17q21.31 sub-haplotypes underlying H1-associated risk for Parkinson's disease and progressive supranuclear palsy converge on altered glial regulation

Authors

KR Bowles[1,2], DA Pugh[1,2], K Farrell[1,2,3], N Han[1,2,3], J TCW[1,2], Y Liu[1,2], SA Liang[4], L Qian[1,2], J Bendl[5,6], JF Fullard[5,6], AE Renton[1,2], A Casella[3], MA Iida[3], S Bandres-Ciga[7], Z Gan-Or[8,9,10], P Heutink[11,12], A Siitonen[13,14], S Bertelsen[1,2], CM Karch[15], SJ Frucht[16], BH Kopell [17,18], I Peter[6,19], YJ Park[17,20], PK Crane[21], JSK Kauwe[22], KL Boehme[22], GU Höglinger[23,24,25,26], PART working group[27], International Parkinson's Disease Genomics Consortium (IPDGC)[28], Progressive Supranuclear Palsy Genetics Consortium[29], A Charney[1,6,17,20] , P Roussos[1,5,6,30], JC Wang[1,2,6], WW Poon[4], T Raj[1,2,6,31], JF Crary[1,2,3] & AM Goate[1,2,6,31*]

Correspondence

alison.goate@mssm.edu (A.M.G)

Affiliations

[1]Nash Family Department of Neuroscience & Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, NY, United States of America

[2]Ronald M. Loeb Center for Alzheimer's disease, Icahn School of Medicine at Mount Sinai, New York, NY, United States of America

[3]Department of Pathology, Icahn School of Medicine at Mount Sinai, New York, NY, United States of America

[4]Institute for Memory Impairments and Neurological Disorders, University of California, Irvine, CA, United States of America

[5]Pamela Sklar Division of Psychiatric Genomics, Department of Psychiatry, Icahn School of Medicine, New York, NY, United States of America

[6] Icahn Institute for Data Science and Genomic Technology, Icahn School of Medicine at Mount Sinai, New York, NY, United States of America

[7]Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, Maryland, United States of America

[8]Department of Human Genetics, McGill University, Montréal, Québec, Canada

[9]Montreal Neurological Institute, McGill University, Montréal, Québec, Canada

[10]Department of Neurology and Neurosurgery, McGill University, Montréal, Québec, Canada

[11]Department for Neurodegenerative Diseases, Hertie Institute for Clinical Brain Research, University of Tübingen, Tübingen, Germany

[12]German Center for Neurodegenerative Diseases (DZNE), Tübingen, Germany

[13]Institute of Clinical Medicine, Department of Neurology, University of Oulu, Oulu, Finland

[14]Department of Neurology and Medical Research Center, Oulu University Hospital, Oulu, Finland.

[15]Department of Psychiatry, Washington University in St Louis, St. Louis, MO, United States of America

[16]Fresco Institute for Parkinson's and Movement Disorders, Department of Neurology, New York University Langone, New York, NY, United States of America

[17]Department of Neurosurgery, Icahn School of Medicine at Mount Sinai, New York, NY, United States of America

[18]Center for Neuromodulation, Icahn School of Medicine at Mount Sinai, New York, NY, United States of America

[19]Institute for Exposomic Research, Icahn School of Medicine at Mount Sinai, New York, NY, United States of America

[20]Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY, United States of America

[21]Department of Medicine, University of Washington, Seattle, WA, United States of America

[22]Department of Biology, Brigham Young University, Provo, UT, United States of America

[23]Department of Neurology, Hannover Medical School, Hannover, Germany

[24]Department of Neurology, Technical University, Munich, Germany

[25]Department of Translational Neurodegeneration, German Centre for Neurodegenerative Diseases (DZNE), Munich, Germany

[26]Munich Cluster for Systems Neurology (SyNergy), Munich, Germany

[27]Primary Age-Related Tauopathy (PART) working group and other contributing brain bank authors & affiliations in Supplementary material 1

[28]IPDGC authors and affiliations in Supplementary material 2

[29]PSP Genetics Consortium authors and affiliations in Supplementary material 3

[30]Mental Illness Research, Education and Clinical Centers, VISN 2, JJ Peters VA Medical Center, Bronx, New York, NY, United States of America

[31]Estelle and Daniel Maggin Department of Neurology, Icahn School of Medicine at Mount Sinai, New York, NY, United States of America

Abstract

Parkinson's disease (PD) and progressive supranuclear palsy (PSP) are clinically similar neurodegenerative movement disorders that display unique neuropathological features (i.e. Lewy body pathology and Tau pathology, respectively). While each disorder has distinct clinical and genetic risk factors, both are associated with the *MAPT* 17q.21.31 locus H1 haplotype. This suggests a pleiotropic effect of this genomic region. To better understand the genetic contribution of this region to these diseases, we fine-mapped the apparent pleiotropy of this locus. Our study indicates that PD and PSP are associated with different sub-haplotypes of the H1 clade. PD-associated sub-haplotypes were associated with altered *LRRC37A* copy number and expression, which, like other PD risk-associated genes, we hypothesize to be most relevant to astroglial function. In contrast, PSP was associated with grossly altered LD structure across the 17q21.31 locus, and risk-associated variants were found to impact chromatin structure in both neurons and microglia. We conclude that the contribution of the 17q21.31 locus to multiple disorders is a result of its structural and haplotypic complexity, which in turn impacts the regulation of multiple genes and neural cell types. This raises the possibility of novel disease-specific pathogenic mechanisms driven by 17q21.31 structural variation and altered epigenetic regulation that appear to converge on glial function and gene expression. By fine-mapping the association of H1 with PD and PSP, we have begun to untangle the apparent pleiotropy of this locus, and gain better insight into the mechanism of each disease, which will guide future functional analyses and disease models for PD and PSP.

Introduction

The *MAPT* 17q21.31 locus lies within a 1.5Mb inversion region of high linkage disequilibrium (LD), conferring two distinct haplotypes; H1, which has a frequency of ~ 0.8 in European ancestry populations, and the less common, inverted H2 (frequency ~ 0.2), which is absent or lower in East and South Asian populations (frequency 0 - 0.09) (Fig 1A, SI Fig 1A). The major haplotype, H1, has been genetically associated with increased risk for multiple neurodegenerative disorders, including *APOE* ε4-negative Alzheimer's disease (AD)[1], corticobasal degeneration (CBD)[2], Parkinson's disease (PD)[3–7] and progressive supranuclear palsy (PSP)[8–10].

PSP is neuropathologically characterized by the presence of aggregated, predominantly 4-repeat (4R) hyper-phosphorylated Tau in astrocytic tufts present throughout the subthalamic nucleus, basal ganglia, brainstem and motor cortex, but can also appear with neurofibrillary tangles and oligodendal coiled bodies[11–14]. Clinically, PSP is defined by early motor dysfunction, specifically regular falls, postural instability and supranuclear vertical gaze palsy[11–14]. Motor symptoms are commonly accompanied by cognitive decline and behavioral abnormalities[15,16]. As Tau pathology is the predominant feature of PSP, and rare autosomal dominant mutations within the *MAPT* gene, such as R5L[17], A152T[18,19], G303V[20], and S305S[21,22], among others (see https://www.alzforum.org/mutations/mapt), are known to be pathogenic for both PSP and frontotemporal dementia (FTD), it is unsurprising that genome-wide association (GWA) studies reveal a significant association between PSP and the 17q21.31 *MAPT* locus, with an odds ratio (OR) of ~ 4-5 for the H1 haplotype (and OR ~ 0.2 for the protective H2 haplotype)[8,9]. The H1 association with PSP has been linked to the H1c sub-haplotype specifically[8,23,24], although additional 17q21.31 sub-haplotypes are also enriched in PSP cases[10,25]. However, the mechanism by which the H1c sub-haplotype confers risk for PSP remains unclear. It has been reported that the H1c sub-haplotype results in increased *MAPT* expression and altered splicing *in vitro*[26,27] due to the location of the H1c tagging single nucleotide polymorphism (SNP) rs242557 within a regulatory region of *MAPT* intron 1[26,28,29], although these *in vitro* studies report conflicting results and have not been replicated in human brain or *in vivo*[30,31].

Similar to PSP, PD is a movement disorder that commonly involves executive dysfunction and dementia[32,33], but is classically characterized by bradykinesia, tremor, rigidity, postural instability, and numerous non-motor symptoms[34]. Neuropathologically, PD is an α-synucleinopathy defined by the presence of intraneuronal accumulation of α-synuclein in Lewy bodies throughout the substantia nigra, brainstem and forebrain[35,36]. The genetic association between the 17q21.31 locus and PD is therefore surprising. While it is not rare for tauopathy to occur alongside α-synuclein inclusions in the substantia nigra[37,38], aggregated Tau is not a typical neuropathological feature of this disorder[36]. Despite the genetic association with 17q21.31, the OR is substantially smaller for PD than for PSP (OR ~ 1.4-1.5 for H1, OR ~ 0.8 for H2)[5,39], and there does not appear to be any association with the H1c sub-haplotype[40,41], indicating that different 17q21.31 locus variants and mechanisms may underlie the relative risk for each disease.

It is not currently understood how variation within this single locus underlies apparently independent and disparate effects, a phenomenon called "pleiotropy", resulting in distinct

disorders with unique neuropathological features. However, the 17q21.31 locus spans multiple genes in addition to *MAPT*, and comprises several sub-haplotypes, defined by complex rearrangements and copy number variation in their distal regions[42,43]. Given that PSP and PD share a similar genetic risk associated with this large complex locus, we sought to determine the specific genes and distinct sub-haplotypes within the 17q21.31 H1 haplotype clade associated with each disease, and investigate the functional impact that these variants may have on disease risk and pathogenesis.

Results

The H1c haplotype is associated with progressive supranuclear palsy

Previous PSP GWA studies have demonstrated a significant association with the 17q21.31 locus, but have not fine-mapped the association with the major H1 haplotype beyond H1c tag SNPs[8,9]. We therefore sought to first confirm the association with 17q21.31 in previously published PSP data[8] using a larger, aged adult control cohort for comparison[44] (Stage 1), as well as in a novel, independent autopsy confirmed PSP/corticobasal degeneration (CBD) and control cohort (Stage 2, SI Methods Table 1). We confirmed the previously reported association of the 17q21.31 locus with PSP risk[8,9] using a case-control analysis in these two datasets (Stage 1 1,071 cases, 11,215 controls; Stage 2 399 PSP cases, 56 CBD cases, 886 controls; SI Fig 1, SI Table 1, SI Methods Table 1). Using the H1/H2 tag SNP rs8070723, we observe ORs similar to previously reported effect sizes[8,9], ranging from 0.19 -0.33 (95% CI 0.16-0.43) for the H2 haplotype, and 3.04-5.18 (95% CI 2.34-6.30) for H1 (SI Fig 1, SI Table 1).

To determine the specific H1 variants associated with increased risk for PSP, we excluded individuals carrying an H2 allele, and repeated the regression analysis in H1 homozygotes only (Fig 1B, SI Fig 2, SI Table 2). In Stage 1 data, we observe a significant ($p < 1x10^{-5}$) association signal spanning both *MAPT* and *KANSL1*, with a peak within *MAPT* intron 1 (SI Fig 2), suggesting the contribution of multiple sub-haplotypes to PSP risk. The top SNP within this peak was rs242562 ($p = 2.39x10^{-12}$, OR 1.44 (95% CI 1.31-1.59) (SI Fig 2, SI Table 2). This SNP is in high LD with the H1c tagging SNP rs242557 (D' = 0.96, $r^2 = 0.96$), which was also significantly associated with increased PSP risk in this dataset with a similar p-value and effect size ($p = 3.6x10^{-12}$, OR 1.45 (95% CI 1.31-1.59) (SI Fig 2, SI Table 2). Our top SNP is therefore likely to be tagging this H1c haplotype. In contrast, we lacked sufficient power to discover this signal in our Stage 2 dataset, although both the top Stage 1 SNP rs242562 and H1c tag SNP rs242557 also showed similar effect sizes with p-values < 0.05 in the Stage 2 data (SI Fig 2, SI Table 2). When Stage 1 and Stage 2 data were pooled together, the H1c tag SNP rs242557 became the top SNP with a strengthened association p-value compared to Stage 1 data alone ($p = 2.57x10^{-13}$, OR 1.39 (95% CI 1.28-1.51)) (Fig 1B, SI Fig 2, SI Table 2).

H1 homozygote 17q21.31 locus LD patterns are altered in PSP compared to controls

In both Stage 1 and Stage 2 data, as well as in the merged dataset, we observed a striking loss of LD in PSP cases in the region of 17q21.31 where we see the genetic association (Fig 1C, SI Fig 3), consistent with our finding that variation in this region is contributing to PSP risk. However, this loss does not encompass *MAPT* intron 1 where the H1c peak association lies, indicating a decoupling or altered recombination between these regions in PSP cases. In contrast, we do not observe any gross differences in LD across this locus in PD cases (N = 3,475) compared to controls (N = 5,177)[3,5,6] (SI Fig 4).

Multiple H1 variants contribute to PSP risk, independent of the H1c haplotype.

Six SNPs are present within the H1H1 *MAPT* intron 1 association peak; rs8076152, rs12947764, rs11867549, rs242557, rs242562 and rs35838379 (SI Table 3). A focused analysis of the LD between these SNPs suggests three independent association signals. The first three SNPs are in high LD with each other ($D' > 0.92$, $r^2 > 0.57$) but only moderately or in low LD with the remaining four SNPs. The next pair is in high LD with each other ($D' > 0.97$, $r^2 = 0.96$). The final SNP is in only low to moderate LD with the other five SNPs and varies between PSP cases and controls ($D' < 0.8$, $r^2 < 0.35$ cases, $D' < 0.7$, $r^2 < 0.3$ controls) (SI Fig 5). Analyses conditioned on each of these SNPs confirmed the presence of two independent association signals in addition to the previously identified H1c association (SI Fig 5, SI Table 4). Analyses incorporating conditioning on all three signals using the tag SNPs rs8076152, rs242557 and rs35838379 did not entirely remove the association signal across *MAPT* and *KANSL1* (SI Fig 5, SI Table 4), but appeared to retain the broad suggestive haplotype signal across the entire region (SI Fig 5, SI Table 4).

Multiple haplotype association signals were apparent in the H1 17q21.31 locus Manhattan plot, characterized by multiple SNPs with similar p-values across the region (Fig 1B). We therefore selected SNPs from the center of each of these four apparent haplotype signals (SI Table 4) and repeated the conditional analyses. All four selected SNPs were in high LD with rs35838379 (SI Fig 6), but had differential effects in the conditional analyses (SI Table 5), and no single SNP alone was able to remove the full association signal in this locus (SI Fig 6, SI Table 5). Two of the four SNPs had the same impact on the association signal as rs35838379, but the other two, rs6503455 and rs58810165, appeared to have independent effects from each other, and from the SNPs present within the association peak (SI Fig 6, SI Table 5). Conditional analysis on all five tag SNPs (rs8076152, rs242557, rs35838379, rs6503455 and rs58810165) was required to fully suppress the association of this locus with PSP risk (SI Fig 6, SI Table 5), indicating that in addition to H1c and the two novel signals we identified in *MAPT* intron 1, there are also a further two, uncharacterized association signals spanning *MAPT* and *KANSL1* that contribute to PSP risk. Tag SNPs for these five signals and their associated effect sizes are summarized in Table 1.

PD does not share the same risk-associated H1 variants as PSP

After identifying multiple variants on the H1 clade contributing to PSP risk, we tested whether those same variants on H1 also conferred risk for PD. We assessed the effect of PSP-associated SNPs in H1 homozygotes from our Stage 1 and Stage 2 PD data (SI Methods Table 2), and observed no significant association between any of the five PSP SNPs in either PD dataset, including H1c (SI Table 6).

To confirm the genetic association of the 17q21.31 H1 haplotype with PD risk, we carried out a case-control association analysis in the region of interest (SI Fig 7, SI Table 7) in two independent datasets (Stage 1 2,780 PD cases, 6,384 controls; Stage 2 2,699 cases, 2,230 controls; SI Methods Table 2). The SNP with the strongest association in Stage 1 (rs17763050, $p = 2.74 \times 10^{-9}$) is in high LD with the known H1/H2 haplotype tag SNP rs8070723 (SI Fig 7E). Both SNPs are associated with ORs ~0.75 (95% CI ± 0.1) for the H2 haplotype, and OR 1.22 (95% CI ± 0.1) for the H1 haplotype (SI Table 6), consistent with previously reported effect sizes[3,5]. Due to a smaller cohort size and consequent lack of power, the association with 17q21.31 is less prominent in Stage 2 data. However, meta-analysis of both cohorts confirms a significant association between PD risk and the 17q21.31 locus (SI Fig 7, SI Table 7), per the H2 haplotype conferring protection with an OR of 0.82 (95% CI 0.76-0.89, random effects meta-analysis $p < 0.01$), and increased risk associated with H1 (OR 1.22 (95% CI 1.08-1.37)).

