery and replication stages were analysed by logistic regression under an additive model. The discovery stage analysis was adjusted for the first multidimensional scaling principle component, and the replication stage was adjusted for age and sex.
†The OR for rs393152 is referencing the risk associated with the major allele (MAPT H1 haplotype). The protective allele (MAPT H2 haplotype) has an OR (95% CI) of 0.29 (0.19–0.44) in the discovery stage and 0.23 (0.11–0.46) in the replication stage. Genome-wide significantceis defined as variants associated with $P < 5 \times 10^{-8}$. Chr, chromosome; CI, confidence interval; lnc, long non-coding (RNA); MAF, minor allele frequency; OR, odds ratio.

In an independent cohort of 67 autopsy-proven CBD cases from the Mayo Clinic Florida Brain Bank and 457 control individuals, a replication stage was performed for a total of seven SNPs: the top five SNPs from the discovery stage ($P < 10^{-5}$) and two PSP GWAS SNPs that showed nominally significant association with CBD (rs1768208 and rs242557) (Table 2). Association testing for these seven SNPs was performed using conditional logistic regression under an additive model with age and sex as covariates. Using PLINK to perform random effects meta-analysis, or the weighted average of the effect sizes of the discovery and replication stages, showed genome-wide significant associations at MAPT (rs393152, $P_{meta} = 1.4 \times 10^{-12}$) (Fig. 1a) and lnc-KIF13B-1 (rs643472, $P_{meta} = 3.4 \times 10^{-8}$; Table 2) (Fig. 1b). SNPs at MOBP (rs1768208, $P_{meta} = 2.1 \times 10^{-7}$) (Fig. 1c), SOS1 (rs963731, $P_{meta} = 1.8 \times 10^{-7}$) (Fig. 1d) and MAPT H1c haplotype (rs242557, $P_{meta} = 7.9 \times 10^{-6}$) had suggestive evidence for association with CBD. Performing the same random effects meta-analysis while including the number of H1 alleles as a covariate, the association was lost (rs242557, OR = 1.18, $P_{meta} = 0.11$). The MAFs for the top SNPs in younger controls of this study did not differ significantly from those of older controls ($N = 3,720$) in three data sets from the National Institute of Health Database for Genotypes and Phenotypes (dbGaP; http://www.ncbi.nlm.nih.gov/gap; Supplementary Table 4). Furthermore, we repeated the discovery stage CBD GWAS with a different control data set, 1,986 controls from dbGaP data set NeuroGenetics Research Consortium (NGRC-PD; phs000196.v2.p1), analysed by logistic regression under an additive model and found similar association results (Supplementary Table 5).

Genotype imputation was performed to increase the resolution of the top susceptibility loci. CBD cases and controls were imputed to samples from the 1,000 Genomes Project (May 2011, European reference population)[20]. Phasing was performed with SHAPEIT v2.r790 (ref. 21; https://mathgen.stats.ox.ac.uk/genetics_software/shapeit/shapeit.html) and genotype imputation was performed

with IMPUTE2 v2.3.1 (ref. 22; https://mathgen.stats.ox.ac.uk/impute/impute_v2.html). Genomic coordinates for imputation were selected at the most significant SNPs and their surrounding gene(s). Testing for association of CBD with imputed genotypes did not improve association results.

Top susceptibility loci were further examined with HaploRegv2 (http://www.broadinstitute.org/mammals/haploreg/documentation_v2.html), a tool for chromatin profiling, which allows one to explore the impact of non-coding variants, such as those identified in GWASs, by using in silico analysis coupled with experiments to validate biochemical and functional activators and repressors[23,24]. Results show that rs643472 is located in an enhancer region in multiple cell lines and alters three transcription factor motifs, rs963731 at 2p22 alters 18 regulatory motifs, and rs1768208 at 3p22 is in high LD ($R^2 = 0.89$–0.98) with five SNPs that are located in enhancer regions in several brain regions (Supplementary Table 6 and Supplementary Fig. 5). Two of the five SNPs (rs631312 and rs1768190) were further genotyped in CBD ($n = 133$) and controls ($n = 457$), and association testing showed similar results with CBD (OR = 1.89 (CI = 1.30–2.75), $P = 8.77 \times 10^{-4}$; OR = 1.91 (CI = 1.31–2.77), $P = 7.24 \times 10^{-4}$, respectively) to rs1768208 (Table 3).

