Genetic determinants of survival in progressive supranuclear palsy: a genome-wide association study

Jabbari et al.

Supplementary Material

Supplementary Table 1: Stage one GWAS – PSP sample sources

Sample source	Number of samples	% of samples with post-mortem confirmation of PSP pathology	Local research ethics committee study approval code
UCL cohort			
Queen Square Brain Bank, UK	249	100	IONMTA20-16
PROSPECT study, UK	180	11*	14/LO/1575
Cambridge Brain Bank, UK	54	100	10/H0308/56
Johns Hopkins Brain Bank, USA	37	100	NA_00032761
MRC London Neurodegenerative Diseases Brain Bank, UK	25	100	18/WA/0206
Multiple Sclerosis and Parkinson's Brain Bank, UK	21	100	18/WA/0238
Newcastle Brain Bank, UK	12	100	08/H0906/136+5
2011 GWAS cohort			
Mayo Clinic Brain Bank, USA	371	100	15-009452
Munich Brain Bank, Germany	84	100	09/07

32 cases (Mayo Clinic Brain Bank, n=31; Queen Square Brain Bank, n=1) were excluded due to either insufficient clinical data or failing genotype data quality control. *=74/180 PROSPECT study cases were deceased at the date of censoring. 20/74 had a post-mortem examination, of which 20/20 had a pathological diagnosis of PSP, therefore 20/180=11%.

Supplementary Methods: Genotype data and imputation methods

In stage one, all brain bank cases had DNA extracted from frozen brain tissue (cerebellum or frontal cortex). PROSPECT study cases had DNA extracted from whole blood. Subsequently, DNA samples from all cases underwent genotyping using the Illumina NeuroChip (UCL cohort cases) or the Illumina Human 660W-Quad Infinium chip (2011 GWAS cohort cases).[1,2] In stage two, genomic DNA was isolated from frozen cerebellar brain tissue from PSP cases using Autogen 245T (Holliston, MA, USA) methods. Lead SNPs of all genome-wide significant loci from the stage one GWAS were genotyped using a TaqMan Allelic Discrimination Assay (assay ID: ANH6JHT) on a QuantStudio 7 Flex Real-Time PCR System (Applied Biosystems, Foster City, CA, USA), and genotypes were called using QuantStudio Real-Time PCR Software (version 1.1; Applied Biosystems).

Standard genotype data quality control steps were applied to remove cases with low overall genotyping rates (<95%), biological and genetic sex mismatch, heterozygosity outliers (>2SDs away from the mean) and related individuals (Identity-By-Descent PIHAT >0.1). Additionally, a principal components analysis (PCA), merged with the European (CEU) HapMap reference dataset, was carried out and all cases of non-white European ancestry were excluded to minimise the impact of population stratification.[3] In stage 1, all cases were screened for known *MAPT, LRRK2 and DCTN1* mutations covered by both genotyping platforms and were excluded if positive as monogenic forms of PSP are genetically distinct from sporadic PSP. The 2011 GWAS and UCL cohort datasets were then merged and approximately 230K SNPs which were present on both genotyping platforms were extracted for SNP imputation against the Haplotype Reference Consortium v1.1 panel using the Sanger Imputation Service (https://imputation.sanger.ac.uk/). The following post-imputation data quality control steps were used to filter out SNPs with: Imputation

information score <0.70, posterior probability <0.90, genetic missingness >0.05, Hardy-Weinberg equilibrium p-value >$1.0x10^{-7}$ and minor allele frequency <0.01. Imputed SNP positions were based on Genome Reference Consortium Human 37/human genome version 19 (GRCh37/hg19) (https://www.ncbi.nlm.nih.gov/grc/human/data?asm=GRCh37).

Supplementary Methods: Coloc

Coloc incorporates a Bayesian statistical framework that computes the posterior probability (PP) of five hypotheses: there is no association with either trait (hypothesis 0, H_0); there is an association with trait 1 only (H_1); there is association with trait 2 only (H_2); there is an association with both traits, but the variants involved are independent (H_3); and finally, there is a shared variant associated with both traits within the analysed region (H_4). We ran coloc using default parameters and priors (p1 = 1×10^{-4}; p2 =1×10^{-4}; p12 = 1×10^{-5}), and restricted analyses to genes within 1Mb of a significant region of interest, as defined by PSP survival $p < 5 \times 10^{-8}$). We excluded all loci in which the summed posterior probabilities of hypothesis H_3 and H_4 (PP3 + PP4) were < 0.8, to exclude loci where we were underpowered to detect colocalisation.

Supplementary Methods: Whole-genome sequencing

Blood-derived DNA samples were normalized with Low-EDTA TE buffer (Quality Biological). PCR-free, paired-end, indexed libraries were constructed by automated liquid handlers (Microlab STAR, Hamilton) using 1µg DNA input (TruSeq DNA PCR-free, Illumina). DNA libraries were diluted to 2.2nM, denatured and clustered using the cBot2 system (Illumina) and following a 'single library' - 'single lane' methodology on patterned flow cells. The

libraries were sequenced on an Illumina HiSeq X Ten sequencer (v2.5 chemistry, Illumina) using 150 bp, paired-end cycles.

Raw genome data in FASTQ file format were transferred to Google Cloud Storage. Paired-end sequences were processed in accordance with the pipeline standard developed by the Centers for Common Disease Genomics (CCDG; https://www.genome.gov/27563570/). Sequence alignment to the human reference genome (GRCh38) was achieved using the Burrows-Wheeler Aligner (BWA-MEM)[4] and was followed by post-alignment processing for variant discovery using the Genome Analysis Toolkit Best Practices pipeline (GATK v4.1.2.0).[5] The average sequencing read-depth after filtering by alignment quality was 35x. Single nucleotide and short indel variants were called from the processed WGS data using GATK HaplotypeCaller to generate individual gVCF files for each subject and followed by joint genotype calling of merged gVCFs to generate a multi-sample VCF file. Standard variant filtering was then performed using GATK4 Variant Quality Score Recalibration (VQSR) tools VariantRecalibrator and ApplyRecalibration.