As the major H1 haplotype is associated with increased risk for PD, we removed individuals carrying an H2 allele and repeated the association analysis in H1 homozygotes only in order to identify variants of H1 that may confer additional risk for PD (SI Fig 8, SI Table 8). While association across the 17q21.31 locus became weaker in H1 homozygotes alone compared to the full analysis, we observe a distinct signal spanning *MAPT* and *KANSL1* genes in Stage 1 and Stage 2 analyses, with much smaller effect sizes compared to PSP. The Stage 1 top SNP, rs41543512, was associated with an OR of 1.21 (95% CI 1.10-1.32, $p < 0.001$, SI Fig 8, SI Table 7), but did not reach statistical significance in our Stage 2 data or meta-analysis. Similarly, the most significant variants identified in the Stage 2 analysis showed little effect in the Stage 1 data (SI Table 8, SI Fig 8).

17q21.31 H1 sub-haplotype blocks spanning *MAPT* & *KANSL1* are associated with PD risk in two independent datasets

Given that we were unable to fine-map the apparent association between PD risk and variants within the H1 haplotype using SNP-based association analyses, we decided to leverage the presence of high LD within this region. In contrast to PSP (Fig 1C, SI Fig 3), there were no gross differences in LD across the 17q21.31 locus between cases and controls (SI Fig 4). We then defined sub-haplotype blocks spanning the whole locus, and performed a logistic regression association analysis on each block (Fig 2A-B, Table 2). This approach greatly improved the power to detect disease-associated H1 variation in both Stage 1 and Stage 2 data and identified greater consistency between datasets. In Stage 1, we observe a peak spanning *MAPT* and the first 5 exons of *KANSL1* (Fig 2A) that reaches the genome-wide suggestive significance threshold of $p=1 \times 10^{-5}$. Within this peak, three sub-haplotype blocks showed substantial overlap of SNPs in Stage 2 (Fig 2B black arrows, Table 2). Using the exact SNPs present within these three blocks from the Stage 1 data, which we named H1.1 ($p = 1.73 \times 10^{-6}$), H1.2 ($p = 2.4 \times 10^{-4}$) and H1.3 ($p = 1.05 \times 10^{-5}$), we also repeated the block association in Stage 2 data and observe replication of this signal (Fig 2B, Table 2); block H1.2 is highly significant in the Stage 2 data ($p = 1.12 \times 10^{-9}$), while both blocks H1.1 and H1.3 are nominally significant ($p < 0.002$ and $p < 0.0003$, respectively).

Blocks H1.1, H1.2 and H1.3 each consist of multiple SNPs in high LD (SI Fig 9) that generate multiple sub-haplotypes (SI Fig 10). Each sub-haplotype is differentially associated with PD susceptibility (SI Fig 10, Table 3), with each block containing both risk- and protective sub-haplotypes with ORs ranging from 0.37 (95% CI 0.32-0.42, $p = 1.3x10^{-49}$, H1.1c) to 2.51 (95% CI 2.2-2.86, $p = 2.4x10^{-45}$, H1.1e). Despite the presence of significant heterogeneity between Stages 1 and 2 (Table 3), the most frequently occurring sub-haplotypes were often replicated in both analysis stages and fixed effects meta-analyses. Specifically, we identify two sub-haplotypes in block H1.1, H1.1b and H1.1e, which increase risk across datasets and analyses when compared against the most common haplotype, with ORs ranging from 1.26-1.6 and 1.45-2.51 respectively (SI Fig 10, Table 3). In the same block we also observe a protective sub-haplotype (H1.1c), associated with an OR ranging 0.37-0.96 (SI Fig 10, Table 3). Blocks H1.2 and H1.3 encompassed multiple sub-haplotypes with frequencies < 0.1, and exhibited greater variability and heterogeneity across stages. However, we identify one risk-associated sub-haplotype in block H1.2 (H1.2c, OR = 1.12-1.31, fixed effects $p < 0.01$) and two protective sub-haplotypes in block H1.3 (H1.3b; OR = 0.95-0.43, fixed effects $p < 0.001$, H1.3g; OR = 0.55-0.83, fixed effects $p < 0.002$) (Table 3, SI Fig 10).

We then assayed the impact of PD-associated sub-haplotype blocks H1.1, H1.2 and H1.3 in the combined PSP data. We found that all three blocks were highly associated with PSP (Table 4), although analysis of the specific sub-haplotypes within each of these blocks revealed that PSP and PD did not share the same risk and protective variants. In block H1.1, the PD risk-associated sub-haplotype H1.1b had no association with PSP risk, whereas the PD protection-associated sub-haplotype H1.1c was significantly associated with increased risk for PSP (OR = 1.43, 95% CI = 1.27-1.60, fisher's exact p-value = $8.17x10^{-10}$; Table 5, SI Fig 11). Similarly, in block H1.2, PSP-associated sub-haplotypes showed no significant effect in the PD data, and PD risk-associated sub-haplotypes showed little to no effect on PSP risk (Table 5, SI Fig 11). Finally, in block H1.3, we again observe the association of sub-haplotypes that are protective against PD, such as H1.3b and H1.3f, with significantly increased PSP risk (ORs ~1.42, 95% CI = 1.21-1.66, fisher's exact p-value = $4.7x10^{-10}$, $2.02x10^{-05}$, respectively; Table 5, SI Fig 11).

PD-associated sub-haplotypes are associated with *LRRC37A* gene expression in human brain

In order to elucidate the functional consequences of our PD-associated H1 block sub-haplotypes and PSP-associated *MAPT* intronic variants, we queried publicly available post-mortem human brain RNA-seq data from dorsolateral prefrontal cortex (PFC) and temporal cortex (TCX) from the AMP-AD and CommonMind consortia (SI Methods Table 3). We identified individuals within these datasets homozygous for our sub-haplotypes of interest and analyzed the expression of each of the genes present within the 17q21.31 locus. Despite the position of the H1 association peaks across *MAPT* and *KANSL1*, we did not observe any differences in the expression of either of these genes between any sub-haplotypes in any block (SI Figs 12, 13) or with any *MAPT* intronic variant (SI Fig 14). We also did not observe any difference in *MAPT* exons 2, 3 or 10 percent spliced in (PSI) values between sub-haplotypes or genotypes (SI Fig 15). The only genes within the 17q21.31 locus that had a significant association with H1 PD-associated sub-haplotypes were *LRRC37A* and its homolog *LRRC37A2* (a.k.a *LRRC37A/2*; Fig 3A-C, SI Fig 12). We observe significantly

increased *LRRC37A/2* expression in PD protective sub-haplotypes, specifically in H1.1c and H1.3b (H1.1c ~ 4.7 fold, p <0.001, H1.3b ~ 5 fold, p <0.001), as well as in sub-haplotypes whose effects were not replicated across PD data-sets; H1.2b and H1.3e (H1.2b ~5, p <0.001, H1.3e ~2.9x, p <0.01), but were protective in the Stage 2 PD analysis (Fig 3A-C). Quantitative reverse-transcription PCR (qRT-PCR) on postmortem prefrontal cortex from a small number of individuals supported the observation of increased *LRRC37A/2* expression in these sub-haplotypes (Fig 3D).

In contrast, nominally reduced *LRRC37A/2* expression was observed in only one sub-haplotype, H1.1d (Fig 3A), although the genetic association of this sub-haplotype with PD was inconsistent between populations, and the direction of effect differed greatly between cohorts. We observed a significant reduction in *LRRC37A/2* expression in the PD risk-associated H1.1b sub-haplotype by qRT-PCR in human brain tissue (p <0.01; Fig 3D). In contrast, the association between PSP risk-associated SNPs and *LRRC37A/2* expression is inconsistent (SI Fig 16), suggesting that *LRRC37A/2* expression may not be the driving mechanism regulating PSP risk in this region.

Interestingly, *LRRC37A/2* is the only gene in the locus that is significantly altered between the major 17q21.31 H1 and H2 haplotypes, with significantly higher expression in the protective H2 (SI Fig 17), suggesting that increased *LRRC37A/2* expression in brain may be associated with protection against PD. However, the interaction between *LRRC37A/2* expression and the structure and context of the H1 haplotype may be more complex for PSP.

PD-associated sub-haplotypes are associated with altered *LRRC37A/2* copy number

The 17q21.31 locus is structurally complex and encompasses regions of copy number variation (CNV) at its distal ends[42,43]. As sub-haplotype blocks within *MAPT* and *KANSL1* were associated with altered *LRRC37A/2* expression, we tested whether they tag structural variants (as defined by Boettger et al (2012)[42]; Fig 3G) in individuals homozygous for sub-haplotypes of interest. Using DNA samples from either blood or brain tissue (SI Methods Table 3) we performed dPCR for *MAPT*, alpha, beta and gamma regions[42], as well as for *LRRC37A/2* specifically (Fig 3H-K, SI Fig 18).

The majority of structural variation in alpha and beta regions is found in the most common sub-haplotype in each block (H1.1a, H1.2a and H1.3a), with each subsequent sub-haplotype carrying fewer copies of these regions (Fig 3H-I, SI Fig 18A-B). However, gamma and *LRRC37A/2* CNVs appear to vary with sub-haplotype within each block; those sub-haplotypes exhibiting increased *LRRC37A/2* expression are also associated with significantly increased gamma and/or *LRRC37A/2* copy number (H1.1c, H1.2b and H1.3b; ~ 3-5 copies gamma compared to 2-3 copies in controls, and ~ 8-11 copies of *LRRC37A/2* compared to 5-10 copies in controls, p<0.05; Fig 3J-K, SI Fig 18), suggesting that structural variation may underlie the increased expression of *LRRC37A/2* in these sub-haplotypes. Interestingly, risk-associated sub-haplotype H1.1b is associated with increased gamma and *LRRC37A/2* copy number (p < 0.01; Fig 3K) but not *LRRC37A/2* expression, indicating that additional factors likely contribute to *LRRC37A/2* expression and PD risk in this locus.

In contrast, there are no associations between any SNPs associated with PSP, including *MAPT* intron 1 variants, and CNVs in either alpha, beta, gamma or *LRRC37A/2* regions (SI Fig 19). Consistent with the CNV haplotypes defined in Boettger et al., 2012[42], we observe no difference in alpha CNVs between the major H1 and H2 haplotypes. However, we do observe increased beta, gamma and *LRRC37A/2* copy number in H2 homozygotes (SI Fig 20), consistent with the increased expression of *LRRC37A/2* in H2.

LRRC37A is a membrane-associated protein implicated in astrocyte cellular migration

As our analyses of PD and PSP sub-haplotypes appear to converge on the expression and/or copy number of *LRRC37A/2*, we explored the likely function of these genes. Very little is known about the role of *LRRC37A* or *LRRC37A2*, other than microdeletions of this region appear to contribute to familial dyslexia[45], and overexpression in HeLa cells induces the formation of filopodia[46], suggesting some relevance of this gene to neural function. As our data suggest that variation in *LRRC37A/2* expression may partially result from copy number variation, we carried out RNA-seq analysis in HEK293T cells overexpressing *LRRC37A/2* in order to elucidate a potential function for this gene. The number of significantly differentially expressed protein-coding genes (fold change ± 1.5, adjusted p-value < 0.05) in the context of *LRRC37A/2* overexpression was minimal (28 upregulated, 21 downregulated; SI Fig 21), suggesting that *LRRC37A/2* is unlikely to play a major regulatory role. However, functional enrichment of gene ontology (GO) terms for significantly differentially expressed genes indicate a role for LRRC37A at the cell membrane (Fig 4A, SI Fig 21), with significant enrichment for cell communication (GO:0007154, $p < 0.05$), and neuroactive ligand-receptor interaction (KEGG:04080, $p < 0.05$), as well as nominal enrichment for membrane-component-related pathways. We confirmed the cellular localization of LRRC37A to the cell membrane in HEK293T cells, iPSC-derived neurons and iPSC-derived astrocytes by isolating cytosolic and membrane-associated proteins from each cell type and analyzing the resulting fractions by western blot (Fig 4B, SI Fig 21).

Ingenuity pathway analysis (IPA) also suggests a role for LRRC37A at the plasma membrane and in the extracellular space; specifically, these analyses indicated that increased *LRRC37A/2* expression results in upregulated cellular movement pathways, such as increased migration of cells ($p < 0.01$, z score = 1.85; Fig 4C) and migration of granulocytes ($p < 0.05$, z score = 1.987), as well as increased recruitment of leukocytes ($p < 0.05$, z score = 2.386), upregulation of chemotaxis ($p < 0.05$, z score = 0.918; Fig 4D) and increased fatty acid metabolism ($p < 0.01$, z score = 2.582). Finally, gene set enrichment analysis (GSEA; SI Table 9, Fig 4E) suggested that increased *LRRC37A/2* expression may also upregulate DNA damage response-associated pathways (normalized enrichment score (NES) = 2.23, adjusted $p < 0.05$), as well as downregulate several glycosaminoglycan-related pathways, such as heparin sulfate metabolism (NES = -2.15, adjusted $p < 0.05$) and aminoglycan biosynthetic processes (NES -2.23, adjusted $p < 0.05$).

In order to confirm that we were not observing spurious changes in gene expression due to gross overexpression in a cell culture model, we carried out a titration of *LRRC37A/2* overexpression in HEK293T cells. We observed dose-dependent changes in the expression of genes that were significantly up or downregulated in our RNA-seq data (SI Fig 21), confirming that the function

of these pathways are likely to be altered by *LRRC37A/2* copy number and expression. As there is increased *LRRC37A/2* expression and copy number in the context of the H2 haplotype compared to H1, we also compared the expression of *LRRC37A/2* and other genes altered by *LRRC37A/2* expression in iPSC-derived neurons and astrocytes homozygous for either H1 or H2 haplotypes (Fig 4F, SI Fig 22). Interestingly, we only observe increased *LRRC37A/2* expression in H2 astrocytes, but not in H2 neuronal cultures (Fig 4F). Consistently, we also only observe differences in several *LRRC37A/2*-associated genes in H2 astrocyte cultures and not in neuronal cultures (SI Fig 22), suggesting that in the brain, H1/H2-associated *LRRC37A/2* expression changes may be specifically impacting astroglial gene expression and function.

PSP-associated *MAPT* intronic variants alter chromatin structure in neurons and microglia

As we did not observe consistent associations between PSP-associated *MAPT* intron 1 variants with either *LRRC37A/2* copy number and expression, or *MAPT* expression and splicing, we examined ATAC-seq data and publicly available epigenetic data in order to determine the likely functional impact of rs8076152 and H1c variants. Analysis of publicly available ChIP-seq data (Roadmap Epigenetics Consortium[47]) reveals that rs242557 lies within a region enriched for H3K27ac, H3K4me1and H3K9ac marks in brain tissue (Fig 5A, SI Table 10), which is consistent with the presence of an active enhancer region in this locus[48–50]. The second PSP-associated variant, rs8076152, is also within a brain-specific weak enhancer region (Fig 5A, SI Table 10). ATAC-seq data derived from neuronal (NeuN+) and glial (NeuN-) cells from human brain[51] shows that while rs8076152 may lie within a brain-specific enhancer region, it does not appear to be associated with an open chromatin peak (Fig 5B, SI Fig 23). In contrast, rs242557 lies within an open chromatin peak that is present only in glial cells (Fig 5B, SI Fig 23, SI Table 11).

Given this surprising finding, we then assayed ATAC-seq data from iPSC-derived neurons, astrocytes and microglia homozygous for either the H1c minor allele (A) or the major allele (G) (Fig 5C-D, SI Fig 23). We find that in neurons homozygous for the major allele, there is no open chromatin spanning *MAPT* intron 1, however in H1c homozygous neurons, the presence of the H1c minor allele creates a peak at the rs242557 locus and at the site of another SNP in high LD with rs242557 that was also significantly associated with PSP in our association analysis; rs242562 (Fig 5C, SI Table 11). In contrast, we observe no such effect in astrocytes, although there is an open chromatin peak at rs242562 in both control and H1c cell lines (SI Fig 23, SI Table 11). We also observe the effect of H1c on chromatin structure in microglia; cells homozygous for the H1c minor allele have a significant open chromatin peak at the rs242562 locus, which was not present in cells homozygous for the major allele (Fig 5D, SI Table 11).

These data suggest that the H1c haplotype alters the chromatin structure of regulatory regions in *MAPT* intron 1 that may be of particular relevance to microglia. Consistent with this notion, rs242562 is associated with CTCF binding in myeloid-lineage CD14+ monocytes and primary astrocytes in data derived from the ENCODE portal[52], but not in H1 neural cells (Fig 5E, SI Fig 23). Taken together, these data indicate that intron 1 of *MAPT* may be a regulatory region for both neuronal and microglial cell types, and therefore both neuronal and glial gene expression may be modified differentially by the presence of the H1c haplotype in a cell-type specific manner.

Discussion

PSP and PD are both neurodegenerative movement disorders genetically associated with the H1 haplotype of the 17q21.31 locus, but present with diverse neuropathological and clinical phenotypes. In order to understand how variation within this locus can underlie such heterogeneity, we tested whether there are distinct sub-haplotypes and variants within the H1 clade that are associated with each disease, and further elucidate the divergent functional impact of these variants on disease risk and pathogenesis.

We identified novel independent association signals for PSP risk within *MAPT* intron 1, in addition to replication of the H1c association[8,23,24]. Paired with the altered pattern of LD between this region and the rest of *MAPT* and *KANSL1*, this suggests that altered recombination or gene conversion across the regulatory *MAPT* intron 1 region and the rest of the locus could be contributing to PSP risk, rather than sub-haplotypic variation, as identified in PD. Furthermore, PSP-associated SNPs in *MAPT* intron 1 showed no significant association with PD.