Expression quantitative trait locus analysis. To further explore potential biological mechanisms, RNA expression quantitative trait loci (eQTL) analysis for the CBD risk loci were interrogated in our expression GWAS (eGWAS)[25] of ~400 samples from cerebellum and temporal cortex from neurodegenerative disorders (Supplementary Table 7) using the Illumina Whole-Genome DASL assay. Expression changes are a possible mechanism of disease-associated functional variants, but other mechanisms also need to be investigated in future studies. The eGWAS data were queried for the top SNPs associated with CBD (rs393152, rs963731, rs1768208, and rs242557). The eGWAS

Table 3 | Association with Mayo CBD cohort and SNPs at *MOBP* predicted to be in regulatory enhancer regions.

SNP	Gene	CBD MAF	Control MAF	LD R^2 rs1768208*	OR (95% CI)	P
rs631312	*MOBP* (Upstream)	0.39	0.25	0.84	1.89 (1.30–2.75)	8.77×10^{-4}
rs1768190	*MOBP* (Intron 1)	0.40	0.26	0.78	1.91 (1.31–2.77)	7.24×10^{-4}
rs1768208	*MOBP* (Intron 2)	0.40	0.29	—	1.89 (1.28–2.77)	1.27×10^{-3}

Mayo CBD cohort is composed of 133 autopsy-proven CBD cases and were tested for association with 457 controls from the replication stage. Association testing was performed using logistic regression under an additive model, with age and sex as covariates.
*Linkage disequilibrium (LD) values from HapMap release 27 with top *MOBP* SNP associated with CBD, rs1768208 (Meta-analysis results, OR = 1.71, $P_{meta} = 2.07 \times 10^{-7}$).

Table 4 | CBD GWAS SNP/transcript associations in cerebellar and temporal cortex human tissue samples.

SNP	Chr	Gene	Probe	Cerebellum		Temporal cortex	
				Beta	P	Beta	P
rs963731	2	*SOS1*	ILMN_1767135	0.154	4.61×10^{-4}	0.178	2.80×10^{-6}
rs1768208	3	*MOBP*	ILMN_2298464	0.363	1.71×10^{-7}	0.325	1.57×10^{-6}
rs1768208	3	*MOBP*	ILMN_2414962	0.243	1.17×10^{-4}	0.180	1.76×10^{-5}
rs1768208	3	*MOBP*	ILMN_1768947	0.143	7.57×10^{-3}	0.107	3.03×10^{-3}
rs1768208	3	*MOBP*	ILMN_1750271	0.074	0.021	0.046	0.016
rs242557	17	*MAPT*	ILMN_1710903	0.183	8.80×10^{-13}	0.181	1.10×10^{-8}
rs242557	17	*MAPT*	ILMN_2298727	0.071	9.78×10^{-3}	0.041	0.11
rs242557	17	*MAPT*	ILMN_2310814	0.002	0.83	−0.026	0.22
rs8070723	17	*MAPT*	ILMN_1710903	−0.473	7.02×10^{-69}	−0.498	8.61×10^{-44}
rs8070723	17	*MAPT*	ILMN_2298727	−0.173	3.36×10^{-7}	−0.137	9.03×10^{-4}
rs8070723	17	*MAPT*	ILMN_2310814	−0.007	0.58	0.018	0.37

Top CBD GWAS SNPs were tested for transcript associations in 374 human cerebellum and 399 temporal cortex samples. Linear regression was employed using an additive model, with the minor allele dosage (0, 1, 2) as the independent variable, and APOE ε4 dosage (0, 1, 2), age at death, gender, PCR plate, RIN, (RIN-RINmean) as biological and technical covariates[14].