For sample-level quality control checks, we first excluded genomes with a high contamination rate (>5% based on VerifyBamID freemix metric).[6] Next, we converted the multi-sample VCF-file to a PLINK 2.0 binary file (cog-genomics.org/plink/2.0)[7] and removed samples with an excessive heterozygosity rate (exceeding +/- 0.15 F-statistic), a sample call rate ≤95%, discordance between reported sex and genotypic sex, and duplicate samples (determined by pi-hat statistic). For variant-level quality control, we excluded variants with an overall missingness rate of >5%, non-autosomal variants and variants that significantly departed from Hardy-Weinberg equilibrium ($p<1.0x10^{-10}$).

Supplementary Table 2: Survival analysis of known PSP risk variants

Chromosome	Postion (BP) – GRCh37	rsID	Nearest gene	PSP survival GWAS hazard ratio (95% confidence interval)	PSP survival GWAS p-value
1	180962282	rs1411478	*STX6*	NA	NA
1	222168434	rs12125383	*DUSP10*	0.98 (0.88-1.11)	0.80
2	88895351	rs7571871	*EIF2AK3*	NA	NA
3	39523003	rs1768208	*MOBP*	0.90 (0.81-1.01)	0.06
6	396321	rs12203592	*IRF4*	1.01 (0.89-1.16)	0.83
6	45495748	rs4438954	*RUNX2*	1.06 (0.94-1.19)	0.37
12	21467215	rs7966334	*SLCO1A2*	1.07 (0.90-1.28)	0.44
17	44019712	rs242557	*MAPT*	0.96 (0.87-1.06)	0.46
17	44081064	rs8070723	*MAPT*	0.80 (0.69-1.02)	0.06

Variants that reached genome-wide significance (in previous PSP case-control GWAS were analysed[8–10]

NA = SNP not present in dataset

Supplementary Table 3: Sub-genome-wide significant signals ($p<5.0x10^{-5}$)

Chromosome	Position (BP) - GRCh37	rsID	Alternate allele	Overlapping gene	Nearest upstream gene	Nearest downstream gene	Coefficient	p-value
1	19323429	rs76602886	T		*IFFO2*	*RP5-1126H10.2*	0.79	3.29E-05
1	38919851	rs113207733	C		*RNU6-753P*	*RP11-329N22.1*	-0.54	1.19E-05
1	58381275	rs114931168	T	*DAB1*			0.70	4.84E-05
1	84263338	rs12144651	C	*RP11-475O6.1*			0.83	1.29E-05
1	96299183	rs144145738	T		*RP11-286B14.1*	*AC092812.1*	0.96	1.26E-05
1	120134108	rs113741465	C		*HSD3BP4*	*GAPDHP33*	0.98	4.69E-05
1	159355851	rs144593458	T	*RP11-550P17.5*			1.01	1.19E-05
1	169592511	rs3917676	C	*SELP*			-0.70	1.46E-05
1	182441275	rs184150470	T	*RGSL1*			0.78	1.91E-05
1	183768258	rs2500095	A	*RGL1*			0.20	3.15E-05
1	187431475	rs80263067	G	*LINC01037*			0.89	2.51E-05
1	208598450	rs696964	G		*RP11-2P2.1*	*RP11-565N2.1*	0.28	3.98E-05
1	228367637	rs138488044	G	*IBA57*			0.72	2.62E-05
1	233668414	rs78571397	T		*RNU4-77P*	*KCNK1*	1.20	4.05E-07
1	234907019	rs35550581	G		*LINC01132*	*RNY4P16*	-0.21	2.40E-05
1	246819159	rs78053612	A	*CNST*			-1.00	6.80E-06
2	28654232	rs13390537	T		*RP11-373D23.3*	*AC104695.4*	0.98	4.79E-05
2	76339690	rs13016424	A		*SUCLA2P2*	*AC073091.2*	0.21	4.48E-05
2	85479989	rs112561287	C	*TCF7L1*			0.83	7.42E-06
2	122907120	rs1864825	A		*AC018737.3*	*AC062020.2*	0.22	1.24E-05
2	143010934	rs2890656	G		*AC078882.1*	*AC016706.1*	0.40	4.28E-05
2	150236492	rs150909501	T	*LYPD6*			-0.92	7.08E-06
2	171025057	rs12470325	A		*RNU6-1006P*	*MYO3B*	1.07	1.79E-06
2	202098131	rs17860416	A		*CASP10*	*CASP8*	0.63	3.41E-05
2	207277125	rs151053679	A		*AC017081.2*	*HNRNPA1P51*	0.52	1.10E-05
3	1138817	rs76590505	T	*CNTN6*			0.88	4.03E-05
3	71068448	rs9843058	G	*FOXP1*			0.52	2.28E-05
3	104597805	rs114013231	G		*RP11-281P11.1*	*ALCAM*	0.77	1.66E-05
3	111877796	rs73218079	T	*SLC9C1*			0.26	2.76E-06
3	129675760	rs66817437	T	*RP11-93K22.6*			0.57	3.04E-05
3	148826658	rs113126478	T		*HLTF-AS1*	*HPS3*	-0.67	3.51E-05
4	30834962	rs1447372	C	*PCDH7*			0.79	1.84E-05
4	79400404	rs72659024	T	*FRAS1*			1.01	1.94E-05
4	90618694	rs115873866	G	*RP11-115D19.1*			1.14	1.73E-07
4	102372561	rs139808768	G	*BANK1*			0.88	1.79E-05
4	105837534	rs34162290	A	*RP11-556I14.2*			0.62	4.47E-05
4	172370813	rs76038806	T		*RP11-717H13.1*	*RP11-97E7.1*	0.61	2.33E-06
5	3309613	rs4499838	C		*CTD-2029E14.1*	*LINC01019*	-0.21	1.72E-05
5	6877162	rs59652007	C	*RP11-332J15.2*			0.68	1.01E-06