We did not observe any association between PSP-associated variants and *MAPT* expression or splicing in human brain RNA-seq data, as has previously been reported[30,31]. However, despite low to no expression of *MAPT* in non-neuronal cell types, expression analyses from bulk tissue may impede the detection of genes expressed in specific cell populations. Unlike PD-associated sub-haplotype blocks, these variants also did not appear to be associated with CNVs within the 17q21.31 locus. We therefore assayed epigenetic data from human brain and from iPSC-derived neurons, astrocytes and microglia in order to elucidate the functional impact of these variants. As has previously been hypothesized, the H1c tag SNPs rs242557 and rs242562 are located within a regulatory region[26,28,29] characterized by histone modifications in multiple brain regions, as is our additional intron 1 association SNP, rs35838379. However, ATAC-seq from human brain and iPSC-derived neural cell types suggests that rather than regulating the expression of *MAPT* in neurons, as has been previously assumed[26,28,29], this may be a regulatory region for other genes within the 17q21.31 locus that are relevant for glial cell types.

Consistent with this idea, both H1c tag SNPs alter the chromatin structure of *MAPT* intron 1 by increasing accessibility in both neurons and microglia. The presence of a CTCF binding site in myeloid cells suggests that this may be a region of chromatin looping in microglia, supporting the assertion that *MAPT* intron 1 is an important regulatory region for non-neuronal cells, although neither rs242557 nor rs242562 minor alleles appear to modify a CTCF binding motif. The impact of chromatin structure variation on gene expression in specific neural sub-types remains to be elucidated, and will require analysis of cell-type specific expression data. However, given the data that is currently available, we propose that variants within the H1 haplotype clade, specifically within *MAPT* intron 1, are likely to contribute to the regulation of other genes in other cell types, either in contrast or in addition to the previously hypothesized regulation of *MAPT* in neurons[26,27].

We identified multiple sub-haplotype blocks that are significantly associated with both PD risk and protection. These blocks also showed an association with PSP risk and protection, but

importantly different sub-haplotypes within each block conferred risk for each disease. Together, this suggests that there is not true pleiotropy of the 17q21.31 locus for PD and PSP, but rather different variants within the same region underlie the H1 association with each disorder.

PD-associated sub-haplotypes did not appear to be associated with the expression or splicing of *MAPT* in human brain, which is consistent with Tau pathology not being a defining feature of PD pathogenesis[36]. However, they were associated with the expression of another gene in the 17q21.31 locus: *LRRC37A/2*. Analysis of CNVs in the 3' distal end of this locus suggested that these sub-haplotypes may be tagging structural variants defined by the gamma region[42] or *LRRC37A/2* copy number. Specifically, protective sub-haplotypes were associated with increased *LRRC37A/2* copy number and expression. Similarly, the protective H2 haplotype clade is also associated with increased *LRRC37A* expression and copy number, suggesting that there may be a shared mechanism of protection between H2 and specific sub-haplotypes of H1.

LRRC37A is a core duplicon on chromosome 17[46] and is present at the inversion breakpoint of the 17q21.31 locus; it has been hypothesized that its propensity for CNVs is responsible for the evolutionary toggling of this region that resulted in the distinct H1 and H2 haplotypes[53]. Due to the complex structural variation surrounding *LRRC37A* and the presence of its paralog *LRRC37A2*, it is challenging to genotype or sequence this region of the genome. As a consequence, this low quality genotype information has been excluded from GWAS analyses, so that the association between *LRRC37A/2* variants and any disease has never been tested. It is therefore likely that additional variation within *LRRC37A/2* itself is contributing to its altered expression and function, and therefore also contributing to PD risk in a more specific manner than we were able to identify here.

Very little is known about the function of *LRRC37A/2*, other than increased copy number has been associated with an increased antibody response to an Anthrax vaccine[54], and overexpression in HeLa cells appeared to promote the formation of filopodia[46], suggesting that *LRRC37A/2* may be involved in the immune and inflammatory response, as well as with cellular migration and synapse formation. Our RNA-seq and biochemical analyses of *LRRC37A/2* expression support these assertions; we find that LRRC37A is a membrane-bound protein that is likely to play a role in cellular migration and chemotaxis, and is expressed in both neurons and astrocytes. It is therefore possible that *LRRC37A/2* may contribute to the regulation of neuronal migration during development, as well as the migration of glia as part of an immune response within the brain. However, despite differences in copy number, the H1/H2 effect on *LRRC37A/2* expression and its associated genes is apparent primarily in astrocytes, suggesting that glial cells may be the most relevant cell type for *LRRC37A/2* function.

The role of glial dysregulation as a contributor to PD pathogenesis has gained attention in recent years. Neuroinflammation of the substantia nigra pars compacta is considered a characteristic feature of PD in addition to neuronal loss[55,56], and many genes associated with PD, such as *GBA*, *LRRK2* and *PINK1*, are expressed in astrocytes and are thought to have a role in glial functions such as inflammatory response, lipid handling, mitochondrial health and lysosomal function[57,58]. Aberrant glial function is therefore thought to be a potent contributory mechanism to the

development of PD and detrimental to the survival of surrounding neurons[58–60]. Altered glial function by variable levels of *LRRC37A/2* expression is therefore relevant to our current understanding of PD pathogenesis; the substantia nigra is considered to be particularly susceptible in PD as dopaminergic neurons in this region are surrounded by the lowest proportion of astrocytes in the brain[61]. Dysfunction of these few cells, and an inability to effectively recruit astrocytes from surrounding regions in the context of α-synuclein accumulation may contribute to neuronal death. The upregulation of chemotaxis and migration in both microglia and astrocytes in the context of increased *LRRC37A/2* expression may therefore allow for improved responsiveness to[62] and subsequent clearance of extracellular α-synuclein, as well as improved support of axonal growth in damaged neurons[60,63], resulting in enhanced neuronal resilience. Indeed, reactive astrocytosis appears absent in the substantia nigra of PD patients[64,65].

Regardless, the exact mechanism underlying *LRRC37A/2* protection against PD remains to be determined. Our current analyses used HEK293T cells as a cell culture system allowing for effective *LRRC37A/2* overexpression, however the role of *LRRC37A/2* in the brain will best be determined by cell-specific analyses, where we may observe alterations in neural pathways not present in HEK293T cells such as phagocytosis or synaptic development, that may also be relevant to the pathophysiology of PD. It is also unclear why haplotypes associated with increased *LRRC37A/2* expression may increase risk for PSP, given that the highly protective H2 haplotype is also associated with higher *LRRC37A/2* copy number and expression. We hypothesize that the sub-haplotype blocks we identify within the *MAPT* and *KANSL1* region may not have the same impact on *LRRC37A/2* expression and function in PSP brain as in PD or control brain, as the LD between this region and the distal ends of the locus is entirely lost in PSP. Consistent with this assertion, we do not find strong or consistent evidence for an association between PSP-associated *MAPT* intron 1 variants and *LRRC37A/2* expression in human brain.

In conclusion, PD and PSP are associated with different sub-haplotypes and variants of the 17q21.31 H1 haplotype clade that result in different functional outcomes and divergent mechanisms of disease pathogenesis, which are likely to underlie the distinct neuropathological and clinical phenotypes of each disease despite a shared genetic association with this locus. Although it has been previously assumed that the genetic association across the 17q21.31 locus is due solely to either *MAPT* or *KANSL1*, and therefore specific to neuronal function, we find evidence for the involvement of other genes within the locus and the implication of cell types other than neurons with PD and PSP. We have identified novel sub-haplotypic variants of H1 associated with PD and PSP, as well as additional association signals for PSP in *MAPT* intron 1, which are independent of the well-known H1c association. We propose novel mechanisms associated with each disease as a result of this variation in 17q21.31, both of which converge on glial function and regulation. By fine-mapping the H1 association with PD and PSP, we have begun to untangle the apparent pleiotropy of this locus, and gain better insight into the mechanism of each disease. We believe this will also be a useful endeavor for other tauopathies that share genetic risk at this locus.

Methods

Autopsy confirmed patient samples and genotyping

For Stage 2 PSP data, fresh frozen autopsy brain tissues of 455 cases (mean age 75.9yrs, 44.1% female) and 886 controls (mean age 83yrs, 52.4% female) were obtained from 21 brain banks (SI methods Table 4). Neuropathological workups were performed at the corresponding center. 368 cases were pure PSP, while 12 had co-morbid corticobasal degeneration (CBD). 56 had pure CBD, but were also included in the analysis given the close overlap and similarities between these two primary tauopathies. A subset of patients (19) had mixed pathology with co-occurring AD or Lewy body pathology. Controls were subjects that did not meet the criteria for any neurodegenerative condition beyond ubiquitous age-related tauopathy (neuronal and glial). High-throughput isolation of genomic DNA from fresh frozen brain tissue was performed using the KingFisher FLEX Magnetic Particle Processor (ThermoFisher Scientific) using the MagMAX DNA Multi-Sample Ultra 2.0 Kit (ThermoFisher Scientific). Genotyping for controls was carried out using the Illumina Infinium OmniExpress-24 v1.3 and cases were genotyped using the Illumina Infinium Global Screening Array v2.4. All post-mortem tissues were collected in accordance with the relevant guidelines and regulations of the respective intuitions.

Genotype data treatment

Case and control data from several cohorts from the International Parkinson's Disease Genetics Consortium (IPDGC; NIA, GER, FIN, NL, SP, McGill)[3,6,7] was kindly shared by Drs. Nalls, Singleton and Bandres-Ciga (NIH, Bethesda, MD; SI methods Tables 2,5). For the Stage 1 PSP cohort, genotype data from PSP cases[8] were kindly provided by Dr. Schellenberg (U. Penn, PA) and the PSP Genetics Consortium following approval from the National Institute on Aging Genetics of Alzheimer's Disease Data Storage Site (NIAGADS). Controls were obtained from the Alzheimer's Disease Genetics Consortium (ADGC); SI methods Tables 1, 6). ADGC data had previously been imputed against the 1000Genomes reference panel[66]; in order to obtain only those SNPs that had been originally genotyped, SNPs with an imputation $r^2 = 1$ were selected for further quality control (QC) and re-imputation.

Pre-imputation QC

Each dataset was obtained with different QC filters already applied, and so were all subsequently passed through the same, more stringent QC pipeline to ensure consistency. Plink v1.9[67] was used to perform quality control for all datasets. First, SNPs were filtered by a 98% call rate, and remaining SNPs with a MAF < 1% were excluded. Individuals with < 98% genotyping rate were then removed. To determine and correct for population stratification, principal components analysis was carried out in combination with Hapmap YRI, CEU and CHB populations[68] using EIGENSOFT[69]. Samples that did not cluster with the CEU European ancestry population were excluded. Identity by descent analysis was then conducted, and related individuals or potential sample duplicates ($Z0 \leq 0.8$) were removed. We were unable to assess discordant sex information on data acquired from other sources, as the required information for this analysis was not provided

to us. Variants that deviated from Hardy Weinberg equilibrium at a significance threshold < $1x10^{-4}$ were then removed. Chromosome 17 was then isolated and screened for strand mismatches.

Imputation and post-imputation QC

Filtered chromosome 17 data from each cohort was submitted individually to the Michigan imputation server[70] (https://imputationserver.sph.umich.edu) and imputed against the HRC r1.1 2016 panel using Eagle v2.3 phasing. Following imputation, SNPs with an r2 < 0.3 were removed, and the remaining SNPs were filtered for a 99% call rate. Genotyping call rates for individuals were again filtered at a 99%, and SNPs that deviated from Hardy-Weinberg equilibrium at a significance threshold < $1x10^{-6}$ were excluded. Individual cohorts were then merged, and finally filtered once more with a SNP call rate of 99%. Prior to analysis, variants were filtered to exclude SNPs with a MAF < 0.01.

Single SNP association analyses

Association analyses were carried out in SNP and Variation Suite v8.8.1 (SVS8) software (Golden Helix, Inc., Bozeman, MT, www.goldenhelix.com). Association analyses using logistic regression with an additive model were carried out. As all potential covariate information was not available, the model was corrected using the first 10 principal components as calculated by SVS8. However, similar to previous PSP analyses[8], the principal components were not adjusted for in PSP H1/H2 comparisons, as this resulted in over-correction and removal of the H1/H2 association signal, but were included in the PSP H1 homozygote analysis. For each dataset, associations were initially carried out on the entire cohort in order to confirm the 17q21.31 H1/H2 haplotype association. The data was then filtered for H1 homozygotes only, using tag SNP rs8070723 and the association analysis was repeated with the same parameters.

Meta-analysis of SNP effects across multiple datasets was carried out using the R package rmeta[71] using both Random Effects (DerSimonian-Laird) and Fixed Effects (Mantel-Haenszel) approaches. Calculation and visualization of linkage disequilibrium (LD) over large genomic ranges was carried out in SVS8 using both r2 and D'. Inspection of LD between individual SNPs of interest was carried out using Haploview[72].

For the PSP analysis, due small sample sizes, data from Stage 1 and Stage 2 were combined and re-analyzed. The first ten principal components were re-computed on the merged data and were included in the combined analysis. To ensure both control datasets had equivalent minor allele frequencies across the 17q21.31 locus and were therefore suitable for merging, we carried out an association analysis as described above, using "Stage" as the outcome variable, and found no significantly variable SNPs between cohorts (SI Fig 24). In addition, we found that the MAFs across this locus were highly and significantly correlated between stages ($r^2 = 0.993$. $p < 2x10^{16}$; SI Fig 24).

Conditional analysis

Conditional analysis was carried out in Plink v1.9 using the logistic flag, and regional plots of the resulting association signals were visualized using LocusZoom

(http://csg.sph.umich.edu/locuszoom). Signal was considered to be retained following conditioning when SNPs passed the suggestive p-value threshold of $1x10^{-5}$.

Haplotype block construction and association

Haplotype blocks were constructed in SVS8 using the D' measure of LD. Blocks were defined using guidelines as described by Gabriel et al (2002)[73]; each block contained a maximum of 15 markers within 160kb of each other, with a D' upper confidence bound ≥ 0.98 and a lower confidence bound ≥ 0.7. Haplotypes were estimated using an expectation-maximization (EM) algorithm with 50 iterations, and a convergence tolerance of 0.0001. Sub-haplotypes with a frequency < 0.01 were excluded from further analysis. Case-control association analyses were carried out per block using a logistic regression model. Odds ratios and associated Fisher's exact p-values were calculated for each sub-haplotype within each block using the R package epitools[74].

Human brain expression analysis

Publicly available RNA-seq expression data from human postmortem prefrontal (PFC) and temporal (TCX) cortices (SI methods Table 3) and associated genotype data were obtained from Synapse (synapse.org; The Religious Orders Study and Memory and Aging Project (ROSMAP) syn3219045; MayoRNAseq syn5550404; CommonMind Consortium syn2759792). Genotype data for chromosome 17 underwent the same QC and imputation pipeline as described above. Data were stratified by 17q21.31 H1/H2 haplotype using the H2 tag SNP rs8070723, and H1 homozygotes were then grouped by genotype for each individual SNP of interest from the PSP analysis. For sub-haplotype analysis, blocks previously defined in the PD analysis were applied to the genotype data and haplotypes were estimated in the same manner. Statistical analysis was carried out in R version 3.4.0. For analysis of *MAPT* splicing, percent spliced in (PSI) values were generated for exons 2, 3 and 10 using Mixture of Isoforms (MISO)[75]. Gene expression and PSI residuals were generated by linear regression using sex, age of death, post-mortem interval and RNA integrity score as covariates. The resulting residuals were then transformed into z-scores and combined across datasets. Statistical differences in gene expression between genotypes and sub-haplotypes were determined by linear regression applied to the adjusted and combined z-scores.

dPCR

Human genomic DNA and accompanying genotype data was kindly provided by Drs. Raj, Crary and Charney (Mount Sinai School of Medicine, NY) and by the Alzheimer's Disease Research Center (ADRC; SI Methods Table 3). Sub-haplotypes were called from these genotype data in the same manner as described above. To examine copy number variation in the 17q21.31 locus, digital PCR was carried out using the ThermoFisher QuantStudio 3D digital PCR chip system. Taqman dPCR probes for loci within the alpha, beta and gamma CNV regions[42], as well as within *LRRC37A* and *MAPT* were selected for analysis (SI Methods Table 7).

Cell lines

Human induced pluripotent stem cells (iPSCs) were obtained from the Knight Alzheimer's Disease Research Center at Washington University[76], the NIH Childhood-onset Schizophrenia study[77], and

the UCI ADRC (SI Methods Table 8). The Icahn School of Medicine at Mount Sinai IRB reviewed the relevant operating protocols as well as this specific study and determined it was exempt from approval.

Cell culture

Unless otherwise specified, all cell culture materials were obtained from ThermoFisher Scientific. Human embryonic kidney cells (HEK293T) were cultured in Dulbecco's Modified Eagle Medium/F-12 with HEPES, supplemented with 10% fetal bovine serum (FBS) and 1% Penicillin-Streptomycin. Cells were passaged every 3-4 days using Trypsin-EDTA (0.25%). For *LRRC37A* overexpression experiments, HEK293T cells were seeded at a density of $1.4x10^5$ cells per well in 6-well plates and transfected with 0.5-2.5ug of *LRRC37A* plasmid (Origene) or empty vector control (Origene) using Lipofectamine 3000. Cells were harvested 48 hours after transfection.

