Online Methods

Sample. The Stage 1 sample consisted of 1,150 autopsy-confirmed cases and 3,658 population-based controls. Patients with a clinical Parkinsonism in life and neuropathological confirmation of PSP were identified from brain banks, research hospitals and neuropathologists (Supplementary Table 1). The neuropathological diagnosis was made according to NINDS diagnostic criteria[39,40]. DNA was extracted from brain tissue from patients who had consented for brain donation.

The Stage 2 sample consisted of 1,150 patients diagnosed with PSP and 3,654 control subjects, none included in Stage 1. The majority of cases were clinically diagnosed (1019), although some were autopsied (131). Clinical cases were recruited through specialized movement disorder hospitals and clinics and referrals from neurologists (Supplementary Table

4). All patients fulfilled the clinical diagnosis of possible or probable PSP[17]. The diagnostic assessments involved review of full medical history, neurological examinations, and mental status evaluations. The demographic data included gender, age at onset of motor symptoms, disease duration, and final clinical diagnosis. Informed consent was obtained for blood collection and genetic analysis. The University of Pennsylvania assembled all DNA samples identified for the study.

Controls for Stage 1 and 2 were young subjects recruited from the Children's Hospital of Philadelphia (CHOP) Health Care Network by CHOP clinicians and nursing staff. Written informed consent was obtained from all subjects. The Stage 1 cohort was largely of European ancestry (89.5%) whereas all Stage 2 controls were of European ancestry. Females comprised 47.7 and 47.0% of the Stage 1 and 2 samples; mean ages for these cohorts were 7.8 and 8.8. The advantage of using these controls is that all were genotyped at the same center using the same protocols as the cases. Although the controls were not selected for absence of neurodegenerative disease, the low population frequency of PSP ensures a negligible number of controls will get PSP later in life. The disadvantage of young controls is that some loci could also have a positive or negative impact on survival. To diminish this disadvantage, for SNPs significantly- or suggestively-associated with PSP, we compared control allele frequencies of our younger samples with those from a set of older controls, N=3,816, obtained from the NIH repository Database for Genotypes and Phenotypes. Only SNPs for which there is no difference between young and old frequencies were reported in Table 2.

Genotyping. Stage 1 cases were genotyped by the Center for Applied Genomics at CHOP using Human 660W-Quad Infinium Beadchips. Control samples were genotyped using the Illumina Human HapMap550 Infinium BeadChip. Stage 2 cases were genotyped for 5,283 SNPs using Infinium HD iSelect Custom BeadChip, most of which were identified by CHOP or University of Pittsburgh at Stage 1 as SNPs showing association by the log-additive model at a P-value ≤ 0.001. Ancestry Informative Markers or AIMs were also included. Stage 2 controls were selected from a larger CHOP control dataset to match cases in terms of genetic ancestry.

Quality Control. Quality Control (QC) procedures were performed at the individual and then SNP level. At the individual level, gender miscalls based on chromosome X and Y genotypes, duplicate samples, and highly-related samples were excluded. At the SNP level, a genotype completion rate of $\geq 98\%$ was required. Hardy-Weinberg Equilibrium (HWE) was evaluated in samples of European ancestry and SNPs failing HWE were excluded ($P < 1 \times 10^{-4}$). To control potential confounding due to variation in genetic ancestry, cases and controls were matched for ancestry based on genetic data using methods described in Lee *et al.* [41] and Crossett *et al.*[42]. All QC and subsequent association analyses were performed independently at CHOP by P.M.S. and University of Pittsburgh by N.M.M and L.K. Results were then compared for agreement. Analytic methods and results described are those used at the University of Pittsburgh unless otherwise noted.