Chromosome	Position (BP) - GRCh37	rsID	Alternate allele	Overlapping gene	Nearest upstream gene	Nearest downstream gene	Coefficient	p-value
5	8789437	rs10491225	G		*RP11-315A16.1*	*RP11-143A12.3*	0.74	4.24E-05
5	60861122	rs10042706	G		*AC008836.1*	*C5orf64*	0.34	4.66E-05
5	66993673	rs10062637	G	*RP11-434D9.1*			-0.43	1.67E-05
5	120211759	rs62381131	T		*CTD-2334D19.1*	*AC008565.1*	-0.22	5.56E-06
5	137769795	rs62382361	T	*KDM3B*			0.69	1.44E-05
5	163093198	rs158832	A		*MAT2B*	*RP11-756G12.1*	0.34	3.20E-06
5	172622516	rs10039302	A		*BNIP1*	*RPL7AP33*	0.26	3.12E-05
5	175671655	rs4242199	G	*SIMC1*			-0.24	4.57E-05
6	21311711	rs62406284	T		*CDKAL1*	*RP1-135L22.1*	0.69	3.88E-05
6	25346742	rs2690097	C	*LRRC16A*			0.24	1.02E-05
6	33382955	rs116199362	C	*PHF1*			0.99	2.54E-05
6	51707176	rs115011904	T	*PKHD1*			0.73	1.87E-05
6	63025542	rs139900383	C		*KHDRBS2*	*RP11-448N11.1*	-0.81	2.87E-05
6	68633843	rs191553496	T		*RP11-301G19.1*	*RP11-301G19.2*	0.88	3.72E-05
6	80540484	rs117259176	T		*AL132875.1*	*RP1-92C4.1*	0.85	1.66E-06
6	85116234	rs10455429	C		*KIAA1009*	*RP1-90L14.1*	0.88	1.86E-05
6	99177471	rs79567149	T		*RP3-453D15.1*	*RP11-67P15.1*	0.32	1.73E-05
6	106616113	rs11751098	T		*PRDM1*	*ATG5*	0.79	2.75E-05
6	123533660	rs9490702	T		*CLVS2*	*TRDN*	-0.52	3.20E-05
6	129931872	rs145903948	C	*ARHGAP18*			0.68	4.88E-05
7	4497566	rs74797028	T		*SDK1*	*CYP3A54P*	0.55	4.70E-06
7	21533758	rs7810841	C	*SP4*			0.23	1.04E-05
7	23192511	rs116922788	A	*KLHL7*			0.53	3.51E-05
7	23287271	rs117207752	A	*GPNMB*			0.94	1.30E-05
7	51699456	rs13236400	G		*RP4-718N17.2*	*RN7SL292P*	-0.21	3.01E-05
7	115176793	rs118151025	A		*RP11-222O23.1*	*SNORA25*	0.89	5.00E-05
7	132739501	rs117588838	C	*CHCHD3*			0.98	1.86E-06
7	146193963	rs35652175	T	*CNTNAP2*			-0.44	2.90E-06
8	295524	rs73173363	T		*RP11-63E5.6*	*FAM87A*	0.79	1.93E-05
8	12866910	rs17123337	A	*KIAA1456*			0.48	3.31E-05
8	26703750	rs523816	T	*ADRA1A*			0.21	4.24E-05
8	40566874	rs12234987	A	*ZMAT4*			0.49	4.14E-05
8	102813956	rs56074625	G	*NCALD*			0.76	4.24E-05
8	109497271	rs79000385	A	*EMC2*			0.95	4.28E-05
8	124549920	rs117766100	T	*FBXO32*			0.41	2.53E-05
8	131441098	rs75476259	C	*ASAP1*			0.93	3.34E-05
9	19811369	rs7864219	C		*SLC24A2*	*AL158077.1*	0.38	4.83E-05
9	81930302	rs10780299	T		*RP11-165H23.1*	*CHCHD2P9*	-0.23	2.76E-05
9	138993315	rs62583536	G		*NACC2*	*RP11-83N9.5*	0.94	2.33E-06
10	1094398	rs7089968	A	*IDI1*			0.77	4.08E-05
10	6185697	rs80141934	C		*RBM17*	*PFKFB3*	0.89	4.39E-05
10	9134927	rs117914243	A		*RP11-428L9.2*	*LINC00709*	0.96	2.91E-05
10	58341064	rs1948426	G		*ZWINT*	*SNORD2*	0.43	4.23E-06
10	66492356	rs1879599	T		*RPL17P35*	*CYP2C61P*	0.45	1.56E-05
10	118591882	rs145575061	C		*RPL5P27*	*RP11-539I5.1*	0.87	1.65E-05
10	132868460	rs2894946	G		*AL607076.1*	*TCERG1L*	-0.21	3.46E-05