genotypes did not contain rs643472 (*KIF13B*/*DUSP4* locus) or any SNPs in LD at $R^2 > 0.5$ with rs643472; and the call rates for the two *DUSP4* expression probes did not pass quality control, precluding eQTL analysis with *DUSP4* expression. SNPs were tested for association with transcript levels in the eGWAS cohort by multivariable linear regression using an additive model, with the minor allele dosage (0, 1 or 2) as the independent variable, and APOE ε4 dosage (0, 1 or 2), age at death, gender, PCR plate, RIN, and RIN-RIN$_{mean}$ as biological and technical covariates. The sample set was analysed as a single cohort for maximum statistical power to detect associations, while including a covariate for AD diagnosis. As we previously reported[25], there were significant SNP/ transcript associations with *MAPT*/rs8070723 (cerebellum: $P = 7.02 \times 10^{-69}$; temporal cortex: $P = 8.61 \times 10^{-44}$) and *MOBP*/rs1768208 (cerebellum: $P = 1.71 \times 10^{-7}$; temporal cortex: 1.57×10^{-6}). We also detected significant associations with *SOS1*/rs963731 (cerebellum: $P = 4.61 \times 10^{-4}$; temporal cortex: $P = 2.80 \times 10^{-6}$) upon eQTL analysis (Table 4). Beta values for SNP/transcript associations showed increased expression levels with each copy of the risk allele.

To determine whether the novel genome-wide significant association at 8p12 (rs643472, an intronic SNP in *lnc-KIF13B-1*) genotype correlates with differential expression levels, we measured *lnc-KIF13B-1* expression in RNA from superior frontal cortex of 22 CBD and 20 normal cases using the mirVana Paris RNA isolation kit and quantitative PCR (qPCR) by SYBR Green. All qPCR reactions were technically replicated. Two different qPCR assays targeting TCONS_00014956 were designed, one targeting exon 1 (primer set A) and the other for exon 2 (primer set B), which showed good correlation upon comparison of expression levels using the two different primer sets ($R^2 = 0.97$). This identified a statistical trend for greater *lnc-KIF13B-1* expression with the minor risk allele (Supplementary Fig. 6).

Discussion

This first CBD GWAS on 152 autopsy-proven CBD patients and 3,311 control individuals has a clear limitation in terms of sample size. Given that CBD is a rare disorder and that it can only be diagnosed upon neuropathologic examination, this was the largest possible cohort at the time of genotyping. To control the false-positive rate due to the small sample size, a stringent significance threshold ($P < 10^{-5}$) was applied in the discovery stage to select SNPs for genotyping in the replication stage. This proved to be a useful approach because the random effects meta-analysis identified two associations that did not replicate (Table 2). The disadvantage of applying stringent criteria is the possibility of false-negative associations, as evidenced from the power analysis, which show low power to detect genome-wide associations of variants with modest effect sizes (Supplementary Tables 8 and 9).

Both CBD and PSP now have confirmed genome-wide significant associations at the *MAPT* locus. In addition, CBD was also associated with rs242557 ($P = 7.91 \times 10^{-6}$), a SNP tagging the H1c subhaplotype at 17q21, shown to be associated with PSP and located in a regulatory region influencing *MAPT* expression[26]. In support of the *MOBP* locus being a shared genetic risk factor between CBD and PSP, CBD cases have an even greater OR estimate compared with PSP (MAF$_{CBD}$ = 0.40; OR = 1.71; MAF$_{PSP}$ = 0.36; OR = 1.39; ref. 17). Importantly, this is the first non-*MAPT* genetic risk factor shared between CBD and PSP. Yamamoto *et al.*[27] described *MOBP* as one of the most abundant oligodendrocyte-expressing proteins, similar to myelin basic protein, with the main difference being that *MOBP* is only expressed in myelin of the central nervous system. The results of the current study and those of the PSP GWAS indicate an overlap in genetic risk factors for PSP and CBD, but it appears that there is additional genetic variation that differs between the two disorders. Additional analyses of the novel susceptibility loci at

chromosomes 8p12 and 2p22 have the potential to advance our understanding of CBD. The chromosome 8p12 locus contains three candidate genes—*lnc-KIF13B-1* (TCONS_00014956, a large intergenic non-coding RNA or lincRNA); kinesin family member 13B (*KIF13B*); and dual-specificity protein phosphatase 4 (*DUSP4*), also known as MAP kinase phosphatase-2 (MKP-2)). Although the *lnc-KIF13B-1* expression levels were not significantly different between CBD and normals, there was a trend for greater expression with the minor risk allele. KIF13B is a plus end-directed microtubule motor protein highly expressed in the brain, and is involved in synaptic vesicle trafficking along microtubules, neurite extension, and caveolin-dependent endocytosis[28–30]. Of interest is that one of the PSP risk variants is in the gene *STX6* and is a SNARE-class protein that regulates vesicle membrane fusion, which raises the possibility that dysfunction of vesicular trafficking may be a common disease mechanism between CBD and PSP.