For Stage 1 there were 3 gender inconsistencies (1 PSP and 2 controls); 32 samples were duplicates (all PSP); and 12 samples had genotyping completion rate < 98%. These samples were eliminated. For SNPs, 4,222 were monomorphic, had uncalled genotypes or had genotyping

completion rate <98%. Ancestry for cases and controls were determined by using 6,490 SNPs with a call rate > 99.9% and with an inter-SNP distance more than 0.4Mb. SpectralGEM[41] estimated 8 significant dimensions of ancestry (Supplementary Fig. 1). Subsequent clustering on these dimensions resulted in 11 clusters: 4 of European ancestry, with N = 691, 682, 939 and 1,116 individuals. The other clusters reflected other major ancestral groups, largely African and African American ancestry. Using the samples of European ancestry to determine Minor Allele Frequency MAF and evaluate HWE, 24,132 SNPs were removed because MAF < 0.01 and 1919 failed HWE.

As determined by genotypes, all cases and controls for Stage 2 were independent of those used in Stage 1. Of the 5,283 genotyped SNPs, 4617 could be aligned with control genotypes: most (3,203) were identified by CHOP's or University of Pittsburgh's Stage 1 analyses as SNPs showing log-additive association at $P_1 \leq 0.001$, as well as a subset of SNPs correlated with the original associated SNPs; Ancestry Informative Markers or AIMs markers (197)[43]; and SNPs identified by exploratory modeling to identify recessive, and dominant or interactive risk loci (896). Sixty control samples were genotyped on both BeadChip and iSelect platforms to assess concordance, which was 100%.

Controls were selected from a larger CHOP control dataset to match cases in terms of genetic ancestry (by P.M.S.). To determine ancestry, multilocus genotype data were decomposed by eigenvector analysis as implemented in Eigensoft[44] into a small set of vectors representing ancestry using smartpca, a part of the EIGENSTRAT package. To do so a trimmed set of 3579 SNP was selected from those on the iSelect panel, excluding all SNPs in LD (pairwise r2 < 0.5)

and all SNPs in the 17q12 inversion. After eigenvector analysis, cases were 1:3 matched to controls using 3 principal components and a distance threshold of 0.071[45].

Association Analysis. At Stage 1 we contrasted the genotypes at 531,451 single nucleotide polymorphisms (SNPs) of PSP subjects and genetically-matched controls *via* conditional logistic regression and a log-additive model. Full matching[45] of cases and controls resulted in 1114 strata, each of which contained one case and one or more controls. To analyze chromosome X, the data were matched conditional on gender. After removing individuals not in close proximity to a discordant (in diagnosis) individual of the same gender, there were 628 PSP cases and 1,686 controls in the male-only set and 485 PSP cases and 1,441 controls in the female-only set for a total of 1,113 cases and 3,127 controls. Cases were full matched in 1,113 distinct strata, each of which contained one case and one or more multiple controls of the same gender. For the analysis of subjects of European ancestry, data consisted of 1,069 PSP cases and 2,964 controls.

Primary analyses were conducted among subjects of European ancestry; the resulting Q-Q plot is Supplementary Fig. 3. Genotypes for any SNPs showing GWA-significant association, or nearly so, were manually inspected for valid genotype clustering. SNPs showing poor clustering were excluded. Exploratory analyses included dominant and recessive models, as well as evaluation of gene-gene interaction by using a model selection procedure called "Screen and Clean"[38]. Analyses using subjects of any ancestry were also conducted and reported in Supplementary Table 2.

At Stage 2, association of SNP genotype and diagnostic status was assessed for 4,099 SNPs remaining after quality control, again by using conditional logistic regression. After the QC analysis described above, 1,051 PSP cases and 3,560 controls remained for the association analysis, all of European ancestry.

Study design and statistical significance. Following Skol *et al.*[14], markers considered genome-wide significant were those that meet the threshold for GWA-significance from the joint analysis of Stages 1 and 2. We take $P \leq 5x10^{-8}$ (joint $z \geq 5.44$) as genome-wide significant and $5.7 \times 10^{-7} \geq P > 5 \times 10^{-8}$ as strongly suggestive (joint z = 5.0 for larger bound).