Chromosome	Position (BP) - GRCh37	rsID	Alternate allele	Overlapping gene	Nearest upstream gene	Nearest downstream gene	Coefficient	p-value
11	5767782	rs7129530	A	*TRIM5*			1.10	1.14E-05
11	6287444	rs11040828	G	*CCKBR*			0.28	1.09E-05
11	18697993	rs58225181	A		*RP11-137N23.1*	*TMEM86A*	-0.46	1.03E-05
11	34857834	rs731727	G		*RP4-743O11.1*	*APIP*	0.24	1.15E-05
11	38326383	rs6484991	G		*RP11-436H16.1*	*RP11-63D14.1*	1.02	1.68E-05
11	65054542	rs481574	G	*POLA2*			-0.40	1.04E-05
11	131086565	rs1793790	G		*AP002856.6*	*AP002856.7*	0.35	8.06E-07
12	40413698	rs2242367	A	*SLC2A13*			0.35	7.46E-10
12	43802292	rs12830181	A	*ADAMTS20*			0.58	9.68E-06
12	105851512	rs113750212	A		*C12orf75*	*CASC18*	0.92	1.05E-05
13	53492963	rs117809677	G		*PCDH8*	*OLFM4*	0.52	1.98E-05
13	77416712	rs4885430	T		*RN7SL571P*	*KCTD12*	-0.42	4.01E-06
13	91651369	rs9515869	T		*LINC00410*	*BRK1P2*	0.22	2.81E-05
13	100233261	rs1417795	A		*LINC01039*	*CFL1P8*	-0.21	4.45E-05
13	103578544	rs9300768	T		*METTL21EP*	*SLC10A2*	0.60	2.98E-05
14	57640614	rs1892219	C		*OTX2-AS1*	*EXOC5*	-0.67	4.58E-05
14	59556752	rs78269377	A		*RP11-112J1.2*	*CTD-2315A10.2*	0.77	2.03E-06
14	59641184	rs148461037	A		*CTD-2315A10.2*	*CTD-2315A10.1*	1.04	9.49E-06
14	84470423	rs17654303	A		*RP11-353P15.1*	*CTD-2320B12.3*	0.69	3.49E-06
14	89778180	rs12888659	T	*FOXN3*			-0.26	2.46E-06
14	93467174	rs4905027	G	*ITPK1*			0.22	1.01E-05
14	95367332	rs12147572	T		*RPL15P2*	*RP11-991C1.1*	0.35	4.36E-05
15	47078862	rs7162159	G	*RP11-627D16.1*			0.24	1.71E-06
15	58694020	rs16940212	T	*ALDH1A2*			0.29	1.82E-05
15	60414139	rs75370883	C		*RP11-82L7.4*	*ANXA2*	0.63	8.76E-06
15	87692992	rs11856418	C		*RP11-138H10.2*	*RP11-648K4.2*	0.34	7.90E-06
15	89920492	rs8032933	T	*LINC00925*			1.09	2.29E-07
16	9319706	rs140558360	A		*RPL21P119*	*AC092122.1*	0.91	2.96E-06
16	11467849	rs76351336	T	*CTD-3088G3.8*			0.99	2.93E-08
16	50980482	rs75243598	A		*RP11-883G14.1*	*RP11-883G14.2*	-0.97	4.57E-05
16	52313507	rs1894879	A	*RP11-142G1.2*			-0.24	4.85E-06
16	65604741	rs141366655	A	*LINC00922*			0.89	4.81E-05
16	70739949	rs78409530	G	*VAC14*			0.51	1.82E-05
17	14718726	rs72813667	C		*AC005863.1*	*AC005863.2*	-0.50	3.68E-05
17	20880491	rs75883746	G	*RP11-344E13.3*			0.95	2.36E-05
19	29839509	rs199747997	G	*CTC-525D6.1*			1.01	1.68E-05
20	16137976	rs34913727	T		*MACROD2*	*PPIAP17*	0.96	7.69E-07
20	33599090	rs6120790	T	*TRPC4AP*			-0.26	8.16E-06
21	35964181	rs117012899	C	*RCAN1*			0.71	4.55E-05
22	43397109	rs9611982	G	*PACSIN2*			-0.75	2.30E-05

Supplementary Table 4: rs2242367 association statistics in PSP case-control and phenotype GWAS

	PSP case-control GWAS*	PSP phenotype GWAS**
Odds ratio (95% confidence interval)	0.95 (0.85-1.12)	0.90 (0.81-1.09)
P-value	0.35	0.21

*= From phase 1 (pathologically diagnosed PSP cases) of Hoglinger *et al.,* 2011[8]

**= From Jabbari *et al.,* 2018[11]

Supplementary Analysis: AIC analyses on the impact of genetic principal components and study site on the Cox model's goodness of fit

Our AIC analysis of ten genetic PCs revealed an increase in the AIC, therefore a decrease in the model's goodness of fit, after the third PC, further justifying our use of the first three PCs as covariates.

We repeated the stage one analysis with 'study site' added in as a covariate. This resulted in rs2242367 remaining as our lead SNP with similar association statistics ($p = 9.3x10^{-11}$, hazard ratio (95% confidence interval) = 1.45 (1.29-1.62)) in comparison to the original analysis, with no new genome-wide significant SNPs identified. Additionally, an AIC analysis of clinical variables revealed an increase in the AIC when study site was included, further justifying our decision to not include it as a covariate in stage one and stage two analyses.