DUSP4 acts through tyrosine and threonine-directed dephosphorylation[31]. Because tau phosphorylation plays a critical role in tauopathies[32], it will be interesting if *DUSP4* is responsible for the genetic association at 8p12. Indeed, in Alzheimer's disease brain tissue there is decreased enzymatic activity of tau phosphatases such as PP-2A and PP-1 (ref. 33), indicating that a potential role for variants in *DUSP4* in CBD pathogenesis would be through aberrant tau phosphorylation. Furthermore, the chromosome 2p22 locus contains son of sevenless homologue 1 (*SOS1*), a guanine nucleotide exchange factor for Ras that has a catalytic region known as the cdc25 domain due to sequence homology to CDC25, a dual-specificity phosphatase[34]. Taken together, the novel susceptibility loci for CBD may link common genetic variation to aberrant tau phosphorylation.

In conclusion, this first CBD GWAS identified *MAPT* and *MOBP* as shared genetic risk factors between CBD and PSP. Given the significant white matter and oligodendrocyte pathology in these primary tauopathies, the genetic association with *MOBP* warrants functional characterization to determine its role in CBD and PSP.

Methods

Samples. The discovery stage cohort was comprised of 152 autopsy-proven CBD cases collected from eight institutions (Table 1 and Supplementary Table 1). Discovery stage controls were recruited from the Children's Hospital of Philadelphia Health Care Network, and although these controls are not age matched to the cases, the justification for using this cohort is that they were all genotyped at the same center using the same protocol as the CBD cases. Furthermore, because CBD is such a rare disorder with an estimated prevalence of 4.9–7.3 cases per 100,000 individuals[35], the chance of any of the young controls developing CBD at a later age is negligible. As described in the PSP GWAS[17], to ensure significant associations were not confounded by using young controls, allele frequencies for the top associated SNPs were compared and found to be similar to older controls ($N = 3,720$) from three data sets downloaded from the National Institute of Health (NIH) Database for Genotypes and Phenotypes (dbGaP).

DNA was extracted from the brain tissue of CBD patients and from the blood samples of controls. Brain autopsies were obtained after consent of the legal next-of-kin and are considered exempt from human subject research. Written informed consent was obtained from control individuals as approved by The Institutional Review Board (University of Pennsylvania) and Mayo Clinic Institutional Review Board.

Power analyses were estimated for the Discovery Stage and Replication Stage for various minor allele frequencies and odds ratios with 1,000 simulations of each scenario (Supplementary Tables 8 and 9). The power assessment assumes that genotype frequencies are in Hardy–Weinberg equilibrium in CBD cases and control individuals separately and that association testing is performed using logistic regression under an additive model.

Quality control. Quality control of genotyping data was performed at the individual level and then at the SNP level. For quality control, 10 individuals were genotyped in duplicate. Exclusion criteria for individual samples included high genotype failure rate (six individuals were removed because of a genotype failure rate $> 2\%$) and cryptic relatedness. Genetic outliers were excluded from further analyses on the basis of identity by state and distance to the nearest-neighbour

analysis was conducted in PLINK[19] (five CBD cases and 341 controls were removed on the basis of Z-score distributions for the first to the fifth neighbour, Z-score < -2). Gender inconsistencies were assessed by chromosome X genotypes, which excluded six individuals on the basis of observed and expected gender. Exclusion criteria for markers included minor allele frequency (24,183 SNPs were removed because of minor allele frequency $< 1\%$), deviation from Hardy–Weinberg equilibrium in controls (1,098 SNPs gave a Hardy–Weinberg expectations test, P value $\leq 10^{-7}$), and high genotype failure rates (3,199 SNPs were removed because of genotype failure rates $> 2\%$).

Population substructure. MDS was applied to a pruned data set ($\sim 140,000$ markers) using PLINK. Scatter plots for the first two principle components were generated using R (Supplementary Fig. 1a). Analysis of population stratification showed that the first principle component was sufficient to reduce the genomic inflation factor (λ) from 1.06 to 1.01, as illustrated by the quantile–quantile (QQ) plots of observed versus expected P values. Using the first two principal components resulted in $\lambda = 1.02$ (Supplementary Fig. 2b); thus, only the first MDS principal component was used as a covariate for all subsequent analyses.