For qRTPCR and protein biochemistry experiments, iPSC lines (SI Methods Table 8) were maintained in complete StemFlex media supplemented with 1% penicillin/streptomycin on Matrigel (BD biosciences), and were differentiated to neural progenitor cells (NPCs) as previously described[78]. Forebrain neuron-enriched cultures and astrocyte cultures were differentiated from NPCs as previously described[78,79]. Neuronal and astrocytic identity was confirmed by immunofluorescence for common neuronal and astrocytic markers (MAP2 (Abcam), Tuj1 (Cell Signaling Technologies), S100β (Sigma Aldrich) and EAAT1 (Abcam); SI Fig 25).

Genomic DNA was extracted using the DNeasy Blood and Tissue kit (Qiagen) and underwent genotyping with Taqman assays for H2 tag SNPs rs8070723 and rs1052553 in order to confirm 17q21.31 haplotype (SI Methods Table 8).

ATAC-seq

ATAC-seq was performed using an established protocol[80] with minor modifications. Following lysis, an aliquot of iPSC-derived astrocyte, neuron and microglial nuclei were stained with DAPI (Thermoscientific) and counted (Countess II FL automated cell counter, Life Technologies). ATAC-seq was performed using 75,000 nuclei each, and the resulting libraries sequenced at the New York Genome Center as 50bp paired-end reads on a HiSeq 2500.

Each set of pair-end reads was aligned by STAR (v2.5.0)[81] to the hg19 reference genome with the pseudoautosomal region masked on chromosome Y using the following parameters:

--alignIntronMax 1 –out FilterMismatchNmax 100 --alignEndsType EndToEnd --out FilterScoreMinOverLread 0.3 –out FilterMatchNminOverLread 0.3

This yielded for each sample a BAM file of mapped paired-end reads sorted by genomic coordinates. From these BAM-files, reads that mapped to multiple loci or to the mitochondrial genome were removed using samtools[82] and duplicated reads were removed with PICARD (v2.24; http://broadinstitute.github.io/picard). The read coverage files (bedGraph) were generated by genomeCoverageBed from BEDTools (v2.25.0)[83] and converted to the indexed binary format files

(bigWig) by bedGraphToBigWig. To complete per-sample quality control, we called peaks for each BAM file with MACS2[84] using the following parameters:

--keep-dup all --shift -100 --extsize 200 --nomodel

Individual peaks were merged across all samples, retaining only peaks found in at least 2 samples. Subsequently, we used featureCounts from Rsubread package[85] to calculate the fraction of reads that overlap consensus peaks (FRiP metrics). Since all samples passed our internal QC thresholds (i.e., FRiP > 5%; number of non-duplicated uniquely-mapped peaks > 5,000,000; ratio of uniquely mapped reads > 0.4, we decided not to exclude any of them (SI Methods Table 9).

To create final peak sets of open chromatin, bigWig files of samples with the same *MAPT* H1c haplotype per each iPSC-derived cell line were merged, and peaks within the *MAPT* gene were called using the following command in MACS2:

bdgpeakcall -i sample.bedgraph --outdir bdgpeaks -o sample_CellType

qRT-PCR

RNA was extracted from HEK293T cells, iPSC-derived neurons, astrocytes and human brain tissue using the RNeasy Mini kit (Qiagen) and reverse transcribed using the High-Capacity RNA-to-cDNA kit (ThermoFisher Scientific). Gene expression was measured by Taqman qRTPCR assays listed in SI Methods Table 10.

RNA-seq

RNA was prepared as described above. Library preparation with poly-A selection and sequencing with 150 base pair paired-end reads was carried out at Genewiz. Sequenced reads were trimmed for Illumina TruSeq adapters, and quantified for gene expression values in TPM (Transcripts Per Kilobase Million) using Salmon[86] guided by the GENCODE human transcriptome model (GRCh38 version 28, Ensembl 92). TPM data was imported into the R (version 3.5.1) programming environment for visualization and analysis, and differential expression of *LRRC37A* overexpression compared to the control was analyzed using the moderated t-test implemented in limma[87]. Differentially expressed genes (DEGs) were defined by $\pm \geq 1.5$ expression fold change and adjusted $p < 0.05$. Gene set enrichment analysis was performed with the Broad Institute's MSigDB annotations[88]. Analysis of GO enrichment terms was carried out using g:Profiler (https://biit.cs.ut.ee/gprofiler/gost)[89,90] and visualized using Cytoscape v3.7.1[91] with the EnrichmentMap[92] plugin. Additional pathway analyses were carried out using Ingenuity Pathway Analysis (QIAGEN Inc., https://www.qiagenbioinformatics.com/products/ingenuitypathway-analysis) using genes with a fold change $\pm \geq 1$.

Protein biochemistry

Membrane and cytosolic proteins were isolated from HEK293T cells, iPSC-derived neurons and iPSC-derived astrocytes using the MEM-PER Plus Membrane Protein Extraction Kit (ThermoFisher Scientific), and protein concentrations were determined by bicinchoninic acid (BCA) assay (ThermoFisher Scientific). For western blotting, protein fractions were subject to

SDS-PAGE electrophoresis through BOLT Bis-Tris gels (ThermoFisher Scientific) and were blotted onto nitrocellulose membranes. Membrane fractions were confirmed by labelling with an anti-pan-Cadherin antibody (Cell Signaling Technology), and cytosolic fractions were confirmed by labelling with anti-HSP90 (Cell Signaling Technology). Membranes were stripped using Restore plus western blot stripping buffer and re-probed with an anti-LRRC37A antibody (ThermoFisher Scientific).

For immunofluorescence, cells were fixed for 15 minutes with Formalin (Sigma Aldrich) and washed with phosphate buffered saline (PBS). Cells were then permeabilized for 30 minutes in 0.1% Triton x-100 in PBS and blocked for an additional 30 minutes in 1% bovine serum albumin (BSA) in PBS. Primary antibodies were diluted in blocking buffer and incubated with the cells at 4°C overnight. Cells were then incubated with secondary AlexaFluor 488 and 568 for two hours at room temperature, before being labelled for 10 minutes with DAPI and storage in PBS. Cells were imaged on a Leica DMIL LED Inverted Routine Fluorescence microscope.

Publicly available epigenetic data analysis

Histone methylation and acetylation CHIP-seq data from human brain (E071, E074, E068, E069, E072, E067, E073, E070) were generated by the Roadmap Epigenomics Project[47] and visualized in the WashU Epigenome browser[93]. Human brain NeuN+ and NeuN- ATAC-seq data is part of the Brain Open Chromatic Atlas developed by Fullard et al. 2018[51] and was downloaded from https://icahn.mssm.edu/boca. The following CTCF CHIP-seq data was downloaded from the Encode Portal[52]; experiment IDs ENCSR822CEA, ENCSR000AOO and ENCSR000ATN. BigWig files were visualized in R v3.4.0 using TxDb.Hsapiens.UCSC.hg19.knownGene annotation[94] and Gviz[95].

Data sharing

All sequencing files and processed peaks for iPSC-derived ATAC-Seq neurons, astrocytes and microglia, as well as aligned read counts and FASTQ files for *LRRC37A*-overexpressing HEK293T cells will be deposited to the Gene Expression Omnibus once the manuscript is accepted for publication.

Acknowledgements

This work was supported by funding from the BrightFocus Foundation (KRB), Association for Frontotemporal Degeneration (KRB) and CurePSP (KRB). The recruitment and clinical characterization of research participants at Washington University were supported by NIH P50 AG05681, P01 AG03991, and P01 AG026276. NIH AG062683 (J.TCW), Rainwater Charitable Organization (CMK), NIH AG046374 (CMK). The McGill cohort was supported by grants from the Michael J. Fox Foundation, the Canadian Consortium on Neurodegeneration in Aging (CCNA), the Canada First Research Excellence Fund (CFREF), awarded to McGill University for the Healthy Brains for Healthy Lives (HBHL) program, and Parkinson Canada. ZGO is supported by the Fonds de recherche du Québec - Santé (FRQS) Chercheurs-boursiers award given with Parkinson Quebec, and is a Parkinson Canada New Investigator awardee. The access to the McGill

participants for this research has been made possible in part thanks to the Quebec Parkinson's Network (http://rpq-qpn.ca/en/). NIH [R01AG054008, R01NS095252, R01AG060961, R01 AG060961to JFC, and F32AG056098 to KF], the Rainwater Charitable Foundation/Tau Consortium (JFC), Genentech/Roche, and an Alexander Saint-Amand Scholarship (JFC). We thank and acknowledge the Alzheimer's disease genetics consortium (ADGC; NIA U01AG032984). Data for this study were prepared, archived and distributed by the National Institute on Aging Alzheimer's Disease Data Storage Site (NIAGADS) at the University of Pennsylvania (U24-AG041689-01), funded by the National Institute on Aging. This work was funded by grants from the CurePSP Foundation, the Peebler PSP Research Foundation. Günter Höglinger was funded by the Deutsche Forschungsgemeinschaft (German Research Foundation) within the framework of the Munich Cluster for Systems Neurology (EXC 2145 SyNergy—ID 390857198), Deutsche Forschungsgemeinschaft (DFG, HO2402/18-1 MSAomics), the German Federal Ministry of Education and Research (BMBF, 01KU1403A EpiPD; 01EK1605A HitTau), the NOMIS foundation (FTLD project).

Author Contributions

Conceptualization: KRB, JFC, AMG. *Methodology*: KRB. *Validation:* KRB, DAP. *Formal Analysis:* KRB, DAP, KF, NH, YL, JB, JFF, AER, SB-C, ZG-O, PH, AS, SB. *Investigation*: KRB, DAP, KF, NH, YL. *Resources:* JTCW, SAL, LQ, AC, MAI, JB, JFF, CMK, SJF, BHK, IP, YJP, AC, PR, PKC, KLB, JSKK, GUH, JCW, WWP, TR, JFC, AMG. *Data Curation*: KRB, KF, NH, YL, JB, JFF, SER, SB-C, ZG-O, PH, AS, SB, PART working group, IPDGC, PSP genetics consortium. *Writing – Original Draft:* KRB. *Writing – Review and Editing*: KRB, KF, JTCW, YL, JB, JFF, AER, SB-C, ZG-O, CMK, IP, GUH, JCW, WWP, TR, JFC, AMG. *Visualization:* KRB. *Supervision*: AMG. *Funding Acquisition:* KRB, JFC, AMG

References

1. Jun, G. *et al.* a Novel Alzheimer Disease Locus Located Near the gene encoding tau protein. *Molecular Psychiatry.* **21**, 108–117 (2016).

2. Kouri, N. *et al.* Genome-wide association study of corticobasal degeneration identifies risk variants shared with progressive supranuclear palsy. *Nat. Commun.* **6**, 1–7 (2015).

3. Bandrés-ciga, S. *et al.* Genome-wide assessment of Parkinson ' s disease in a Southern Spanish population. *Neurobiol. Aging* **45**, 213.e3-213.e9 (2016).

4. Desikan, R. S. *et al.* Genetic overlap between Alzheimer ' s disease and Parkinson ' s disease at the MAPT locus. *Mol. Psychiatry* 1588–1595 (2015). doi:10.1038/mp.2015.6

5. Nalls, M. A. *et al.* Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. *Nat. Genet.* **46**, 989–993 (2014).

6. Bandres-Ciga, S. *et al.* The genetic architecture of Parkinson disease in Spain: characterizing population-specific risk, differential haplotype structures, and providing etiologic insight. *Mov. Disord.* (2019). doi:10.1101/609016

7. Nalls, M. A. *et al.* Identification of novel risk loci , causal insights , and heritable risk for Parkinson ' s disease : a meta-analysis of genome-wide association studies. *Lancet Neurol.* 1091–1102 (2019). doi:10.1016/S1474-4422(19)30320-5

8. Höglinger, G. U. *et al.* Identification of common variants influencing risk of the tauopathy progressive supranuclear palsy. *Nat. Genet.* **43**, 699–705 (2011).

9. Chen, J. A. *et al.* Joint genome-wide association study of progressive supranuclear palsy identifies novel susceptibility loci and genetic correlation to neurodegenerative diseases. *Mol. Neurodegener.* **13**, 1–11 (2018).

10. Pastor, P. *et al.* Novel Haplotypes in 17q21 Are Associated with Progressive Supranuclear Palsy. *Ann. Neurol.* **56**, 249–258 (2004).

11. Hoglinger, G. U. *et al.* Clinical Diagnosis of Progressive Supranuclear Palsy : The Movement Disorder Society Criteria. *Mov. Disord.* **32**, 853–864 (2017).

12. Kovacs, G. G. Invited review : Neuropathology of tauopathies : principles and practice. *Neuropathol. Appl. Neurobiol.* **41**, 3–23 (2015).

13. Dickson, D. W. Neuropathologic differentiation of progressive supranuclear palsy and corticobasal degeneration. *J Neurol* **246**, 6–15 (1999).

14. Litvan, I. *et al.* Which clinical features differentiate progressive supranuclear palsy (Steele – Richardson – Olszewski syndrome) from related disorders ? A clinicopathological study. *Brain* **120**, 65–74 (1997).

15. Caso, F. *et al.* Parkinsonism and Related Disorders Cognitive impairment in progressive supranuclear palsy-Richardson ' s syndrome is related to white matter damage. *Park. Relat. Disord.* **31**, 65–71 (2016).

16. Gerstenecker, A., Duff, K., Mast, B. & Litvan, I. Behavioral abnormalities in progressive supranuclear palsy. *Psychiatry Res.* **210**, 1205–1210 (2013).

17. Poorkaj, P. *et al.* An R5L τ mutation in a subject with a progressive supranuclear palsy phenotype. *Ann. Neurol.* **52**, 511–516 (2002).

18. Coppola, G. *et al.* Evidence for a role of the rare p.A152T variant in mapt in increasing the risk for FTD-spectrum and Alzheimer's diseases. *Hum. Mol. Genet.* **21**, 3500–3512 (2012).

19. Kovacs, G. G. *et al.* Unclassifiable tauopathy associated with an A152T variation in MAPT exon 7. *Clin. Neuropathol.* **30**, 3–10 (2011).

20. Ros, R. *et al.* A New Mutation of the Tau Gene, G303V, in Early-Onset Familial Progressive Supranuclear Palsy. *Arch Neurol* **62**, (2005).

21. Stanford, P. M. *et al.* Progressive supranuclear palsy pathology caused by a novel silent mutation in exon 10 of the tau gene Expansion of the disease phenotype caused by tau gene mutations. *Brain* 880–893 (2000).

22. Skoglund, L. *et al.* The tau S305S mutation causes frontotemporal dementia with

parkinsonism. *Eur. J. Neurol.* **15**, 156–161 (2008).

23. Pittman, A. *et al.* Linkage disequilibrium fine mapping and haplotype association analysis of the tau gene in progressive supranuclear palsy and corticobasal degeneration. *J Med Genet* **42**, 837–846 (2005).

24. Zhang, C. *et al.* Meta-analysis of the association between variants in MAPT and neurodegenerative diseases. *Oncotarget* **8**, 44994–45007 (2017).

25. Heckman, M. G. *et al.* Association of MAPT Subhaplotypes With Risk of Progressive Supranuclear Palsy and Severity of Tau Pathology. *JAMA Neurol.* 1–8 (2019). doi:10.1001/jamaneurol.2019.0250

26. Myers, a. J. *et al.* The H1c haplotype at the MAPT locus is associated with Alzheimer's disease. *Hum. Mol. Genet.* **14**, 2399–2404 (2005).

27. Majounie, E. *et al.* Variation in tau isoform expression in different brain regions and disease states. *Neurobiol. Aging* **34**, 1922.e7-1922.e12 (2013).

28. 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).

29. Anaya, F., Lees, A. & de Silva, R. Tau gene promoter rs242557 and allele-specific protein binding. *Transl. Neurosci.* **2**, 176–205 (2011).

30. Hayesmoore, J. B. G. *et al.* The effect of age and the H1c MAPT haplotype on MAPT expression in human brain. *Neurobiol. Aging* **30**, 1652–1656 (2009).

31. Trabzuni, D. *et al.* MAPT expression and splicing is differentially regulated by brain region: Relation to genotype and implication for tauopathies. *Hum. Mol. Genet.* **21**, 4094–4103 (2012).

32. Galvin, J. E., Pollack, J. & Morris, J. C. Clinical phenotype of Parkinson disease dementia. *Neurology* **67**, 1605–1612 (2006).

33. Aarsland, D., Zaccai, J. & Brayne, C. A Systematic Review of Prevalence Studies of Dementia in Parkinson ' s Disease. *Mov. Disord.* **20**, 1255–1263 (2005).

34. Massano, J. & Bhatia, K. P. Clinical Approach to Parkinson's Disease: Features, Diagnosis, and Principles of Management. *Cold Spring Harb. Perspect. Med.* 1–15 (2012).

35. Spillantini, M. G. *et al.* alpha-Synuclein in Lewy bodies. *Nature* **388**, 839–840 (1997).

36. Dickson, D. W. Neuropathology of Parkinson Disease. *Parkinsonism Relat. Disord.* **46**, 1–11 (2018).

37. Zhang, X. *et al.* Tau Pathology in Parkinson ' sDisease. *Front. Neurol.* **9**, 1–7 (2018).

38. Espay, A. J. *et al.* Revisiting protein aggregation as pathogenic in sporadic Parkinson and Alzheimer diseases. *Neurology* **92**, 329–337 (2019).