Supplementary Table 5: Coloc analysis results

Dataset	Gene Ensembl ID	Gene HGNC symbol	PP H0	PP H1	PP H2	PP H3	PP H4	PP H3 + PP H4
eQTLGen	ENSG00000038532	CLEC16A	6.77E-41	3.00E-41	0.562132362	0.248725303	0.189142335	0.43787
eQTLGen	ENSG00000048462	TNFRSF17	1.01E-15	4.98E-16	0.609005149	0.300345354	0.090649497	0.39099
eQTLGen	ENSG00000048471	SNX29	5.63E-304	2.68E-304	0.658293712	0.313890811	0.027815477	0.34171
eQTLGen	ENSG00000103274	NUBP1	1.14E-10	5.04E-11	0.670316213	0.294905241	0.034778547	0.32968
eQTLGen	ENSG00000103342	GSPT1	8.71E-29	4.29E-29	0.632992368	0.312036091	0.054971541	0.36701
eQTLGen	ENSG00000122299	ZC3H7A	3.41E-43	1.65E-43	0.657629717	0.318158804	0.02421148	0.34237
eQTLGen	ENSG00000139116	KIF21A	2.08E-05	0.056888849	3.44E-04	0.941961794	7.84E-04	0.94275
eQTLGen	ENSG00000151229	SLC2A13	1.19E-07	3.27E-04	3.64E-04	0.999236797	7.22E-05	0.99931
eQTLGen	ENSG00000153066	TXNDC11	1.07E-33	5.20E-34	0.631703501	0.308470529	0.05982597	0.36830
eQTLGen	ENSG00000166669	ATF7IP2	7.90E-290	3.94E-290	0.650210589	0.324309013	0.025480398	0.34979
eQTLGen	ENSG00000166676	TVP23A	0.619505661	0.274427257	0.069138988	0.03062073	0.006307365	0.03693
eQTLGen	ENSG00000171490	RSL1D1	6.17E-05	3.02E-05	0.627846467	0.307062909	0.064998814	0.37206
eQTLGen	ENSG00000173208	ABCD2	3.08E-81	8.43E-78	3.65E-04	0.999621777	1.36E-05	0.99964
eQTLGen	ENSG00000175604		0.56272362	0.271933745	0.104929532	0.050697025	0.009716079	0.06041
eQTLGen	ENSG00000175643	RMI2	3.11E-304	1.40E-304	0.671441769	0.3016548	0.026903431	0.32856
eQTLGen	ENSG00000179583	CIITA	2.27E-28	1.04E-28	0.661943285	0.304692433	0.033364282	0.33806
eQTLGen	ENSG00000182108	DEXI	1.70E-292	7.81E-293	0.648792233	0.298624111	0.052583656	0.35121
eQTLGen	ENSG00000184602	SNN	0.275580257	0.139646151	0.324328542	0.164252418	0.096192632	0.26045
eQTLGen	ENSG00000185338	SOCS1	2.71E-13	1.21E-13	0.674718362	0.300083748	0.02519789	0.32528
eQTLGen	ENSG00000188897		3.50E-30	1.67E-30	0.653719217	0.312033973	0.034246811	0.34628
eQTLGen	ENSG00000188906	LRRK2	1.26E-307	3.44E-304	3.64E-04	0.998951702	6.84E-04	0.99964
eQTLGen	ENSG00000189067	LITAF	2.09E-25	1.11E-25	0.610498701	0.323984831	0.065516468	0.38950
eQTLGen	ENSG00000223914		2.38E-04	0.652593437	1.15E-04	0.31375501	0.033298908	0.34705
eQTLGen	ENSG00000225342		5.74E-29	1.57E-25	3.62E-04	0.993372209	0.006265869	0.99964
eQTLGen	ENSG00000234719		2.88E-302	1.42E-302	0.651466257	0.321365198	0.027168545	0.34853
eQTLGen	ENSG00000241641	RPS23P6	1.03E-06	4.88E-07	0.623577578	0.296632091	0.079788816	0.37642
eQTLGen	ENSG00000260943		0.00E+00	4.80E-305	8.44E-05	0.230349299	0.769566338	0.99992
eQTLGen	ENSG00000261560		1.71E-11	8.36E-12	0.649539271	0.316673444	0.033787284	0.35046
eQTLGen	ENSG00000262151		0.534137872	0.245125258	0.141978993	0.065143041	0.013614835	0.07876
eQTLGen	ENSG00000262222		6.81E-05	3.13E-05	0.668507284	0.307300894	0.024092424	0.33139
eQTLGen	ENSG00000262488		2.57E-141	1.19E-141	0.666852564	0.309812721	0.023334716	0.33315
eQTLGen	ENSG00000262636		1.18E-49	5.40E-50	0.667028656	0.304655161	0.028316182	0.33297
eQTLGen	ENSG00000262703		4.81E-35	2.16E-35	0.661115609	0.297027083	0.041857308	0.33888
eQTLGen	ENSG00000262999		2.61E-17	1.25E-17	0.635633372	0.30571268	0.058653948	0.36437
eQTLGen	ENSG00000263013		6.18E-13	2.84E-13	0.664770493	0.305705007	0.0295245	0.33523
eQTLGen	ENSG00000263033		5.53E-07	2.44E-07	0.603962002	0.265944773	0.130092428	0.39604
eQTLGen	ENSG00000263307		0.564472337	0.276385779	0.099983883	0.048945461	0.01021254	0.05916
PsychENCODE	ENSG00000018236	CNTN1	6.55E-04	0.784268818	1.61E-04	0.192599921	0.022314877	0.21491
PsychENCODE	ENSG00000038532	CLEC16A	0.62233891	0.221264679	0.105568331	0.037520169	0.013307911	0.05083
PsychENCODE	ENSG00000048462	TNFRSF17	6.14E-33	2.29E-33	0.705732441	0.262922946	0.031344612	0.29427
PsychENCODE	ENSG00000048471	SNX29	0.527009718	0.195786301	0.185262192	0.068802535	0.023139254	0.09194
PsychENCODE	ENSG00000103274	NUBP1	0.608921653	0.204987697	0.129598768	0.043615319	0.012876563	0.05649
PsychENCODE	ENSG00000103342	GSPT1	0.558113708	0.208252462	0.154181508	0.057508768	0.021943554	0.07945
PsychENCODE	ENSG00000122299	ZC3H7A	0.613172362	0.223640041	0.108681339	0.039624052	0.014882205	0.05451
PsychENCODE	ENSG00000122304	PRM2	0.637683016	0.219408612	0.093602956	0.032189002	0.017116414	0.04931
PsychENCODE	ENSG00000139116	KIF21A	3.93E-04	0.463626299	4.43E-04	0.522119562	0.013417671	0.53554
PsychENCODE	ENSG00000151229	SLC2A13	2.42E-09	2.86E-06	8.45E-04	0.999115799	3.67E-05	0.99915
PsychENCODE	ENSG00000153060	TEKT5	0.197893411	0.07429749	0.513423625	0.192739126	0.021646348	0.21439
PsychENCODE	ENSG00000153066	TXNDC11	2.41E-19	8.76E-20	0.717956139	0.260893378	0.021150483	0.28204
PsychENCODE	ENSG00000166669	ATF7IP2	9.83E-51	2.84E-51	0.755525473	0.218196111	0.026278416	0.24447
PsychENCODE	ENSG00000166676	TVP23A	0.003135981	0.0011207	0.717515609	0.256395383	0.021832328	0.27823
PsychENCODE	ENSG00000171490	RSL1D1	0.291019608	0.107480154	0.420578788	0.155303685	0.025617765	0.18092
PsychENCODE	ENSG00000173208	ABCD2	3.40E-04	0.402925222	4.91E-04	0.581163121	0.015080945	0.59624
PsychENCODE	ENSG00000175604		0.582501999	0.213031302	0.137561563	0.050292088	0.016613048	0.06691
PsychENCODE	ENSG00000175643	RMI2	1.07E-183	3.67E-184	0.725432034	0.249226612	0.025341354	0.27457
PsychENCODE	ENSG00000175646	PRM1	0.390602095	0.134471807	0.317412297	0.109226614	0.048287188	0.15751