Association analysis. CBD versus control association was tested by conditional logistic regression under an additive model using the first MDS principle component as a covariate. We also tested for association from a subset of samples from stage 1, excluding outliers based on population substructure (Supplementary Fig. 1b), which yielded similar results. For rs242557, additional association testing was performed, conditioning on *MAPT* haplotype (that is, using 0, 1 or 2 as a covariate for the number of H1 alleles per individual). Linkage disequilibrium values were derived from HapMap 3 release 27.

CBD cases and controls were imputed up to the 1000 Genomes European reference[20], May 2011 release (phase 1, version 3; hg19). SHAPEIT v2.r790 (ref. 21) was used for phasing and IMPUTE2 v2.3.1 (ref. 22) was used for genotype imputation. Imputed variants with MAF < 0.01 or dosage $R^2 < 0.3$ were excluded, and all downstream analyses were based on the most likely genotype[36]. Regional genome-wide association plots were created with LocusZoom[37] (http://csg.sph.umich.edu/locuszoom/). HaploRegv2 (Broad Institute) uses linkage disequilibrium information from 1000 Genomes Project, SNPs (based on dbSNP 137), motif instances (based on PWMs discovered from ENCODE experiments), and enhancer annotations (adding 90 cell types from the Roadmap Epigenome Mapping Consortium)[23,24].

eQTL analysis. These data were generated and previously reported in a brain expression GWAS (eGWAS) by Zou et al.[25], where the detailed mRNA extraction and quality assessment using RNAqueous kit (Ambion, Grand Island, NY) and RNA 6,000 Nano Chip (Agilent, Santa Clara, CA), Whole Genome DASL assay (Illumina, San Diego, CA), and data quality control methods can be found. This eGWAS cohort included 374 cerebellar and 399 temporal cortex mRNA samples from the Mayo Clinic Florida Brain Bank. Neuropathologic diagnoses for the cohort include Alzheimer's disease, PSP, CBD, Lewy body disease, frontotemporal lobar degeneration, and other diagnoses (Supplementary Table 7). The Bonferroni P value threshold for SNP/transcript association significance was $P < 1.56 \times 10^{-3}$ (four SNPs and eight probes).

RNA preparation for *lnc-KIF13B-1* expression studies. Superior frontal cortex tissue was isolated from CBD cases and normal controls from the Mayo Clinic Florida brain bank and matched for age at death and gender. Total RNA from brain tissue samples was extracted with acid–phenol: chloroform chemistry using mir-Vana PARIS isolation kit (Ambion). Total RNA concentration was determined with a NanoDrop ND-1,000 spectrophotometer (NanoDrop, Wilmington, DE). RNA quality control was performed using a 2,100 Bioanalyzer (Agilent, Santa Clara, CA) to measure RNA integrity number (RIN) and only samples with an RNA integrity value > 5 were included in this study. The mean RIN in frontal cortex for 22 CBD cases (5.6 ± 0.38) were not different from 19 controls (6.4 ± 0.77).

Real-time quantitative PCR for lincRNA. Five hundred nanograms of total RNA from the human superior frontal of CBD patients ($n = 22$) and normal controls ($n = 19$) were reverse transcribed with High Capacity cDNA Reverse Transcription kit (Applied Biosystems, Carlsbad, CA). Quantitative PCR (qPCR) was performed using Fast SYBR Green PCR Master Mix (Applied Biosystems) and StepOnePlus (Applied Biosystems) according to the company's suggested protocol and thermal cycling programme. A final dissociation step was performed at the end of the qPCR assay to evaluate the specificity of amplification. All qPCR reactions were technically replicated. Two different qPCR assays targeting TCONS_00014956 were designed, one targeting exon 1 (Primer set A) and the other for exon 2 (Primer set B). All the primers were purchased from Sigma Life Science (Sigma, St Louis, MO) with the sequences shown in Supplementary Fig. 6. The relative levels of lincRNA were calculated by comparative Ct method using StepOne Software v2.3 (Applied Biosystems, Grand Island, NY) and GeneEx 5.3.2 (Multid Analyses, Göteborg, Sweden). GAPDH was used as a normalization control. Statistical analysis was performed using GraphPadPrism v5.04 (GraphPad, La Jolla, CA).