Effect of diagnostic misclassification on power. Samples from Stage 1 consist of autopsy-confirmed subjects whereas a large fraction of Stage 2 was diagnosed clinically. For Stage 2, then, a fraction 1 - π of subjects could be misdiagnosed as PSP when they have some other diseases, often Parkinson Disease (PD). Here we assume all misdiagnoses are PD, whereas in truth it will be a mixture of diseases; most, like PD, are unlikely to share risk loci with PSP. For this reason for most SNPs, denoted by *j*, the expected odds ratio OR = 1 and $\theta = \log(OR_{PDj})$, which is assumed to be normally distributed with mean 0 and variance $\sigma_{PD}^{\ 2}$, the usual variance of the log odds ratio.

The distribution of the data in Stage 2 is a normal mixture. As noted above, typically θ_{PDj} is 0; however, the inversion polymorphism of the *MAPT* locus is known to be a risk factor for PSP and PD, with an estimated odds ratio of 5.6 for PSP, Stage 1 ($\theta_P = 1.72$), 4.7 for PSP, Stage 2 ($\theta_P = 1.56$), and 1.32 ($\theta_{PD} = 0.28$) for PD[12]. Of note, the expected value of θ at

Stage 2 for any locus j is $\pi\theta_{Pj} + (1-\pi)\theta_{PDj}$. This observation gives us a way to estimate the fraction of misdiagnosis 1 - π by using the estimated parameters for the MAPT locus and the method of moments. Plugging in estimates of log(OR) and solving for 1 - π, our estimate is 0.12, or 12% of the Stage 2 sample is likely to be PD. Note that if the misclassification were not all PD, it would likely diminish the estimated fraction of subjects misclassified.

We can also use this result to ask what impact 12% misclassification would have on the OR at Stage 2. For a true OR = 1.4 for a pure PSP sample, the adjusted odds ratio – assuming the locus has no impact on risk for PD or other misclassified diagnoses – would be 1.34; for OR = 1.3, it would be 1.26. Because we are more concerned about the OR from the joint analysis of Stages 1 and 2, an OR = 1.4 is expected to be reduced to 1.37 in the joint analysis and OR = 1.3 to 1.28 by 12% misclassification. Applying the CaTS power calculator[14] for an alpha of 5×10^{-8} and an OR reduction from 1.4 to 1.37, the power for the joint analysis is reduced from 97% to 92%; reducing OR from1.3 to 1.28, the power is reduced from 61% to 48%. Thus misclassification does not have a large impact on power in this setting.

Expression Quantitative Trait Loci (eQTL) Analysis. The sample for eQTL analysis was composed of 387 neurologically normal Caucasian subjects with no clinical history of neurological or cerebrovascular disease or a diagnosis of cognitive impairment (249 brains from the University of Maryland Brain Bank; 36 from Johns Hopkins University Brain Bank; 25 from the Baltimore Longitudinal Study on Aging; 24 from the University of Miami Brain Bank; and 78 from the Sun Health Brain Bank). All individuals were of non-Hispanic European ancestry.

RNA from the cerebellum (CRBLM) and frontal cortex (FCTX) was obtained and expression measured using Illumina HumanHT-12 v3 Expression Beadchips. DNA was genotyped using the Illumina genotyping arrays (HumanHap550 v3, Human610-Quad v1 or Human660W-Quad v1). Processing and analysis of these data were performed in a manner similar to methods previous described for a study including 150 of these 387 subject[46]. In brief, expression profiles were adjusted for these covariates: subject age, gender, post-mortem interval, originating tissue bank, principal components 1 and 2 based on identity-by-state pairwise genetically-estimated distances between subjects[47], and the mRNA sample preparation and hybridization batch. eQTL analysis was then performed using linear regression of the residuals for every trait in each brain tissue with the allele dosage. For each trait analyzed, only SNPs that are *cis* to the trait were considered in the analysis, where *cis* is the region near the gene, ± 1Mb from the mRNA transcript start or end site. SNPs associated with PSP by joint analysis were then evaluated for their impact on expression of genes proximate to the SNPs.