39. Nalls, M. A. *et al.* Expanding Parkinson's disease genetics: novel risk loci, genomic context, causal insights and heritable risk. *bioRxiv* (2019). doi:http://dx.doi.org/10.1101/388165

40. Zabetian, C. P. *et al.* Association Analysis of MAPT H1 Haplotype and Subhaplotypes in Parkinson's Disease. *Ann. Neurol.* **62**, 137–144 (2007).

41. Vandrovcova, J. *et al.* Association of MAPT haplotype-tagging SNPs with sporadic Parkinson's disease. *Neurobiol. Aging* **30**, 1477–1482 (2009).

42. Boettger, L. M., Handsaker, R. E., Zody, M. C. & McCarroll, S. A. Structural haplotypes and recent evolution of the human 17q21.31 region. **44**, 881–885 (2012).

43. Steinberg, K. M. *et al.* Structural diversity and African origin of the 17q21 . 31 inversion polymorphism. *Nat. Publ. Gr.* **44**, 872–880 (2012).

44. Naj, A. C. *et al.* Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. *Nat. Genet.* **43**, 436–443 (2011).

45. Veerappa, A. M., Saldanha, M., Padakannaya, P. & Ramachandra, N. B. Family Based Genome-Wide Copy Number Scan Identifies Complex Rearrangements at 17q21.31 in Dyslexics. *Am. J. Med. Genet. PART B-NEUROPSYCHIATRIC Genet.* **165**, 572–580 (2014).

46. Giannuzzi, G. *et al.* Evolutionary dynamism of the primate LRRC37 gene family. *Genome Res.* **23**, 46–59 (2013).

47. Roadmap Epigenomics Consortium *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–329 (2015).

48. Creyghton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci.* **107**, 21931–21936 (2010).

49. Local, A. *et al.* Identification of H3K4me1-associated proteins at mammalian enhancers. *Nat. Genet.* **50**, 73–82 (2018).

50. Gates, L. A. *et al.* Acetylation on histone H3 lysine 9 mediates a switch from transcription initiation to elongation. *J. Biol. Chem.* **292**, 14456–14472 (2017).

51. Fullard, J. F. *et al.* An Atlas of Chromatin Accessibility in the Adult Human Brain. *GENOME Res.* (2018).

52. Davis, C. A. *et al.* The Encyclopedia of DNA elements (ENCODE): Data portal update. *Nucleic Acids Res.* **46**, D794–D801 (2018).

53. Zody, M. C. *et al.* Evolutionary toggling of the MAPT 17q21.31 inversion region. *Nat. Genet.* **40**, 1076–1083 (2008).

54. Falola, M. I. *et al.* Genomic Copy Number Variants: Evidence for Association with Antibody Response to Anthrax Vaccine Adsorbed. *PLoS One* **8**, (2013).

55. Miklossy, J. *et al.* Role of ICAM-1 in persisting inflammation in Parkinson disease and

MPTP monkeys. *Exp. Neurol.* **197**, 275–283 (2006).

56. Koprich, J. B., Reske-Nielsen, C., Mithal, P. & Isacson, O. Neuroinflammation mediated by IL-1β increases susceptibility of dopamine neurons to degeneration in an animal model of Parkinson's disease. *J. Neuroinflammation* **5**, 1–12 (2008).

57. Zhang, Y. *et al.* Purification and characterization of progenitor and mature human astrocytes reveals transcriptional and functional differences with mouse. *Neuron* **89**, 37–53 (2016).

58. Booth, H. D. E., Hirst, W. D. & Wade-Martins, R. The Role of Astrocyte Dysfunction in Parkinson's Disease Pathogenesis. *Trends Neurosci.* **40**, 358–370 (2017).

59. di Domenico, A. *et al.* Patient-Specific iPSC-Derived Astrocytes Contribute to Non-Cell-Autonomous Neurodegeneration in Parkinson's Disease. *Stem Cell Reports* **12**, 213–229 (2019).

60. Joe, E.-H. *et al.* Astrocytes, Microglia, and Parkinson's Disease. *Exp. Neurobiol.* **27**, 77 (2018).

61. Mena, M. A. & García De Yébenes, J. Glial cells as players in parkinsonism: The 'good,' the 'bad,' and the 'mysterious' glia. *Neuroscientist* **14**, 544–560 (2008).

62. Wang, S. *et al.* α-Synuclein, a chemoattractant, directs microglial migration via H 2 O 2 -dependent Lyn phosphorylation . *Proc. Natl. Acad. Sci.* **112**, E1926–E1935 (2015).

63. Tom, V. J., Doller, C. ., Malouf, A. . & Silver, J. Astrocyte-Associated Fibronectin Is Critical for Axonal Regeneration in Adult White Matter. *J. Neurosci.* **24**, 9282–9290 (2004).

64. Mirza, B., Hadberg, H., Thomsen, P. & Moos, T. The absence of reactive astrocytosis is indicative of a unique inflammatory process in Parkinson's disease. *Neuroscience* **95**, 425–432 (1999).

65. Tong, J. *et al.* Low Levels of Astroglial Markers in Parkinson's Disease: Relationship to α-Synuclein Accumulation. *Neurobiol. Dis.* **82**, 243–253 (2015).

66. Boehme, K. L. *et al.* ADGC 1000 Genomes combined data workflow *(electronic document).* 1–12 (2014) http://kauwelab.byu.edu/Portals/22/adgc_combined_1000G_12032014.pdf

67. 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).

68. Consortium, T. I. H. The International HapMap Project. *Nature* **426**, 789–796 (2003).

69. Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).

70. Das, S. *et al.* Next-generation genotype imputation service and methods Sayantan. **48**, 1284–1287 (2016).

71. Lumley, T. rmeta: Meta-Analysis. R package version 3.0. https://CRAN.R-project.org/package=rmeta (2018).

72. Barrett, J. C., Fry, B., Maller, J. & Daly, M. J. Haploview: analysis and visualization of LD and haplotype maps. *BIOINFORMATICS* **21**, 263–265 (2005).

73. Gabriel, S. B. *et al.* The structure of haplotype blocks in the human genome. *Science (80-.).* **296**, 2225–9 (2002).

74. Aragon, T. J. epitools: Epidemiology Tools. R package version 0.5-10. Https://CRAN.R-project.org/package=epitools (2017).

75. Katz, Y., Wang, E. T., Airoldi, E. M. & Burge, C. B. Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nat. Methods* **7**, 1009–1015 (2010).

76. Karch, C. M. *et al.* A Comprehensive Resource for Induced Pluripotent Stem Cells from Patients with Primary Tauopathies. *Stem cell reports* **13**, (2019).

77. Hoffman, G. E. *et al.* Transcriptional signatures of schizophrenia in hiPSC-derived NPCs and neurons are concordant with post-mortem adult brains. *Nat. Commun.* **8**, (2017).

78. Bowles, K. R., Julia, T. C. W., Qian, L., Jadow, B. M. & Goate, A. M. Reduced variability of neural progenitor cells and improved purity of neuronal cultures using magnetic activated cell sorting. *PLoS One* **14**, 1–18 (2019).

79. TCW, J. *et al.* An Efficient Platform for Astrocyte Differentiation from Human Induced Pluripotent Stem Cells. *Stem Cell Reports* **9**, 600–614 (2017).

80. Buenrostro, J., Wu, B., Chang, H. & Greenleaf, W. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Curr. Protoc. Mol. Biol.* **109**, 1–10 (2016).

81. Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

82. Li, H. *et al.* The Sequence Alignment / Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

83. Quinlan, A. R. & Hall, I. M. BEDTools : a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).

84. Zhang, Y. *et al.* Open Access Model-based Analysis of ChIP-Seq (MACS). *Genome Biol.* (2008). doi:10.1186/gb-2008-9-9-r137

85. Liao, Y., Smyth, G. K. & Shi, W. The Subread aligner : fast , accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res.* **41**, (2013).

86. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon: fast and bias-aware quantification of transcript expression using dual-phase inference. *Nat. Methods* **14**, 417 (2017).

87. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing

and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).

88. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15545–50 (2005).

89. Reimand, J., Kull, M., Peterson, H., Hansen, J. & Vilo, J. G:Profiler-a web-based toolset for functional profiling of gene lists from large-scale experiments. *Nucleic Acids Res.* **35**, 1–8 (2007).

90. Raudvere, U. *et al.* g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res.* 1–8 (2019). doi:10.1093/nar/gkz369

91. Shannon, P. *et al.* Cytoscape : A Software Environment for Integrated Models of Biomolecular Interaction Networks Cytoscape : A Software Environment for Integrated Models of Biomolecular Interaction Networks. *GENOME Res.* 2498–2504 (2003). doi:10.1101/gr.1239303

92. Merico, D., Isserlin, R., Stueker, O., Emili, A. & Bader, G. D. Enrichment map: A network-based method for gene-set enrichment visualization and interpretation. *PLoS One* **5**, (2010).

93. Zhou, X. *et al.* Epigenomic annotation of genetic variants using the Roadmap EpiGenome Browser. **33**, 345–346 (2015).

94. Carlson, M. & Maintainer, B. P. TxDb.Hsapiens.UCSC.hg19.knownGene: Annotation package for TxDb object(s). (2015).

95. Hahne, F. & Ivanek, R. Visualizing Genomic Data Using Gviz and Bioconductor. in *Statistical Genomics. Methods in Molecular Biology* (eds. Mathe, E. & Davis, S.) (HUMANA PRESS INC, 2016). doi:https://doi.org/10.1007/978-1-4939-3578-9_16

Figure Legends

Figure 1. Multiple H1 signals are associated with PSP risk and are associated with altered LD. A. The17q21.31 locus confers two distinct sub-haplotypes defined by gross structural inversion; H1 and H2. Direction of gene orientation in each haplotype is indicated by arrows. Each gene or partial gene is labeled with a distinct color and connected with a crossed rectangle between H1 and H2 to aid visualization of altered gene position between haplotypes. **B.** –log10 regression p-values of SNP association with PSP in H1 homozygotes in merged Stage 1 and Stage 2 data spanning the 17q21.31 locus. Black dotted line indicates genome-wide suggestive p-value of $1x10^{-5}$, and grey dotted line indicates genome-wide significant p-value of $5x10^{-8}$. **C.** Patterns of D' LD across the 17q21.31 locus in Stage 1 and Stage 2 merged data, split by control (top) and PSP (bottom). Red indicates high D' and strong LD, blue indicates low D' and no LD. Aligned to Refseq gene annotation.

Figure 2. H1 sub-haplotypes within the *MAPT* 17q21.21 inversion region are associated with Parkinson's disease risk. A-B. H1 sub-haplotype block association (-log10 p-value) with PD plotted above H1

homozygote D' LD structure and sub-haplotype blocks generated from ***A***. Stage 1 data and ***B.*** Stage 2 data, spanning Hg19 Chr17:43384997-44913630. **B**. Plotted in orange is the association (-log10 p-value) of exact blocks H1.1, H1.2 and H1.3 (as defined in Stage 1 analysis) with PD in Stage 2 data. In LD plots, red indicates high D' and blue indicates low. Black arrows indicate similar blocks generated across Stage 1 and Stage 2 data. Dotted black lines indicate genome wide suggestive significance p-value of $1x10^{-5}$.

Figure 3. PD-associated sub-haplotypes are associated with *LRRC37A/2* expression and copy number. A-C. Expression of *LRRC37A* in human brain, measured by RNA-seq across three different cohorts (SI Methods Table 6), split by sub-haplotype in blocks ***A.*** H1.1, ***B***. H1.2 and ***C.*** H1.3. **D-F.** qRTPCR for *LRRC37A* in human PFC brain tissue for sub-haplotypes across blocks ***D.*** H1.1, ***E.***H1.2 and ***F.*** H1.3. G. Schematic of the regions of copy number variation in the 3' distal end of the 17q21.31 locus, as defined by Boettger et al 2012[37]. Black arrows indicate the location of dPCR probes for (left to right) beta, alpha, *LRRC37A* and gamma. **H-K**. Copy number of ***H***. alpha, ***I.*** beta, ***J***. gamma and ***K***. *LRRC37A* in individuals homozygous for sub-haplotypes in block H1.1. All statistical comparisons are against the most common sub-haplotype. ns = not significant, *p < 0.05, **p < 0.01, ***p < 0.001.

Figure 4. LRRC37A is a membrane associated protein that plays a role in cellular migration and communication. A. Enriched GO terms for significantly differentially expressed genes following *LRRC37A* overexpression in HEK293T cells. Paler node colors indicate less significant enrichment p-values, and edge thickness indicates the proportion of shared genes between GO terms. **B.** Western blots for LRRC37A in cytosolic (C) and membrane (M) fractions from HEK293T cells overexpressing *LRRC37A* or an empty vector control. Cytosolic fractions were confirmed by labelling with an anti-HSP90 antibody, and membrane fractions were confirmed by labelling with an anti-Pan-Cadherin antibody, N=3. **C-D.** Significantly enriched pathways associated with *LRRC37A* overexpression, derived from Ingenuity pathway analysis. Red genes indicate upregulation, green indicates downregulation. **E.** Normalized enrichment scores for top enriched pathways following *LRRC37A* overexpression using GSEA. Deeper reds specify lower associated p-values. **F.** qRTPCR for *LRRC37A* expression in H1 and H2 homozygote iPSC-derived neurons and astrocytes. N=4-5. ns = not significant, *p < 0.05.

Figure 5. The H1c haplotype is associated with altered chromatin structure and CTCF binding in neurons and microglia. A. H3K27ac, H3K4me1 and H3K9ac CHIP-seq peaks (-log10 p-value) spanning the *MAPT* gene in multiple human brain regions; yellow = Hippocampus, orange = Substantia Nigra, dark blue = Anterior Caudate, dark pink = Cingulate Gyrus, red = Inferior Temporal lobe, green = Angular Gyrus, pale blue = Prefrontal Cortex, pale pink = Germinal Matrix. N = 127 **B**. ATAC-seq peaks (-log10 p-value) from NeuN+ (pink) and NeuN- (blue) cells derived from human Hippocampus (HIPP) and Dorsolateral Prefrontal Cortex (DLPFC). N = 5. **C-D.** ATAC-seq peaks (-log10 p-value) from ***C***. iPSC-derived neurons (N=2) and ***D.*** iPSC-derived microglia (N=1-2), separated by H1c major and minor alleles **E.** CTCF CHIP-seq peaks (-log10 p-value) spanning the *MAPT* gene in H1-derived neural cells and CD14+ monocytes (N=1). Yellow highlights denote location of SNPs of interest; from left to right, rs8076152, rs242557 and rs242562.

Table 1. Tag SNPs required to remove the 17q21.31 H1 haplotype association with PSP

rsID	Position	Gene	p-value	-log$_{10}$ p-value	FDR p-value	Major Allele	Minor Allele	MAF (case/control)	OR (95% CI)	OR fisher p-value
rs8076152	17:43995932	MAPT	2.66E-10	9.57	3.79E-05	C	G	0.44 (0.50/0.42)	1.35 (1.24-1.47)	5.68E-13
rs242557	17:44019712	MAPT	2.57E-13	12.59	6.45E-08	G	A	0.48 (0.54/0.46)	1.39 (1.28-1.51)	4.85E-15
rs35838379	17:44040184	MAPT	8.60E-10	9.07	1.16E-06	A	G	0.21 (0.25/0.2)	1.34 (1.22-1.48)	3.33E-09
rs6503455	17:44140748	KANSL1	1.50E-03	2.82	2.66E-01	T	C	0.48 (0.51/0.48)	1.14 (1.02-1.24)	2.30E-03
rs58810165	17:44151005	KANSL1	1.24E-06	5.91	1.07E-03	T	C	0.18 (0.15/0.18)	0.77(0.68-0.86)	4.35E-06

Table 2. 17q21.31 H1 sub-haplotype blocks associated with PD susceptibility

							Stage 1			Stage 2		
Block	#Sub-haplotypes	#SNPs	Start	Stop	Gene	Location	p-value	-log10 p-value	FDR p-value	p-value	-log10 p-value	FDR p-value
H1.1	5	5	44040184	44041992	*MAPT*	Intronic	1.35E-08	7.87	1.73E-06	1.94E-03	2.71	4.50E-03
H1.2	8	8	44090196	44097249	*MAPT*	Intronic/Exonic	6.58E-06	5.18	2.40E-04	1.12E-09	5.95	7.84E-06
H1.3	7	9	44119987	44131305	*KANSL1*	Intronic	1.23E-07	6.91	1.05E-05	2.39E-04	3.47	1.12E-03

Table 3. Sub-haplotypes associated with PD susceptibility in two independent datasets