References

1. Dickson, D. W. et al. Office of rare diseases neuropathologic criteria for corticobasal degeneration. J. Neuropathol. Exp. Neurol. 61, 935–946 (2002).
2. Rebeiz, J. J., Kolodny, E. H. & Richardson, Jr. E. P. Corticodentatonigral degeneration with neuronal achromasia. Arch. Neurol. 18, 20–33 (1968).
3. Armstrong, M. J. et al. Criteria for the diagnosis of corticobasal degeneration. Neurology 80, 496–503 (2013).
4. Kouri, N. et al. Neuropathological features of corticobasal degeneration presenting as corticobasal syndrome or Richardson syndrome. Brain 134, 3264–3275 (2011).
5. Kouri, N., Whitwell, J. L., Josephs, K. A., Rademakers, R. & Dickson, D. W. Corticobasal degeneration: a pathologically distinct 4R tauopathy. Nat. Rev. Neurol. 7, 263–272 (2011).
6. Ling, H. et al. Does corticobasal degeneration exist? A clinicopathological re-evaluation. Brain 133, 2045–2057 (2010).
7. Boeve, B. F. et al. Pathologic heterogeneity in clinically diagnosed corticobasal degeneration. Neurology 53, 795–800 (1999).
8. Fujioka, S. et al. A familial form of parkinsonism, dementia, and motor neuron disease: a longitudinal study. Parkinsonism Relat. Disord. 20, 1129–1134 (2014).
9. Fujioka, S. et al. Three sib-pairs of autopsy-confirmed progressive supranuclear palsy. Parkinsonism Relat. Disord. (2014).
10. Kouri, N. et al. Novel mutation in MAPT exon 13 (p.N410H) causes corticobasal degeneration. Acta Neuropathol. 127, 271–282 (2014).
11. Baker, M. et al. Association of an extended haplotype in the tau gene with progressive supranuclear palsy. Hum. Mol. Genet. 8, 711–715 (1999).
12. Ezquerra, M. et al. Identification of a novel polymorphism in the promoter region of the tau gene highly associated to progressive supranuclear palsy in humans. Neurosci. Lett. 275, 183–186 (1999).
13. Higgins, J. J. et al. An extended 5'-tau susceptibility haplotype in progressive supranuclear palsy. Neurology 55, 1364–1367 (2000).
14. Hoenicka, J. et al. The tau gene A0 allele and progressive supranuclear palsy. Neurology 53, 1219–1225 (1999).
15. Di Maria, E. et al. Corticobasal degeneration shares a common genetic background with progressive supranuclear palsy. Ann. Neurol. 47, 374–377 (2000).
16. Houlden, H. et al. Corticobasal degeneration and progressive supranuclear palsy share a common tau haplotype. Neurology 56, 1702–1706 (2001).
17. Hoglinger, G. U. et al. Identification of common variants influencing risk of the tauopathy progressive supranuclear palsy. Nat. Genet. 43, 699–705 (2011).
18. Josephs, K. A. & Dickson, D. W. Diagnostic accuracy of progressive supranuclear palsy in the Society for Progressive Supranuclear Palsy brain bank. Mov. Disord. 18, 1018–1026 (2003).
19. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81, 559–575 (2007).
20. Abecasis, G. R. et al. An integrated map of genetic variation from 1,092 human genomes. Nature 491, 56–65 (2012).
21. Delaneau, O., Howie, B., Cox, A. J., Zagury, J. F. & Marchini, J. Haplotype estimation using sequencing reads. Am. J. Hum. Genet. 93, 687–696 (2013).
22. Howie, B., Marchini, J. & Stephens, M. Genotype imputation with thousands of genomes. G3 (Bethesda) 1, 457–470 (2011).
23. Ernst, J. et al. Mapping and analysis of chromatin state dynamics in nine human cell types. Nature 473, 43–49 (2011).
24. Ward, L. D. & Kellis, M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res. 40, D930–D934 (2012).
25. Zou, F. et al. Brain expression genome-wide association study (eGWAS) identifies human disease-associated variants. PLoS Genet. 8, e1002707 (2012).
26. Rademakers, R. et al. High-density SNP haplotyping suggests altered regulation of tau gene expression in progressive supranuclear palsy. Hum. Mol. Genet. 14, 3281–3292 (2005).
27. Yamamoto, Y. et al. Cloning and expression of myelin-associated oligodendrocytic basic protein. A novel basic protein constituting the central nervous system myelin. J. Biol. Chem. 269, 31725–31730 (1994).
28. Hanada, T., Lin, L., Tibaldi, E. V., Reinherz, E. L. & Chishti, A. H. GAKIN, a novel kinesin-like protein associates with the human homologue of the Drosophila discs large tumor suppressor in T lymphocytes. J. Biol. Chem. 275, 28774–28784 (2000).
29. Horiguchi, K., Hanada, T., Fukui, Y. & Chishti, A. H. Transport of PIP3 by GAKIN, a kinesin-3 family protein, regulates neuronal cell polarity. J. Cell Biol. 174, 425–436 (2006).
30. Kanai, Y., Wang, D. & Hirokawa, N. KIF13B enhances the endocytosis of LRP1 by recruiting LRP1 to caveolae. J. Cell Biol. 204, 395–408 (2014).
31. Misra-Press, A., Rim, C. S., Yao, H., Roberson, M. S. & Stork, P. J. A novel mitogen-activated protein kinase phosphatase. Structure, expression, and regulation. J. Biol. Chem. 270, 14587–14596 (1995).
32. Forman, M. S. et al. Signature tau neuropathology in gray and white matter of corticobasal degeneration. Am. J. Pathol. 160, 2045–2053 (2002).
33. Gong, C. X., Singh, T. J., Grundke-Iqbal, I. & Iqbal, K. Phosphoprotein phosphatase activities in Alzheimer disease brain. J. Neurochem. 61, 921–927 (1993).
34. Margarit, S. M. et al. Structural evidence for feedback activation by Ras.GTP of the Ras-specific nucleotide exchange factor SOS. Cell 112, 685–695 (2003).
35. Togasaki, D. M. & Tanner, C. M. in Corticobasal Degeneration. (eds Litvan, I., Goetz, C. G. & Lang, A. E.) 53–59 (Lippincott, Williams & Wilkins, 2000).
36. Browning, B. L. & Browning, S. R. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. Am. J. Hum. Genet. 84, 210–223 (2009).
37. Pruim, R. J. et al. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics 26, 2336–2337 (2010).