Dataset	Gene Ensembl ID	Gene HGNC symbol	PP H0	PP H1	PP H2	PP H3	PP H4	PP H3 + PP H4
PsychENCODE	ENSG00000178279	TNP2	0.630735621	0.217356366	0.102578034	0.035335188	0.013994791	0.04933
PsychENCODE	ENSG00000179583	CIITA	0.574851513	0.204516692	0.151250323	0.053795207	0.015586265	0.06938
PsychENCODE	ENSG00000180116	C12orf40	6.38E-04	0.75563901	1.84E-04	0.217943457	0.025595797	0.24354
PsychENCODE	ENSG00000182108	DEXI	2.03E-06	7.21E-07	0.696323703	0.247560154	0.056113394	0.30367
PsychENCODE	ENSG00000184602	SNN	0.184410219	0.071569423	0.398195422	0.15434777	0.191477167	0.34582
PsychENCODE	ENSG00000185338	SOCS1	0.633007037	0.21747707	0.099071414	0.034020735	0.016423744	0.05044
PsychENCODE	ENSG00000188897		5.78E-04	2.03E-04	0.720429933	0.252904168	0.025884844	0.27879
PsychENCODE	ENSG00000188906	LRRK2	5.71E-29	6.76E-26	8.44E-04	0.999124605	3.17E-05	0.99916
PsychENCODE	ENSG00000189067	LITAF	1.64E-06	6.47E-07	0.699110284	0.276273542	0.02461389	0.30089
PsychENCODE	ENSG00000199571		6.67E-04	0.796388608	1.50E-04	0.178886533	0.023907612	0.20279
PsychENCODE	ENSG00000205592	MUC19	5.67E-04	0.671556173	2.41E-04	0.284893379	0.042742995	0.32764
PsychENCODE	ENSG00000213853	EMP2	3.06E-16	1.18E-16	0.636553894	0.245074372	0.118371734	0.36345
PsychENCODE	ENSG00000223914		5.81E-04	0.687491783	1.59E-04	0.18847766	0.12328984	0.31177
PsychENCODE	ENSG00000225342		6.20E-04	0.734439107	2.04E-04	0.241334088	0.0234032	0.26474
PsychENCODE	ENSG00000229014	RPL30P13	4.26E-04	0.503646324	1.42E-04	0.167522234	0.328263969	0.49579
PsychENCODE	ENSG00000229899		6.81E-04	0.805788391	1.37E-04	0.162443686	0.030950183	0.19339
PsychENCODE	ENSG00000229917	RPL7P46	0.506796998	0.176731881	0.216508018	0.075476887	0.024486215	0.09996
PsychENCODE	ENSG00000234719		1.71E-56	6.43E-57	0.704497129	0.264217977	0.031284894	0.29550
PsychENCODE	ENSG00000241641	RPS23P6	0.530690387	0.193706453	0.189126345	0.069015245	0.017461571	0.08648
PsychENCODE	ENSG00000255991		6.81E-04	0.804264417	1.47E-04	0.17333685	0.021571029	0.19491
PsychENCODE	ENSG00000256013		0.583609931	0.225559644	0.127369884	0.049212991	0.014247551	0.06346
PsychENCODE	ENSG00000257237		7.43E-04	0.777709727	1.88E-04	0.196265972	0.025093246	0.22136
PsychENCODE	ENSG00000257680		6.55E-04	0.785042338	1.60E-04	0.191643586	0.022498882	0.21414
PsychENCODE	ENSG00000258144		3.08E-10	3.54E-07	8.69E-04	0.999103151	2.80E-05	0.99913
PsychENCODE	ENSG00000260224	UBL5P4	0.529666573	0.197222185	0.184651213	0.068735446	0.019724583	0.08846
PsychENCODE	ENSG00000260310		0.458231648	0.176503536	0.251965646	0.097036893	0.016262276	0.11330
PsychENCODE	ENSG00000260318	COX6CP1	0.59131542	0.22043719	0.124850612	0.04652634	0.016870438	0.06340
PsychENCODE	ENSG00000260468		0.567117654	0.219004879	0.14310506	0.055247622	0.015524785	0.07077
PsychENCODE	ENSG00000260488		0.61622182	0.228670048	0.10368206	0.038461788	0.012964284	0.05143
PsychENCODE	ENSG00000260723		0.592198827	0.211407003	0.126308241	0.04506532	0.025020609	0.07009
PsychENCODE	ENSG00000261216		0.478944581	0.178546329	0.228313575	0.085084182	0.029111332	0.11420
PsychENCODE	ENSG00000261293		0.451048896	0.164976505	0.25891619	0.094671293	0.030387116	0.12506
PsychENCODE	ENSG00000261394		0.584695496	0.208728415	0.141243826	0.050407218	0.014925045	0.06533
PsychENCODE	ENSG00000261560		2.94E-06	1.10E-06	0.707190694	0.263156199	0.029649068	0.29281
PsychENCODE	ENSG00000261904		0.599577984	0.202429819	0.136939065	0.046218601	0.014834531	0.06105
PsychENCODE	ENSG00000262020		0.61450576	0.21620066	0.111877927	0.039343775	0.018071879	0.05742
PsychENCODE	ENSG00000262117	BCAR4	0.60512589	0.224171792	0.112540482	0.041674675	0.016487162	0.05816
PsychENCODE	ENSG00000262151		0.605548057	0.215522693	0.121184762	0.043116658	0.01462783	0.05774
PsychENCODE	ENSG00000262158		0.627010316	0.211659681	0.110250281	0.037203277	0.013876445	0.05108
PsychENCODE	ENSG00000262222		0.336995831	0.119880117	0.382169769	0.135924892	0.025029391	0.16095
PsychENCODE	ENSG00000262259		0.628136799	0.212121414	0.109288793	0.036893199	0.013559795	0.05045
PsychENCODE	ENSG00000262322		0.607796629	0.205252551	0.128624503	0.043421511	0.014904807	0.05833
PsychENCODE	ENSG00000262381		0.617237396	0.20822802	0.120551681	0.040655365	0.013327538	0.05398