			Stage 1			Stage 2			Meta					
Block	Sub-haplotype ID	Sub-haplotype	Frequency (Case/Control)	OR (95% CI)	OR Fisher's exact psig	Frequency (Case/Control)	OR (95% CI)	OR Fisher's exact psig	RE OR (95% CI)	RE Heterogeneity (p)	RE p^{sig}	FE OR (95% CI)	FE Heterogeneity (p)	FE p^{sig}
	H1.1a	ACTCT	0.25 (0.26/0.24)	1		0.21 (0.2/0.23)	1		1			1		
	H1.1b	ACTTG	0.27 (0.38/0.22)	1.6 (1.45-1.78)	2.04E-19***	0.38 (0.39/0.36)	1.26 (1.09-1.46)	2.0E-03**	1.43 (1.13-1.81)	7.06 (0.0079)	3.0E-03**	1.48 (1.36-1.61)	7.06 (0.0079)	0***
H1.1	H1.1c	GCCTG	0.18 (0.09/0.23)	0.37 (0.32-0.42)	1.3E-49***	0.14 (0.13/0.16)	0.96 (0.79-1.15)	6.40E-01	0.59 (0.23-1.52)	67.79 (<0.0001)	2.70E-01	0.51 (0.46-0.57)	67.85 (<0.0001)	0***
	H1.1d	ATCTG	0.18 (0.06/0.24)	0.22 (0.19-0.26)	6.5E-92***	0.23 (0.24/0.21)	1.29 (1.1-1.52)	2.0E-03**	0.54 (0.1-3.01)	231.51 (<0.0001)	4.90E-01	0.49 (0.44-0.55)	232.74 (<0.0001)	0***
	H1.1e	ACCTG	0.12 (0.21/0.08)	2.51 (2.2-2.86)	2.4E-45***	0.02 (0.02/0.02)	1.45 (0.98-2.17)	5.70E-02	1.97 (1.15-3.36)	7.13 (0.0079)	1.3E-02*	2.37 (2.1-2.68)	7.13 (0.0076)	0***
	H1.2a	TTTCGATG	0.48 (0.44/0.5)	1		0.49 (0.49/0.48)	1		1			1		
	H1.2b	TCTCGATG	0.17 (0.18/0.16)	1.27 (1.14-1.4)	2.02E-05***	0.17 (0.17/0.18)	0.93 (0.8-1.09)	3.56E-01	1.09 (0.81-1.47)	10.57 (0.001)	5.70E-01	1.14 (1.04-1.24)	10.58 (0.0011)	3.0E-03**
	H1.2c	TTAAAATA	0.15 (0.16/0.14)	1.31 (1.16-1.47)	4.52E-06***	0.19 (0.21/0.18)	1.12 (0.97-1.3)	1.21E-01	1.22 (1.05-1.42)	2.65 (0.103)	9.0E-03**	1.23 (1.13-1.35)	2.65 (0.1037)	4.41E-06***
H1.2	H1.2d	TTAAGATG	0.07 (0.08/0.06)	1.36 (1.16-1.59)	1.2E-04***	0.04 (0.03/0.05)	0.52 (0.38-0.69)	3.46E-06***	0.84 (0.33-2.19)	35.25 (<0.0001)	7.27E-01	1.07 (0.94-1.23)	35.28 (<0.0001)	2.90E-01
	H1.2e	CTTCGATG	0.06 (0.06/0.06)	1.17 (0.99-1.39)	5.64E-02	0.08 (0.08/0.08)	1.05 (0.85-1.3)	6.76E-01	1.12 (0.99-1.28)	0.72 (0.40)	7.00E-02	1.12 (0.99-1.28)	0.72 (0.397)	7.00E-02
	H1.2f	TTTCGGTG	0.05 (0.05/0.05)	1.16 (0.96-1.39)	1.18E-01	0.02 (0.02/0.03)	0.58 (0.41-0.83)	7.1E-03**	0.83 (0.42-1.64)	12.41 (0.0004)	6.00E-01	0.99(0.84-1.16)	12.41 (0.0004)	9.10E-01
	H1.2g	TTAAAATG	0.01 (0.01/0.01)	1.1 (0.77-1.54)	6.07E-01	-	-	-	-	-	-	-	-	-
	H1.2h	TTTCGACG	0.01 (0.01/0.01)	1.28 (0.89-1.84)	1.64E-01	-	-	-	-	-	-	-	-	-
	H1.3a	GACTGAGAT	0.28 (0.32/0.27)	1		0.32 (0.32/0.32)	1		1			1		
H1.3	H1.3b	CATTAGGGC	0.18 (0.11/0.22)	0.43 (0.38-0.49)	1.66E-41***	0.18 (0.18/0.19)	0.95 (0.81-1.11)	5.22E-01	0.64 (0.29-1.44)	58.53 (<0.0001)	2.80E-01	0.57 (0.52-0.63)	58.57 (<0.0001)	0***
	H1.3c	GATTGAGAT	0.17 (0.21/0.16)	1.11 (0.99-1.24)	7.37E-02	0.16 (0.16/0.17)	0.95 (0.80-1.12)	5.33E-01	1.04 (0.89-1.21)	2.24 (0.134)	6.10E-01	1.06 (0.96-1.16)	2.24 (0.1343)	0.23

H1.3d	CTTTGGTGC	0.15 (0.12/0.16)	0.61 (0.54-0.7)	1.58E-14***	0.2 (0.21/0.18)	1.16 (0.99-1.36)	5.81E-02	0.84 (0.45-1.56)	35.27 (<0.0001)	5.80E-01	0.77 (0.7-0.85)	35.28 (<0.0001)	2.96E-07***
H1.3e	CATTGGGGC	0.09 (0.14/0.07)	1.67 (1.45-1.92)	2.79E-13***	0.02 (0.02/0.02)	0.81 (0.54-1.21)	2.83E-01	1.19 (0.58-2.42)	12.01 (0.0004)	6.30E-01	1.54 (1.35-1.75)	12.01 (0.0005)	4.57E-11***
H1.3f	CATTGGGGT	0.08 (0.07/0.09)	0.6 (0.51-0.70)	9.34E-11***	0.09 (0.1/0.08)	1.15 (0.93-1.42)	2.00E-01	0.81 (0.44-1.5)	21.37 (<0.0001)	5.10E-01	0.74 (0.66-0.84)	21.38 (<0.0001)	3.59E-06***
H1.3g	GATCGAGAT	0.04 (0.04/0.04)	0.83 (0.67-1.03)	9.39E-02	0.03 (0.02/0.02)	0.55 (0.38-0.78)	5.58E-04***	0.7 (0.49-1.02)	3.48 (0.062)	6.60E-02	0.75 (0.63-0.9)	3.48 (0.062)	1.6E-03**

Table 4. Association of 17q21.31 sub-haplotype blocks with PSP risk

	PSP Stage 1 + 2		
Block	p-value	$-\log_{10}$ p-value	FDR p-value
H1.1	5.08E-07	6.29	5.08E-07
H1.2	2.81E-15	14.55	8.40E-15
H1.3	8.53E-14	13.07	1.28E-13

Table 5. Association of 17q21.31 sub-haplotypes with PSP susceptibility

	Sub-haplotype	Frequency (Case/Control)	OR (95% CI)	Fisher's exact p^{sig}
H1.1	H1.1a	0.29 (0.26/0.30)	-	-
	H1.1b	0.25 (0.23/0.25)	1.08 (0.97-1.22)	1.70E-01
	H1.1c	0.22 (0.21/0.22)	1.43 (1.27-1.6)	8.17E-10
	H1.1d	0.21 (0.24/0.20)	1.1 (0.98-1.24)	9.93E-02
	H1.1e	0.04 (0.06/0.03)	2.3 (1.88-2.8)	1.01E-15
H1.2	H1.2a	0.39 (0.36/0.40)	-	-
	H1.2b	0.20 (0.24/0.19)	1.46 (1.3-1.63)	2.41E-11
	H1.2c	0.18 (0.15/0.18)	0.91 (0.8-1.04)	1.73E-01
	H1.2d	0.07 (0.07/0.08)	0.99 (0.83-1.17)	8.95E-01
	H1.2e	0.08 (0.1/0.08)	1.43 (1.23-1.67)	4.73E-06
	H1.2f	0.06 (0.08/0.06)	1.47 (1.23-1.74)	1.55E-05
	H1.2g	0.01 (0.01/0.01)	1.07 (0.73-1.54)	7.06E-01
	H1.2h	-	-	-
H1.3	H1.3a	0.32 (0.29/0.32)	-	-
	H1.3b	0.20 (0.25/0.19)	1.44 (1.28-1.61)	4.70E-10
	H1.3c	0.16 (0.16/0.16)	1.15 (1.01-1.31)	3.36E-02
	H1.3d	0.18 (0.15/0.18)	0.9 (0.79-1.03)	1.23E-01
	H1.3e	0.02 (0.02/0.03)	0.81 (0.59-1.1)	1.91E-01
	H1.3f	0.08 (0.09/0.07)	1.41 (1.21-1.66)	2.02E-05
	H1.3g	0.04 (0.04/0.04)	1.06 (0.85-1.32)	5.78E-01

SI Table 1. Association of the 17q21.31 H1/H2 haplotype tag SNP with PSP risk

	H2 Tag SNP	Position	p-value	$-\log_{10}$ p-value	FDR p-value	Major Allele	Minor Allele	MAF (case/control)	H2 OR (95% CI)	H1 OR (95% CI)	OR fisher p-value
Stage 1			1.07E-74	73.97	2.93E-71			0.20 (0.05/0.22)	0.19 (0.16-0.24)	5.18 (4.26-6.30)	5.90E-96
Stage 2	rs8070723	17:44081064	8.03E-20	19.1	7.88E-18	A	G	0.17 (0.09/0.23)	0.33 (0.25-0.43)	3.04 (2.34-3.94)	4.38E-19
Meta			-	-	-			-	0.25 (0.15-0.42)	3.99 (2.37-6.74)	2.13E-07[Ŧ]

[Ŧ]Random effects Meta p-value

SI Table 2. Variants within the H1 haplotype clade associated with PSP risk in two independent datasets

	Stage 1 Top SNP	Position	p-value	-log$_{10}$ p-value	FDR p-value	Major Allele	Minor Allele	MAF (case/control)	OR (95% CI)	OR fisher p-value
Stage 1			2.39E-12	11.62	6.10E-07			0.47 (0.55/0.46)	1.44 (1.31-1.59)	4.18E-14
Stage 2	rs242562	17:44026739	9.01E-02	1.05	8.71E-01	G	A	0.48 (0.51/0.46)	1.24 (1.02-1.52)	3.66E-02
Meta			-	-	-			-	1.38 (1.2-1.58)	8.9E-06[Ŧ]

	Stage 2 Top SNP	Position	p-value	-log$_{10}$ p-value	FDR p-value	Major Allele	Minor Allele	MAF (case/control)	OR (95% CI)	OR fisher p-value
Stage 1			4.11E-02	1.39	7.20E-01			0.02 (0.018/0.024)	0.75 (0.53-1.06)	1.10E-01
Stage 2	rs191237882	17:44058967	1.70E-03	2.77	3.31E-01	G	A	0.03 (0.01/0.04)	0.34 (0.16-0.7)	2.12E-03
Meta			-	-	-			-	0.54 (0.25-1.15)	0.11[Ŧ]

	H1c Tag SNP	Position	p-value	-log$_{10}$ p-value	FDR p-value	Major Allele	Minor Allele	MAF (case/control)	OR (95% CI)	OR fisher p-value
Stage 1			3.60E-12	11.44	8.93E-09			0.47 (0.55/0.46)	1.45 (1.31-1.59)	4.22E-14
Stage 2	rs242557	17:44019712	5.40E-02	1.27	8.15E-01	G	A	0.49 (0.52/0.46)	1.26 (1.03-1.53)	2.12E-02
Meta			-	-	-			-	1.39 (1.22-1.57)	3.58E-07[Ŧ]

[Ŧ]Random effects Meta p-value

SI Table 3. Significant SNPs in H1 clade 17q21.31 region associated with PSP risk in Stage 1 and Stage 2 combined data.

	rsID	Position	p-value	-log$_{10}$ p-value	FDR p-value	Major Allele	Minor Allele	MAF (case/control)	OR (95% CI)	OR fisher p-value
Peak	rs8076152	17:43995932	9.89E-10	9	1.21E-06	C	G	0.44 (0.50/0.42)	1.35 (1.24-1.47)	5.68E-13
	rs12947764	17:44004442	4.41E-09	8.36	4.65E-06	T	G	0.35 (0.41/0.34)	1.33 (1.46-1.23)	1.93E-11
	rs11867549	17:44013235	6.39E-10	9.19	8.06E-07	A	G	0.33 (0.39/0.32)	1.34 (1.46-1.23)	2.71E-11
	rs242557	17:44019712	7.12E-13	12.15	1.45E-09	G	A	0.48 (0.54/0.46)	1.39 (1.28-1.51)	4.85E-15
	rs242562	17:44026739	1.32E-12	11.88	2.64E-09	G	A	0.47 (0.54/0.50)	1.38 (1.50-1.27)	1.28E-14
	rs35838379	17:44040184	2.97E-08	7.53	2.76E-05	A	G	0.21 (0.25/0.2)	1.34 (1.22-1.48)	3.33E-09
Broad	rs9303524	17:44136576	2.79E-09	8.55	3.08E-06	C	T	0.20 (0.25/0.19)	1.37 (1.51-1.25)	2.56E-10
	rs6503455	17:44140748	2.15E-03	2.67	2.90E-01	T	C	0.48 (0.51/0.48)	1.14 (1.02-1.24)	2.30E-03
	rs58810165	17:44151005	9.82E-07	6.01	7.69E-04	T	C	0.18 (0.15/0.18)	0.77 (0.68-0.86)	4.35E-06
	rs4792835	17:44219555	2.20E-08	7.66	2.11E-05	A	G	0.21 (0.26/0.21)	1.33 (1.47-1.21)	5.30E-09

SI Table 4. Association analysis conditioning on top 6 SNPs from the H1 PSP combined analysis

Conditioned SNP	Test SNP	p-value
rs8076152	rs12947764	0.13
	rs11867549	0.08
	rs242557	1.06E-08
	rs242562	1.85E-08
	rs35838379	1.12E-03
rs12947764	rs8076152	2.23E-03
	rs11867549	0.37
	rs242557	1.19E-07
	rs242562	2.22E-07
	rs35838379	2.22E-03
rs11867549	rs8076152	6.50E-04
	rs12947764	0.13
	rs242557	1.86E-07
	rs242562	3.47E-07
	rs35838379	2.70E-03
rs242557	rs8076152	2.24E-06
	rs12947764	1.05E-03
	rs11867549	3.74E-03
	rs242562	0.92
	rs35838379	1.59E-03
rs242562	rs8076152	1.35E-06
	rs12947764	6.50E-04
	rs11867549	2.28E-03
	rs242557	0.14
	rs35838379	1.27E-03
rs35838379	rs8076152	1.07E-07
	rs12947764	6.16E-06
	rs11867549	1.52E-05
	rs242557	6.96E-10
	rs242562	1.62E-09
rs8076152 + rs242557 + rs35838379	rs9303524	3.19E-02
	rs6503455	8.40E-01
	rs58810165	8.38E-05
	rs4792835	0.18

SI Table 5. Association analysis conditioning on 4 broad region SNPs from the H1 PSP combined analysis, and on all 5 SNPs required to fully remove 17q21.31 H1 association signal across the locus

Conditioned SNP	Test SNP	p-value
rs9303524	rs6503455	0.72
	rs58810165	1.50E-03
	rs4792835	0.57
rs6503455	rs9303524	4.46E-08
	rs58810165	4.86E-11
	rs4792835	7.32E-07
rs58810165	rs9303524	9.55E-08
	rs6503455	9.04E-09
	rs4792835	1.28E-06
rs4792835	rs9303524	2.20E-02
	rs6503455	8.70E-01
	rs58810165	1.20E-03
rs8076152 + rs242557 + rs35838379 + rs6503455 + rs58810165	rs12947764	2.69E-01
	rs11867549	2.20E-01
	rs242562	9.35E-01
	rs9303524	5.64E-01
	rs4792835	5.33E-01

SI Table 6. PSP-associated H1 haplotype clade SNPs are not associated with PD risk

Stage 1

rsID	p-value	-log$_{10}$ p-value	FDR p-value	MAF (case/control)	OR (95% CI)	OR fisher p-value
rs8076152	3.82E-01	0.42	0.97	0.40 (0.41/0.39)	1.06 (0.98-1.15)	1.25E-01
rs242557	8.62E-01	0.06	0.99	0.45 (0.45/0.45)	1.01 (0.94-1.1)	7.23E-01
rs35838379	1.78E-02	1.75	0.9	0.18 (0.20/0.17)	1.16 (1.05-1.28)	2.98E-03
rs6503445	6.46E-01	0.19	0.98	0.49 (0.49/0.50)	0.98 (0.90-1.06)	5.71E-01
rs58810165	4.10E-03	2.39	0.72	0.15 (0.16/0.14)	1.16 (1.04-1.29)	5.51E-03

Stage 2

rsID	p-value	-log$_{10}$ p-value	FDR p-value	MAF (case/control)	OR (95% CI)	OR fisher p-value
rs8076152	9.77E-01	0.01	1	0.27 (0.25/0.29)	0.81 (0.72-0.92)	9.30E-04
rs242557	3.29E-01	0.48	0.97	0.44 (0.43/0.45)	0.94 (0.84-1.05)	2.82E-01
rs35838379	5.68E-01	0.25	0.98	0.14 (0.13/0.16)	0.81 (0.69-0.94)	5.96E-03
rs6503445	9.89E-02	1	0.96	0.49 (0.50/0.47)	1.11 (0.99-1.24)	4.89E-02
rs58810165	4.51E-02	1.35	0.94	0.19 (0.21/0.18)	1.21 (1.05-1.39)	7.59E-03