Author contributions

Co-first authors N.K. and O.A.R. and senior authors G.D.S. and D.W.D. contributed equally to this project. N.K., G.D.S. and D.W.D. initiated this study and were involved in all the aspects of data analysis and manuscript preparation. N.K. conducted the analysis, wrote the manuscript, and prepared RNA samples for both eQTL and lincRNA expression studies. O.A.R. and A.S.-O. generated the replication stage data. B.D. organized the CBD DNA bank, and sample distribution for genotyping and phenotype distribution for analysis. C.S.Y., N.E.-T. and S.G.Y. provided the eQTL data and analysis. D.J.S. performed the genotype imputation analysis. N.K., O.A.R., M.B., N.A.F. and A.S.-O. participated in Mayo Clinic sample preparation. H.Y. and J.K. performed lincRNA expression studies. L.B.C. coordinated sample acquisition and selection. C.A.M., B.G., S.S., M.R.H., S.B., E.T.G., H.A.K., S.R., M.G., J.L.J., J.P.G.V., V.M.V.D., M.G., H.I.H., R.G.G., S.E.A., J.Q.T., V.M.L., G.K.W., C.L.W., G.U.H., U.M., B.F.B., K.A.J., J.E.P., Z.K.W., R.J.U., N.R.G.R., I.L., R.R., G.D.S. and D.W.D. participated in characterization, preparation, and contribution of CBD cases used in this study. B.D. performed the initial association analysis and contributed to discussions of the manuscript. O.A.R., L.I.G., J.C., R.R. and N.E.T. participated in data analysis discussions and reviewed the manuscript. L.-S.W. participated in the initial association analysis. H.H. supervised genotyping as well as platform and sample selection.

Additional information

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Kouri, N. et al. Genome-wide association study of corticobasal degeneration identifies risk variants shared with progressive supranuclear palsy. Nat. Commun. 6:7247 doi: 10.1038/ncomms8247 (2015).