Supplementary Table 6: PSP survival GWAS association statistics for *LRRK2* locus lead SNPs from Parkinson's disease (PD), Crohn's disease (CD) and leprosy case-control GWAS

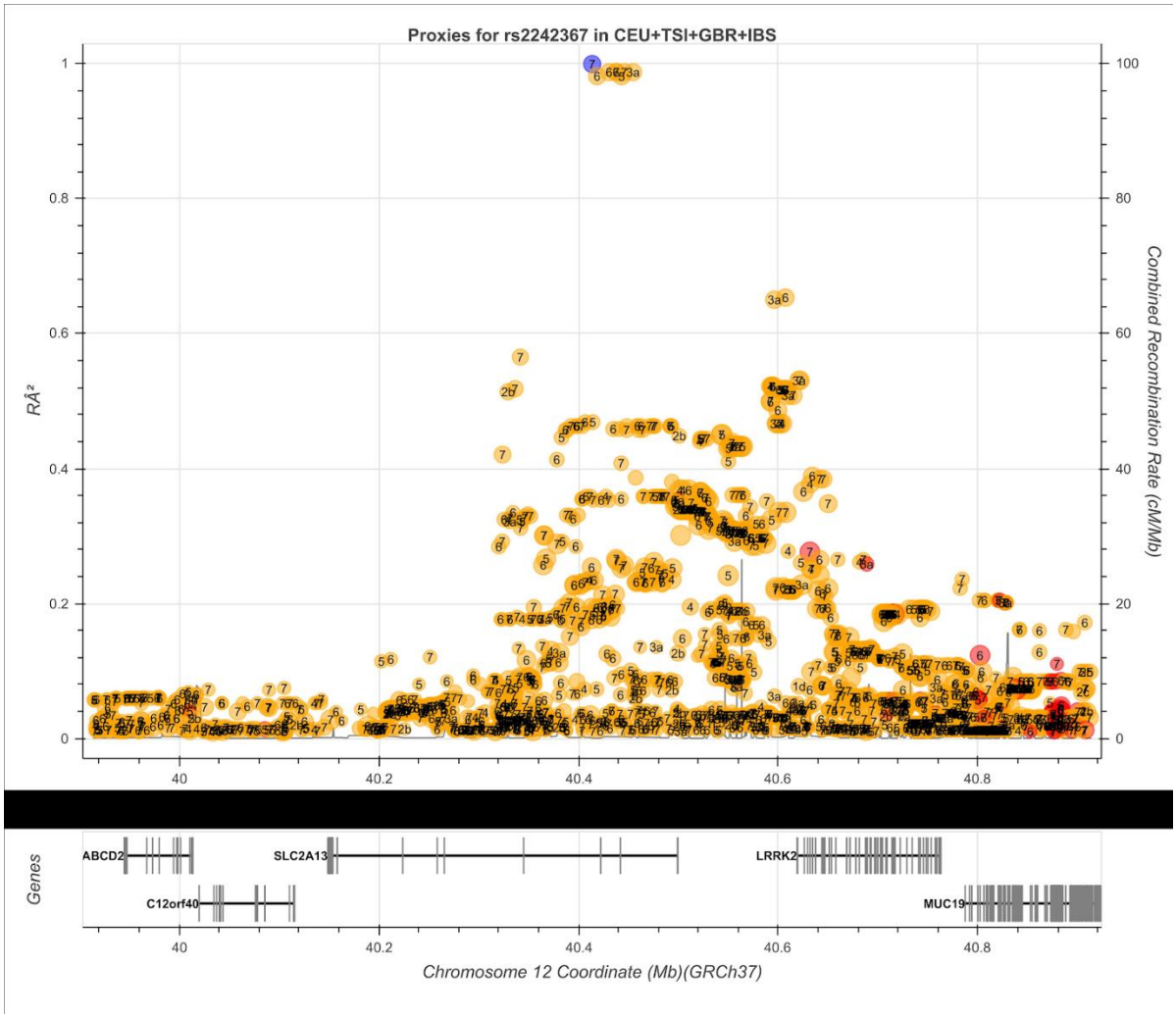
	PD-associated SNP (rs76904798)	CD-associated SNP (rs11175593)	Leprosy-associated SNP (rs1873613)
Minor allele frequency (%)*	12.8	2.1	28.8
r^2 LD with rs2242367**	0.05	0.01	0.09
D' LD with rs2242367**	0.90	1.0	0.32
PSP survival GWAS hazard ratio (95% confidence interval)	0.96 (0.89-1.08)	1.11 (0.95-1.21)	NA
PSP survival GWAS p-value	0.60	0.57	NA

PD = Parkinson's disease, CD = Crohn's disease, NA = SNP not present in dataset, r^2 LD = r-squared linkage disequilibrium, D' LD = D-prime linkage disequilibrium.