SI Table 7. Top SNPs in the 17q21.31 locus associated with PD in two independent data sets

	Stage 1 Top SNP	Position	p-value	-log$_{10}$ p-value	FDR p-value	Major Allele	Minor Allele	MAF (case/control)	OR (95% CI)	OR fisher p-value
Stage 1			2.74E-09	8.562	5.44E-04			0.21 (0.17/0.22)	0.75 (0.69-0.81)	8.50E-13
Stage 2	rs17763050	17:43903336	6.10E-01	0.22	9.70E-01	G	A	0.26 (0.25/0.28)	0.87 (0.8-0.96)	3.77E-03
Meta			-	-	-			-	0.82 (0.73-0.93)	1.42E-03[Ŧ]

	Stage 2 Top SNP	Position	p-value	-log$_{10}$ p-value	FDR p-value	Major Allele	Minor Allele	MAF (case/control)	OR (95% CI)	OR fisher p-value
Stage 1			2.59E-03	2.59	1.57E-01			0.03 (0.04/0.03)	1.42 (1.12-1.7)	1.13E-04
Stage 2	rs151036546	17:44122049	8.7E10-04	3.06	2.70E-01	T	C	0.02 (0.01/0.03)	0.51 (0.37-0.68)	2.72E-06
Meta			-	-	-			-	0.85 (0.31-2.35)	7.6E-01[Ŧ]

	H2 Tag SNP	Position	p-value	-log$_{10}$ p-value	FDR p-value	Major Allele	Minor Allele	MAF (case/control)	H2 OR (95% CI)	H1 OR (95% CI)	OR fisher p-value
Stage 1			1.70E-07	6.77	1.96E-05			0.2 (0.16/0.21)	0.77 (0.71-0.83)	1.29 (1.19-1.40)	1.66E-10
Stage 2	rs8070723	17:44081064	4.70E-01	0.33	9.50E-01	A	G	0.25 (0.24/0.27)	0.86 (0.78-0.94)	1.14 (1.04-1.25)	1.07E-03
Meta			-	-	-			-	0.82 (0.76-0.89)	1.22 (1.08-1.37)	6.32E-07[Ŧ]

[Ŧ]Random effects Meta p-value

SI Table 8. Top H1 haplotype clade SNPs associated with PD risk in two independent data sets

	Stage 1 Top SNP	Position	p-value	-log$_{10}$ p-value	FDR p-value	Major Allele	Minor Allele	MAF (case/control)	OR (95% CI)	OR fisher p-value
Stage 1			8.4E10-4	3.08	0.55			0.23 (0.25/0.21)	1.21 (1.10-1.32)	4.44E-05
Stage 2	rs41543512	17:44090778	0.61	0.22	0.99	T	A	0.26 (0.24/0.24)	1.02 (0.89-1.16)	0.822
Meta			-	-	-			-	1.11 (0.94-1.32)	0.21[Ŧ]

	Stage 2 Top SNP	Position	p-value	-log$_{10}$ p-value	FDR p-value	Major Allele	Minor Allele	MAF (case/control)	OR (95% CI)	OR fisher p-value
Stage 1			-	-	-			-	-	-
Stage 2	rs139217062	17:43849514	2.80E-03	2.55	0.66	T	C	0.02 (0.03/0.02)	1.82 (1.25-2.69)	1.01E-02
Meta			-	-	-			-	-	-

	Stage 2 2nd SNP	Position	p-value	-log$_{10}$ p-value	FDR p-value	Major Allele	Minor Allele	MAF (case/control)	OR (95% CI)	OR fisher p-value
Stage 1			0.2	0.7	0.94			0.05 (0.05/0.04)	1.11 (0.92-1.34)	0.26
Stage 2	rs16940711	17:43930006	8.80E-03	2.06	0.79	A	C	0.03 (0.02/0.04)	0.54 (0.39-0.76)	2.51E-04
Meta			-	-	-			-	0.79 (0.39-1.59)	0.5[Ŧ]

[Ŧ]Random effects Meta p-value

SI Table 9. MSigDB Annotations resulting from Gene Set Enrichment Analsysis associated with *LRRC37A* overexpression

MSigDB Annotation	ES	NES	Pvalue	P.adj
NABA_BASEMENT_MEMBRANES	-0.74525	-2.32167	6.48E-06	0.02188
GO_REGULATION_OF_CENTROSOME_DUPLICATION	0.53846	2.30563	6.48E-06	0.02188
GO_AMINOGLYCAN_BIOSYNTHETIC_PROCESS	-0.57664	-2.23452	1.30E-05	0.02188
GO_INTRA_S_DNA_DAMAGE_CHECKPOINT	0.70901	2.23379	1.30E-05	0.02188
REACTOME_CHONDROITIN_SULFATE_DERMATAN_SULFATE_METABOLISM	-0.6186	-2.2161	1.51E-05	0.02188
PID_FANCONI_PATHWAY	0.50129	2.16649	3.02E-05	0.03126
REACTOME_GLYCOSAMINOGLYCAN_METABOLISM	-0.54135	-2.15238	3.67E-05	0.03126
REACTOME_HEPARAN_SULFATE_HEPARIN_HS_GAG_METABOLISM	-0.62113	-2.14604	3.89E-05	0.03126
GO_LYSOSOMAL_LUMEN	-0.56321	-2.1452	3.89E-05	0.03126
REACTOME_METABOLISM_OF_CARBOHYDRATES	-0.4869	-2.13494	4.97E-05	0.03268
GO_GLUCOSE_6_PHOSPHATE_METABOLIC_PROCESS	-0.71837	-2.13657	4.97E-05	0.03268
GO_REGULATION_OF_MRNA_POLYADENYLATION	0.75112	2.1203	6.48E-05	0.03907
GO_CELL_CYCLE_CHECKPOINT	0.34394	2.10909	7.78E-05	0.03941
REACTOME_CDC6_ASSOCIATION_WITH_THE_ORC_ORIGIN_COMPLEX	0.63267	2.09992	8.21E-05	0.03941
REACTOME_ACTIVATION_OF_ATR_IN_RESPONSE_TO_REPLICATION_STRESS	0.53506	2.09346	8.86E-05	0.03941
GO_PENTOSE_METABOLIC_PROCESS	-0.76215	-2.09532	8.86E-05	0.03941
GO_REGULATION_OF_DOUBLE_STRAND_BREAK_REPAIR_VIA_HOMOLOGOUS_RECOMBINATION	0.58715	2.08406	0.0001	0.03941
REACTOME_HS_GAG_BIOSYNTHESIS	-0.63593	-2.08083	0.00011	0.03941
REACTOME_DAG_AND_IP3_SIGNALING	-0.63953	-2.06978	0.00011	0.03941
GO_FILOPODIUM_ASSEMBLY	-0.77604	-2.06753	0.00012	0.03941
REACTOME_GENERIC_TRANSCRIPTION_PATHWAY	0.35621	2.06716	0.00012	0.03941
GO_BASEMENT_MEMBRANE	-0.54943	-2.06718	0.00012	0.03941
GO_REGULATION_OF_DNA_RECOMBINATION	0.48487	2.06316	0.00013	0.03941
GO_PENTOSE_PHOSPHATE_SHUNT	-0.76058	-2.04918	0.00016	0.04203
GO_SULFUR_COMPOUND_BIOSYNTHETIC_PROCESS	-0.47933	-2.04265	0.00017	0.04203
REACTOME_CA_DEPENDENT_EVENTS	-0.65457	-2.04014	0.00017	0.04203
REACTOME_G2_M_CHECKPOINTS	0.48289	2.03953	0.00017	0.04203
GO_NEGATIVE_REGULATION_OF_CALCIUM_ION_TRANSMEMBRANE_TRANSPORT	-0.68211	-2.04261	0.00017	0.04203
GO_RESPIRATORY_CHAIN	-0.51681	-2.04074	0.00017	0.04203
REACTOME_A_TETRASACCHARIDE_LINKER_SEQUENCE_IS_REQUIRED_FOR_GAG_SYNTHESIS	-0.67809	-2.02806	0.00018	0.04249
KEGG_ARGININE_AND_PROLINE_METABOLISM	-0.56789	-2.02532	0.00019	0.04249
REACTOME_RESPIRATORY_ELECTRON_TRANSPORT_ATP_SYNTHESIS_BY_CHEMIOSMOTIC_COUPLING_AND_HEAT_PRODUCTION_BY_UNCOUPLING_PROTEINS_	-0.49546	-2.02688	0.00019	0.04249
GO_CARBOHYDRATE_CATABOLIC_PROCESS	-0.51301	-2.01809	0.0002	0.04357
GO_VACUOLAR_LUMEN	-0.52006	-2.01552	0.00021	0.04413
KEGG_GLYCEROLIPID_METABOLISM	-0.57845	-2.01164	0.00023	0.04647
GO_MITOTIC_SISTER_CHROMATID_COHESION	0.64016	2.01021	0.00024	0.04647
GO_PROTEIN_LOCALIZATION_TO_CENTROSOME	0.62206	2.00483	0.00024	0.04647
PID_INTEGRIN1_PATHWAY	-0.54084	-2.00427	0.00024	0.04647
GO_CELL_COMMUNICATION_BY_ELECTRICAL_COUPLING	-0.76521	-1.99786	0.00025	0.04655
GO_OXIDOREDUCTASE_ACTIVITY_ACTING_ON_A_HEME_GROUP_OF_DONORS	-0.61705	-1.99492	0.00027	0.04655
KEGG_PENTOSE_PHOSPHATE_PATHWAY	-0.6456	-1.99106	0.00027	0.04655
REACTOME_AXON_GUIDANCE	-0.43246	-1.98631	0.00029	0.04655
KEGG_OXIDATIVE_PHOSPHORYLATION	-0.45821	-1.98226	0.00029	0.04655
GO_NEGATIVE_REGULATION_OF_CATION_CHANNEL_ACTIVITY	-0.64444	-1.98413	0.00029	0.04655
GO_MONOSACCHARIDE_BIOSYNTHETIC_PROCESS	-0.55483	-1.97777	0.00029	0.04655
GO_MITOTIC_CELL_CYCLE_CHECKPOINT	0.36794	1.97634	0.0003	0.04655
GO_NEGATIVE_REGULATION_OF_DNA_RECOMBINATION	0.66436	1.97506	0.0003	0.04655
GO_EXTRACELLULAR_MATRIX	-0.43099	-1.97284	0.00031	0.04658

GO_HEXOSE_CATABOLIC_PROCESS	-0.59107	-1.96976	0.00033	0.04658
GO_CENTRIOLE_ASSEMBLY	0.58792	1.96901	0.00033	0.04658
GO_PROTEOGLYCAN_METABOLIC_PROCESS	-0.52334	-1.96909	0.00033	0.04658
GO_CARBOHYDRATE_DERIVATIVE_BIOSYNTHETIC_PROCESS	-0.39811	-1.96209	0.00036	0.04987
KEGG_GLYCOSAMINOGLYCAN_BIOSYNTHESIS_HEPARAN_SULFATE	-0.65664	-1.95811	0.00038	0.04987
GO_5_3_EXONUCLEASE_ACTIVITY	0.673	1.95573	0.00039	0.04987
GO_MITOCHONDRIAL_MEMBRANE_PART	-0.43705	-1.95389	0.00039	0.04987
GO_STRUCTURAL_CONSTITUENT_OF_CYTOSKELETON	-0.48043	-1.95241	0.00039	0.04987
GO_PROTEIN_TARGETING_TO_PLASMA_MEMBRANE	-0.62103	-1.94986	0.0004	0.04987
GO_CARBOHYDRATE_BIOSYNTHETIC_PROCESS	-0.46643	-1.94959	0.00041	0.04987
GO_NUCLEOSIDE_TRIPHOSPHATE_METABOLIC_PROCESS	-0.42977	-1.9494	0.00041	0.04987
GO_POLARIZED_EPITHELIAL_CELL_DIFFERENTIATION	-0.77137	-1.94812	0.00041	0.04987
GO_REGULATION_OF_CENTROSOME_CYCLE	0.52059	1.94172	0.00044	0.04987
GO_SKELETAL_MUSCLE_ADAPTATION	0.87488	1.94102	0.00044	0.04987
GO_OXIDATIVE_PHOSPHORYLATION	-0.48491	-1.93979	0.00045	0.04987
GO_POSITIVE_REGULATION_OF_MONOOXYGENASE_ACTIVITY	-0.63647	-1.93803	0.00045	0.04987
GO_FIBRINOLYSIS	-0.76589	-1.93615	0.00046	0.04987
GO_RIBONUCLEOSIDE_TRIPHOSPHATE_BIOSYNTHETIC_PROCESS	-0.52225	-1.93504	0.00047	0.04987
GO_AMINOGLYCAN_METABOLIC_PROCESS	-0.47287	-1.93482	0.00047	0.04987
GO_EXTRACELLULAR_MATRIX_COMPONENT	-0.48173	-1.93487	0.00047	0.04987

SI Table 10. Chromatin marks from publicly available CHIP-seq data spanning *MAPT* intron 1 across multiple brain regions (Roadmap Epigenetics Consortium)

Chromatin mark	Start	End	SNP	Signal value	p-value (-log10)	Q Value (-log10)	Brain region
H3K27ac	43981947	44003184	rs8076152	3.18029	5.11397	3.54803	HPP
	43994962	44003602	rs8076152	2.32427	2.64757	1.22123	SN
	43995823	44002902	rs8076152	2.28686	2.57752	1.20401	AC
	43995806	44003069	rs8076152	2.38463	2.82814	1.38775	CG
	43982143	44004659	rs8076152	2.76309	4.29747	2.76714	ITL
	43981878	44002910	rs8076152	2.46659	3.18897	1.67477	AG
	43981890	43996265	rs8076152	2.63143	3.26713	1.72236	DLPFC
	44009783	44033634	rs242557	6.57366	19.51962	17.5781	HPP
	44004660	44033678	rs242557	5.5659	14.49617	12.55186	SN
	44006609	44033870	rs242557	4.69771	10.13337	8.26917	AC
	44006737	44033722	rs242557	4.80705	11.25418	9.2885	CG
	44010251	44033768	rs242557	5.1329	13.26008	11.25681	ITL
	44011338	44039151	rs242557	3.98398	8.56048	6.66871	AG
	44005926	44033785	rs242557	3.50053	5.97218	4.18216	DLPFC
H3K4me1	43970803	44003261	rs8076152	3.54324	6.85326	5.11976	HPP
	43981850	43988652	rs8076152	3.20474	4.70157	2.90342	SN
	43981778	44003143	rs8076152	3.15869	4.66118	2.87483	AC
	43981843	43994881	rs8076152	3.23384	4.71858	2.91229	CG
	43981813	43994486	rs8076152	3.30926	4.14553	2.27068	ITL
	43981866	44003476	rs8076152	3.0397	4.6035	2.7164	AG
	43981854	44003599	rs8076152	3.14093	4.37474	2.45768	DLPFC
	44006304	44058497	rs242557	3.93424	8.16521	6.31087	HPP
	44004007	44033581	rs242557	3.50446	5.63458	3.68874	SN
	44006292	44033716	rs242557	3.32282	5.23157	3.35548	AC
	44009641	44033836	rs242557	3.90516	6.63762	4.54595	CG
	44010603	44033156	rs242557	4.04251	5.92398	3.78184	ITL
	44006626	44033705	rs242557	3.10584	4.74337	2.83516	AG
	44004846	44033707	rs242557	3.00681	4.26102	2.39422	DLPFC
	44016881	44027507	rs242557	2.38941	1.92806	0.32607	GM
H3K9ac	44011767	44031905	rs242557	3.28186	4.17032	2.35737	SN
	44011576	44029818	rs242557	2.7198	3.48366	1.82626	AC
	44015567	44031744	rs242557	2.93022	3.29985	1.56061	CG
	44015315	44033655	rs242557	2.85524	3.75238	2.07793	ITL
	44010087	44032998	rs242557	2.59319	3.33418	1.65882	AG
	44016670	44029569	rs242557	2.86184	3.24344	1.49267	DLPFC
	43990857	44002046	rs8076152	2.00058	1.86481	0.46506	AC
	43995932	43998872	rs8076152	2.87777	2.87744	1.23631	CG
	43990830	44003343	rs8076152	2.11439	2.07993	0.63774	ITL
	43990572	44003351	rs8076152	2.18025	2.37967	0.8728	AG

HPP; Hippocampus. SN; Substantia nigra. AC; Anterior caudate. CG; Cingulate gyrus. ITL; Inferior temporal lobe. AG; Angular gyrus. DLPFC; Dorsolateral prefrontal gyrus. GM; Germinal matrix.

SI Table 11. Open chromatin peaks from publicly available ATAC-seq data across MAPT intron 1 in NeuN+ and NeuN- cells derived from human brain, from ATAC-seq data in iPSC-derived neurons, astrocytes and microglia, and CTCF binding sites derived from publicly available CHIP-seq data (Encode Portal)

Brain

Cell type	Brain region	Start	End	Fold Enrichment	p-value (-log10)	Q Value (-log10)	SNP
NeuN-	ACC	44019569	44020113	6.68	33.93	30.79	rs242557
NeuN-	AMY	44019602	44020083	5.56	21.56	18.42	rs242557
NeuN-	AMY	44026534	44026740	3.51	7.82	5.2	rs242562
NeuN-	DLPFC	44019567	44020103	8.72	61.59	58.21	rs242557
NeuN-	DLPFC	44024164	44024616	3.02	8.97	6.51	-
NeuN-	DLPFC	44026516	44026739	3.79	10.38	7.86	rs242562
NeuN-	HIPP	44019558	44020246	7.48	35.63	32.58	rs242557
NeuN-	HIPP	44026549	44026751	3.12	5.94	3.68	rs242562
NeuN-	INS	44019573	44020124	6.82	25.27	21.97	rs242557
NeuN-	ITC	44019557	44020113	8.6	52.72	49.35	rs242557
NeuN-	ITC	44024174	44024496	3.33	9.54	7.01	-
NeuN-	MDT	44019555	44020457	8.59	110.49	107.02	rs242557
NeuN-	MDT	44026422	44026824	5.47	29.31	26.57	rs242562
NeuN-	NAC	44019557	44020115	9.63	36.32	32.6	rs242557
NeuN-	NAC	44024169	44024369	2.81	4.3	1.93	-
NeuN-	OFC	44019557	44020128	8.23	52.13	48.89	rs242557
NeuN-	OFC	44024175	44024399	3.48	11.03	8.51	-
NeuN-	PMC	44019557	44020335	8.63	61.68	58.32	rs242557
NeuN-	PMC	44026530	44026740	3.58	8.74	6.3	rs242562
NeuN-	PUT	44019557	44020271	7.85	58.62	55.34	rs242557
NeuN-	PUT	44024200	44024597	2.49	6.31	4.05	-
NeuN-	PUT	44026470	44026824	4.85	16.9	14.26	rs242562
NeuN-	PVC	44019557	44020390	7.91	59.49	56.17	rs242557
NeuN-	STC	44019557	44020083	8.61	73.72	70.19	rs242557
NeuN-	STC	44026483	44026730	3.55	10.58	8.04	rs242562
NeuN-	VLPFC	44019557	44020098	8.11	45.24	42.07	rs242557
NeuN-	VLPFC	44024157	44024369	3.07	7.99	5.58	-
NeuN-	VLPFC	44026485	44026734	4.04	10.04	7.54	rs242562

iPSC-derived cell lines

Cell type	Haplotype	Start	End	Fold Enrichment	SNP
Neurons	Control	-	-	-	rs242557
Neurons	Control	-	-	-	rs242562
Neurons	H1c	44019540	44020050	15.76	rs242557
Neurons	H1c	44026519	44026801	9.58	rs242562
Astrocytes	Control	44026493	44026784	18.45	rs242562
Astrocytes	H1c	-	-	-	rs242562
Microglia	Control	-	-	-	rs242562
Microglia	H1c	44026422	44026836	25.36	rs242562

CTCF binding

Cell type	ENCODE dataset ID	Start	End	Signal Value	SNP
H1 Neural cells	ENCSR822CEA	-	-	-	rs242562
Primary astrocytes	ENCSR000AOO	44026436	44026786	177	rs242562
CD14+ monocytes	ENCSR000ATN	44026367	44027080	638	rs242562

SI Methods Table 1. Data summary for Stage 1 and Stage 2 PSP analyses

	Cohort	Array	Total N (Case/Control)	H1H1 (Case/Control)	H2H2 (Case/Control)	#SNPs used for imputation (Chr17)	#Imputed SNPs (Chr17) R2 ≥ 0.3
Stage 1	ACT1	Illumina Human660	1348 (0/1348)	829 (0/829)	54 (0/54)	12201	587642
	ACT2	Illumina Human660	5 (0/5)	3 (0/3)	0 (0/0)	4090	395069
	ADC1	Illumina Human660	543 (0/543)	342 (0/342)	16 (0/0)	12291	779330
	ADNI	Illumina Human660	53 (0/53)	23 (0/23)	3 (0/3)	4707	457101
	BIOCARD	Illumina OmniExpress	123 (0/123)	67 (0/67)	3 (0/3)	2372	403814
	CHAP2	Illumina OmniExpress	164 (0/164)	109 (0/109)	4 (0/4)	6082	479797
	EAS	Illumina OmniExpress	209 (0/209)	134 (0/134)	7 (0/7)	3010	395203
	GSK	Affymetrix 500/ Illumina 550/ Illumina 1M	764 (0/764)	456 (0/456)	36 (0/36)	5399	287017
	MAYO	Illumina HumanHap300	925 (0/925)	583 (0/583)	43 (0/43)	5765	541592
	MIRAGE	Illumina 610/330	294 (0/294)	187 (0/187)	14 (0/14)	4474	510744
	MTV	Illumina OmniExpress	193 (0/193)	130 (0/130)	8 (0/8)	4686	516179
	NBB	Illumina 1M	85 (0/85)	56 (0/56)	2 (0/2)	2857	347197
	NIALOAD	Illumina 610	801 (0/801)	473 (0/473)	46 (0/46)	21257	609927
	OHSU	Illumina Human CNV370v1_C	109 (0/109)	69 (0/69)	3 (0/3)	2822	405048
	RMAYO	Illumina OmniExpress	271 (0/271)	176 (0/176)	5 (0/5)	3873	376279
	ROSMAP	Illumina 1M	853 (0/853)	562 (0/562)	39 (0/39)	8626	493678
	ROSMAP2	Illumina OmniExpress	237 (0/237)	149 (0/149)	14 (0/14)	6983	623370
	TARCC1	Affymetrix 6.0	144 (0/144)	80 (0/80)	10 (0/10)	3888	430156
	TGEN2	Affymetrix 6.0	488 (0/488)	279 (0/279)	24 (0/24)	7389	513778
	UKS	Illumina 550	973 (0/973)	619 (0/619)	35 (0/35)	8581	527402
	UMVUMSSM	Illumina Human660/1M Duo/Affymetrix 6.0	1006 (0/1006)	613 (0/613)	51 (0/51)	4686	472189
	UPITT	Illumina Human Omni-Quad	834 (0/834)	490 (0/490)	41 (0/41)	14611	578298
	WASHU	Illumina Human610	166 (0/166)	106 (0/106)	10 (0/10)	4746	447398
	WASHU2	Illumina OmniExpress	65 (0/65)	44 (0/44)	3 (0/3)	2336	447792
	WHICAP	Illumina OmniExpress	562 (0/562)	317 (0/317)	37 (0/37)	5763	433983
	Hoglinger et al 2011 Stage 1	Illumina Human660W-Quad	1071 (1071/0)	966 (966/0)	3 (3/0)	13904	487411
	Merged	-	12286 (1071/11215)	7862 (966/6896)	511 (3/508)	-	263401

		Array	Total N (Case/Control)	H1H1 (Case/Control)	H2H2 (Case/Control)	#SNPs used for imputation (Chr17)	#Imputed SNPs (Chr17) R2 >= 0.3
Stage 2	Mt.Sinai Controls	Illumina Infinium Human Genome Screening Array v2.4	676(0/676)	393 (0/393)	29 (0/29)	18033	431349
	Mt.Sinai PSP	Illumina Infinium Human Genome Screening Array v2.4	455 (455/0)	377 (377/0)	3 (3/0)	13265	402125
	Merged	-	1131 (455/676)	770 (377/393)	32 (3/29)	-	199577

SI Methods Table 2. Data summary for Stage 1 and Stage 2 PD analyses

	Cohort	Array	Total N (Case/Control)	H1H1 (Case/Control)	H2H2 (Case/Control)	#SNPs used for imputation (Chr17)	#Imputed SNPs (Chr17) R2 >= 0.3
Stage 1	NIA (IPDGC)	Illumina 550K v1 & v3	3906 (905/3001)	2447 (606/1841)	161 (31/130)	11941	634666
	NL (IPDGC)	Illumina 550K & 610K	2725 (767/1958)	1654 (500/1154)	144 (39/105)	13457	530635
	FIN (IPDGC)	Illumina 370K	860 (368/492)	732 (316/416)	7 (4/3)	7970	357255
	GER (IPDGC)	Illumina 550K v1	1673 (740/933)	1121 (527/594)	51 (19/32)	11803	481302
	Merged	-	9164 (2780/6384)	5932 (1937/3995)	378 (94/284)	-	295782
Stage 2	McGill (IPDGC)	Illumina Omini Express & NeuroChip v.1.0	1485 (582/903)	859 (354/505)	70 (19/51)	17467	515155
	SP (IPDGC)	NeuroChip v.1.0 & v.1.1	3444 (2117/1327)	1861 (1184/677)	243 (130/113)	16590	493744
	Merged	-	4929 (2699/2230)	2925 (1538/1182)	313 (149/164)	-	368729

SI Methods Table 3. Sample summary for RNA-seq expression analyses, dPCR copy number variation analyses and qRTPCR.

		Brain Region/Tissue	N	#H1H1	#Homozygote block1 H1 sub-haplotypes	#Homozygote block2 H1 sub-haplotypes	#Homozygote block3 H1 sub-haplotypes	#H2H2
RNA-seq Analysis	AMP-AD ROSMAP	PFC	450	289	64	61	56	19
	AMP-AD MAYO	TCX	276	185	49	45	38	7
	CommonMind	PFC	624	347	70	34	28	21
	Total	-	1350	821	183	140	122	47
dPCR	MSMD	Blood	66	49	37	30	32	17
	ADRC	Blood	33	21	16	9	13	12
	Charney LB	Blood	20	17	13	12	9	3
	Charney PD	PFC	41	34	18	18	15	4
	Total	-	160	121	84	69	69	36
qRTPCR	Charney PD	PFC	41	34	18	18	15	4

SI Methods Table 4. Sources for brain tissue/DNA for Mt.Sinai PSP/CBD and control cohorts

Institution	Country	Cases	Controls	Total
Banner Sun Health Sciences	USA	67	110	177
Boston University	USA	7	28	35
Columbia University Medical Center	USA	42	26	68
Emory	USA	19	10	29
Imperial College London	UK	20	-	20
Mount Sinai NDBB	USA	3	202	205
Netherlands Brain Bank	Netherlands	42	-	42
Neurobiobank - Harvard / McLean Hospital / Harvard Brain Tissue Resource Center	USA	35	-	35
Neurobiobank - University of Maryland	USA	7	-	7
Neurobiobank - University of Miami	USA	6	-	6
Neurobiobank - University of Pittsburgh	USA	30	21	51
Neurobiobank - VA LA	USA	27	-	27
Newcastle University	UK	-	24	24
Northwestern	USA	-	6	6
Oregon Health Sciences University	USA	23	169	192
University of Michigan	USA	28	-	28
University of Kentucky	USA	-	75	75
University of Washington	USA	9	18	27
University of California, Irvine	USA	14	67	81
University of California, San Diego	USA	26	9	35
University of California, San Francisco	USA	50	-	50
University of Pennsylvania	USA	-	47	47
University of Texas, Southwestern	USA	-	44	44
University of Vienna	Austria	-	15	15
Washington University in St.Louis	USA	-	15	15
Total Samples	-	**455**	**886**	**1341**

SI Methods Table 5. Quality control summary for Stage 1 and Stage 2 PD data analysis

	Stage 1					Stage 2		
	NIA	NL	GER	FIN	Total	McGILL	SPAIN	Total
Initial sample size	4005	2796	1686	883	9370	1847	3444	5291
Missingness	37	3	10	21	71	5	0	5
Relatedness	6	68	3	2	79	0	0	0
Ancestry	56	0	0	0	56	0	0	0
Missing phenotype	0	0	0	0	0	357	0	357
Excluded (n)	99	71	13	23	206	362	0	362
Excluded (%)	2.47	2.54	0.77	2.60	2.20	19.60	0.00	19.60
Final sample size	**3906**	**2725**	**1673**	**860**	**9164**	**1485**	**3444**	**4929**
Cases	905	767	740	368	2780	582	2117	2699
Controls	3001	1958	933	492	6384	903	1327	2230

SI Methods Table 6. Quality control summary for Stage 1 and Stage 2 PSP data analysis

	Stage 1		Stage 2	
	ADGC combined	Hoglinger et al 2011	MSEC Controls	MSEC PSP
Initial sample size	11303	1112	886	455
Missingness	0	3	90	4
Relatedness	0	2	97	12
Ancestry	0	36	23	20
Missing phenotype	88	0	0	0
Excluded (n)	88	41	210	36
Excluded (%)	0.78	3.69	23.7	7.32
Final sample size	11215	1071	676	419
Cases	0	1071	0	419
Controls	11215	0	676	0

SI Methods Table 7. dPCR probe design

Probe ID	Source	Reference/Catalog ID	Label	Forward	Reverse	Probe
Alpha	ThermoFisher	Hs03955205_cn	FAM	-	-	-
Beta	ThermoFisher	Hs03971091_cn	FAM	-	-	-
Gamma	Custom	Boettger et al 2012	FAM	GTTGTTGACCATGGCTTCCT	GTGAGAAGACGGCCTTTGAG	CACATGTGTTCTGGAATGCC
LRRC37A	Custom	-	FAM	TGTGTGTGTGTGTGTGTTTGTG	CTGCTCTGCTTTCATTCAAACCTTT	TTTCCTTTTTGTGTCCATCTCTCTCC
MAPT	ThermoFisher	Hs07226271_cn	FAM	-	-	-
RNase P	ThermoFisher	4403326	VIC	-	-	-

SI Methods Table 8. Summary of iPSC line sources and 17q21.31 haplotypes.

Protein analysis

Line ID	Source	Sex	Haplotype
F11349	ADRC	Male	H2H2
F0510.2Δ2H1	ADRC	Male	H1H1
FA12455	ADRC	Female	H2H2
F13505	ADRC	Female	H1H1
3182-3	NIHCZ	Female	H2H2
F11421Δ2A07	ADRC	Female	H1H1

ATAC-seq data

Line ID	Source	Sex	Haplotype
iPS6	UCI	Male	Control
iPS4	UCI	Male	H1c
iPS13	UCI	Female	H1c
iPS14	UCI	Male	Control

ADRC: Knight Alzheimer's Disease Research Center at Washington University
NIHCZ: NIH Childhood-onset Schizophrenia study (Hoffman et al, 2017)
UCI: University of California Alzheimer's Disease Research Center

SI Methods Table 9. Summary of iPSC line ATAC-seq quality control metrics

Sample name	iPSC-derived cell line	Subject ID	Sex	Ethnicity	rs242557 genotype	Number of initial paired-end reads	Fraction of uniquely mapped paired-end reads to initial paired-end reads	Number of uniquely mapped paired-end reads	Fraction of mitochondrial pair-end reads to uniquely mapped, non-duplicated paired-end reads	Fraction of duplicated pair-end reads to uniquely mapped, non-duplicated paired-end reads	Number of non-duplicated, non-mitochondrial paired-end reads	PBC: PCR bottleneck coefficient	Number of peaks per sample	FRiP: Fraction of reads in peaks (upon consensus generated across all samples)
iPS14_C5_2_APOE_33_astros	Astrocyte	7028	Male	Caucasian	**GG**	58,581,279	0.818	47,922,790	0.134	0.675	13,193,607	0.377	35,052	0.141
iPS14_C5_2_APOE_33_neurons	Neuron	7028	Male	Caucasian	**GG**	45,862,944	0.869	39,856,214	0.035	0.155	32,258,539	0.863	20,419	0.094
7595_imgl	Microglia	7595	Male	Caucasian	**GG**	50,261,934	0.824	41,424,216	0.177	0.750	7,942,332	0.321	67,122	0.375
iPS6_C2_2_APOE44_neurons	Neuron	7595	Male	Caucasian	**GG**	39,819,560	0.868	34,545,824	0.037	0.178	27,126,736	0.846	21,366	0.097
iPS6_C2_2_astros_repeat	Astrocyte	7595	Male	Caucasian	**GG**	54,035,626	0.824	44,502,473	0.112	0.642	13,729,506	0.412	60,061	0.183
10100_imgl	Microglia	10100	Female	Caucasian	**AA**	54,139,167	0.824	44,588,251	0.057	0.425	23,410,423	0.645	77,477	0.373
iPS13_C2_1_APOE_33_astros	Astrocyte	10100	Female	Caucasian	**AA**	61,900,812	0.850	52,581,690	0.090	0.541	21,712,066	0.504	29,489	0.114
iPS13_C2_1_neurons	Neuron	10100	Female	Caucasian	**AA**	46,652,450	0.853	39,772,163	0.042	0.304	26,181,284	0.745	84,631	0.205
10104_imgl	Microglia	10104	Male	Caucasian	**AA**	57,641,250	0.850	48,971,593	0.095	0.590	17,676,114	0.466	77,615	0.370
iPS4_C15_3_APOE_33_astros	Astrocyte	10104	Male	Caucasian	**AA**	44,900,557	0.828	37,153,981	0.125	0.672	10,287,938	0.396	30,953	0.154
iPS4_C15_3_APOE_33_neurons	Neuron	10104	Male	Caucasian	**AA**	43,779,975	0.867	37,954,902	0.038	0.206	28,746,166	0.817	30,192	0.111

SI Methods Table 10. Taqman qRTPCR assays

Gene	Taqman assay ID
LRRC37A	Hs04191152_gH
CEP76	Hs00950370_m1
TMEM67	Hs00402347_m1
PLK4	Hs00179514_m1
KCNN4	Hs01069779_m1
ADGRG1	Hs00938474_m1
MFAP5	Hs00969608_g1
ACTB	Hs99999903_m1