*= Data taken from non-Finnish white European cases in gnomAD browser (https://gnomad.broadinstitute.org/)

**= Data taken from non-Finnish white European cases in LDlink (https://ldlink.nci.nih.gov/?tab=ldpair)

Supplementary Figure 1: Linkage disequilibrium structure around rs2242367



Linkage disequilibrium plot from LDlink (https://ldlink.nci.nih.gov/) highlighting that there are no coding variants in linkage disequilibrium with the lead SNP rs2242367, as defined by the region encompassed by variants with an r^2>0.3. Blue dot = lead SNP rs2242367, yellow dots = non-coding SNPs, red dots = coding SNPs. Regulatory potential of each SNP (based on RegulomeDB scores - http://www.regulomedb.org/help#score) indicated by numerical value between 1 (high) to 7 (low).

PSP Genetics Group collaborators

Kin Y. Mok, David P. Murphy, Safa Al-Sarraj, Claire Troakes, Steve M. Gentleman, Kieren S.J. Allinson, Zane Jaunmuktane, Janice L. Holton, Andrew J. Lees, Christopher M. Morris, Yaroslau Compta, Ellen Gelpi, John C. van Swieten, Alex Rajput, Leslie Ferguson, Mark R. Cookson, J. Raphael Gibbs, Cornelis Blauwendraat, Jinhui Ding, Ruth Chia, Bryan J. Traynor, Alexander Pantelyat, Coralie Viollet, Olga Pletnikova, Juan C. Troncoso, Liana S. Rosenthal, Adam L. Boxer, Gesine Respondek, Thomas Arzberger, Sigrun Roeber, Armin Giese, David J. Burn, Nicola Pavese, Alexander Gerhard, Christopher Kobylecki, P. Nigel Leigh, Alistair Church, Michele T.M. Hu.

Affiliations: Department of Neurodegenerative Disease, UCL Queen Square Institute of Neurology, London, UK (Kin Y. Mok, David P. Murphy); MRC London Neurodegenerative Diseases Brain Bank, Institute of Psychiatry, King's College London, London, UK (Safa Al-Sarraj, Claire Troakes); Parkinson's UK Brain Bank, Department of Brain Sciences, Imperial College London, London, UK (Steve M. Gentleman); Department of Clinical Neurosciences and Cambridge University Hospitals NHS Trust, University of Cambridge, Cambridge, UK (Kieren S.J. Allinson); Queen Square Brain Bank for Neurological Disorders, UCL Queen Square Institute of Neurology, London, UK (Zane Jaunmuktane, Janice L. Holton, Andrew J. Lees); Translational and Clinical Research Institute, and Newcastle Brain Tissue Resource, Faculty of Medical Sciences, Newcastle University, Newcastle, UK (Christopher M. Morris); Movement Disorders Unit, Neurology Department, Hospital Clínic de Barcelona, University of Barcelona, Barcelona, Catalonia, Spain (Yaroslau Compta); Neurological Tissue Bank and Neurology Department, Hospital Clínic de Barcelona, University of Barcelona, Barcelona,

Catalonia, Spain (Ellen Gelpi); Department of Neurology, Erasmus Medical Centre, Rotterdam, The Netherlands (John C. van Swieten); Division of Neurology, Royal University Hospital, University of Saskatchewan, Canada (Alex Rajput, Leslie Ferguson); Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, Maryland, USA (Mark R. Cookson, J. Raphael Gibbs, Cornelis Blauwendraat, Jinhui Ding, Ruth Chia, Bryan J. Traynor); Department of Neurology, Johns Hopkins University Medical Center, Baltimore, Maryland, USA (Alexander Pantelyat, Liana S. Rosenthal); Department of Anatomy, Physiology and Genetics, Uniformed Services University of the Health Sciences, Bethesda, Maryland, USA (Coralie Viollet); Department of Pathology, The Johns Hopkins School of Medicine, Baltimore, Maryland, USA (Olga Pletnikova, Juan C. Troncoso); Memory and Aging Center, University of California, San Francisco, USA (Adam L. Boxer); German Centre for Neurodegenerative Diseases (DZNE), Munich; Department of Neurology, Hannover Medical School, Hannover, Germany (Gesine Respondek, Thomas Arzberger); Center for Neuropathology and Prion Research, Munich LMU, Munich, Germany (Sigrun Roeber, Armin Giese); Clinical Ageing Research Unit, Newcastle University, Newcastle, UK (David J. Burn, Nicola Pavese); Division of Neuroscience and Experimental Psychology, University of Manchester, Manchester, UK (Alexander Gerhard, Christopher Kobylecki); Department of Neuroscience, Brighton and Sussex Medical School, Brighton, UK (P. Nigel Leigh); Department of Neurology, Royal Gwent Hospital, Newport, UK (Alistair Church); Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, UK (Michele T.M. Hu).

Supplementary References

1. Höglinger GU, Melhem NM, Dickson DW, et al. Identification of common variants influencing risk of the tauopathy progressive supranuclear palsy. *Nat Genet* 2011; 43: 699-705.

2. Blauwendraat C, Faghri F, Pihlstrom L, et al. NeuroChip, an updated version of the NeuroX genotyping platform to rapidly screen for variants associated with neurological diseases. *Neurobiol Aging* 2017; 57: e9-e247.

3. Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT. Data quality control in genetic case-control association studies. *Nat Protoc* 2010; 5: 1564-1573.

4. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009; 25: 1754-1760.

5. DePristo MA, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genetics* 2011; 43: 491-498.

6. Jun G, Flickinger M, Hetrick KN, et al. Detecting and estimating contamination of human DNA samples in sequencing and array-based genotype data. *Am J Hum Genet* 2012; 91: 839-848.

7. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 2015; 4: 7.

8. Höglinger GU, Melhem NM, Dickson DW, et al. Identification of common variants influencing risk of the tauopathy progressive supranuclear palsy. *Nat Genet* 2011; 43: 699-705.

9. 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: 37.

10. 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: 41.

11. Jabbari E, Woodside J, Tan MMX, et al. Variation at the TRIM11 locus modifies progressive supranuclear palsy phenotype. *Ann Neurol* 2018; 84: 485-496.