

α -synuclein genetic variability: a biomarker for dementia in Parkinson's disease

Ilaria Guella¹, PhD, Daniel M. Evans¹, BSc, Chelsea Szu Tu¹, MSc, Ekaterina Nosova¹, PhD, Stephanie Bortnick¹, BSc, Silke Appel-Cresswell², MD, Joseph K Tsui², MD, Martin McKeown², MD, A. Jon Stoessl², MD, Jan O. Aasly^{3,4}, MD, Alex Rajput⁵, MD, Ali H. Rajput⁵, MD, Jennifer G. Goldman⁶, MD, John Dalrymple-Alford⁷, PhD, Jeroen Hoogland⁸, MD, Gert J. Geurtsen⁸, PhD, Irene Litvan⁹, MD, on behalf of the MDS PD-MCI Study Group, Alexandra I. Soto-Ortolaza¹⁰, BSc, Owen A. Ross¹⁰, PhD, Dennis W. Dickson¹⁰, MD, Lefkos T. Middleton¹¹, MD, Hazal Haytural¹², MSc, Laura Parkkinen¹², PhD, and Matthew J. Farrer, PhD*¹.

¹ Djavad Mowafaghian Centre for Brain Health, Department of Medical Genetics, University of British Columbia, Vancouver, BC, Canada;

² Pacific Parkinson's Research Centre, Department of Medicine (Neurology) and Djavad Mowafaghian Centre for Brain Health, University of British Columbia Vancouver, BC, Canada;

³ Department of Neuroscience, Norwegian University of Science and Technology, Trondheim, Norway;

⁴ Department of Neurology, St. Olav's Hospital, Trondheim, Norway;

⁵ Division of Neurology, University of Saskatchewan and Saskatoon Health Region, Saskatoon, SK, Canada;

⁶ Department of Neurological Sciences, Rush University Medical Center, Chicago, IL, USA;

⁷ New Zealand Brain Research Institute, Christchurch, New Zealand;

⁸ Department of Neurology, Academic Medical Center Amsterdam, The Netherlands;

⁹ Department of Neurosciences University of California, Movement Disorder Center, San Diego, CA, USA;

¹⁰ Department of Neuroscience, Mayo Clinic, Jacksonville, FL 32224, USA;

¹¹ School of Public Health, Faculty of Medicine, Imperial College, St. Mary's Campus, London, UK;

¹² Oxford Parkinson's Disease Centre, Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, UK.

* Corresponding Author:

Department of Medical Genetics, University of British Columbia, Vancouver, BC, Canada, V6T 1Z3. Tel: (604) 822 7753. Fax: (604) 822-0361; E-mail: mfarrer@can.ubc.ca

Running head: SNCA variability and cognitive decline

Keywords: SNCA, cognitive decline, Lewy body disorders.

Title character count (with spaces): 80, Running head character count: 38

Abstract word count: 250, Text word count: xxx; Tables: 2; Figures: 2; Supplementary tables 6 and Supplementary figure 2; References: 57

Abstract

Objective:

The relationship between Parkinson's disease (PD), PD with dementia (PDD) and dementia with Lewy bodies (DLB) is long debated. Although PD is primarily considered a motor disorder, cognitive impairment is often present at diagnosis, and only ~20% of patients remain cognitively intact in the long term. Alpha-synuclein (*SNCA*) was first implicated in the pathogenesis of the disease when point mutations and locus multiplications were identified in familial parkinsonism with dementia. In world-wide populations *SNCA* genetic variability remains the most reproducible risk factor for idiopathic PD. However, few investigators have looked at *SNCA* variability in terms of cognitive outcomes.

Methods

We have used targeted high-throughput sequencing to characterize the 135kb *SNCA* locus in a large multi-national cohort of patients with PD, PDD, DLB and healthy controls.

Results

An analysis of 44 tagging single nucleotide polymorphisms across the *SNCA* locus shows two distinct association profiles for symptoms of parkinsonism and/or dementia, respectively towards the 3' or the 5' of the *SNCA* gene. In addition, we define a specific haplotype in intron 4 that is directly associated with PDD. The PDD risk haplotype has been interrogated at single nucleotide resolution and is uniquely tagged by an expanded TTTC_n repeat.

Interpretation

Our data show that PD, PDD and DLB, rather than a disease continuum, have distinct genetic aetiologies albeit within one genomic locus. Such results may serve as prognostic biomarkers to

these disorders, to inform physicians and patients and to assist in the design and stratification of clinical trials aimed at disease modification.

Introduction

Parkinson's disease (PD) has been traditionally defined by characteristic clinical motor hallmarks of bradykinesia, tremor, muscular rigidity and postural instability. However, the non-motor aspects of PD, including cognitive impairment, are now increasingly recognized as a common feature of the disease. At the time of diagnosis, approximately 24% of PD patients have mild cognitive impairment (PD-MCI) ^{1, 2} and approximately 80% of longitudinally followed patients with PD develop dementia (PDD) during the course of the disease ^{3, 4}. The presence of cognitive impairment in patients with PD is associated with lower quality of life, increased nursing home placement and mortality ⁵. Clinically, the cognitive features of PDD are similar to and often indistinguishable from dementia with Lewy bodies (DLB) ^{6, 7}. The two dementia syndromes are arbitrarily differentiated based on the timing of the motor PD signs relative to the onset of dementia (i.e. diagnosis of DLB is assigned when motor symptoms and dementia appear together or within one year of each other)⁸. However, the overlap in the clinical presentation often causes difficulties in the diagnostic process. In addition to the clinical phenotypic similarities, PDD and DLB also share common neuropathological features, since the burden of cortical Lewy bodies and neurites is often indistinguishable ^{7, 9-12}. Recent studies suggest increased cortical LB and A β deposition in temporal and parietal regions may be distinguishing features of DLB, compared to PDD ¹³, and potentially mediated by APOE ϵ 4 ¹⁴. The aetiopathogenic mechanisms of DLB and PDD still remain unclear and they are often considered as two manifestations of one continuous spectrum of disease. Genetic factors may play a role in the expression of cognitive deficits in PDD and DLB, as suggested by dominant familial forms of PDD/DLB. Notably, missense mutations in the α -synuclein gene (*SNCA*) and

locus multiplications are associated with clinical and pathological phenotypes ranging from PD to PDD to DLB ¹⁵. In world-wide populations *SNCA* genetic variability remains the most reproducible risk factor for idiopathic PD. However, only few investigators have looked at *SNCA* variability in terms of these different clinico-pathological groups. In this study, we have used targeted high-throughput sequencing to comprehensively characterize the 135kb *SNCA* locus in a large multi-national cohort of patients with PD, PDD, DLB and healthy controls.

Methods

Subjects

All sites received approval from an ethical standards committee on human experimentation before study initiation and obtained written informed consent for research from all individuals participating in the study. A total of 1492 PD, 922 DLB and 978 healthy controls (HC) samples, originating from 8 cohorts, were included in the study (Table 1). All samples are of self-declared European or North American ancestry. Clinical examinations were performed by movement disorders specialty-trained neurologists and diagnoses made using established criteria ¹⁶, the UK Brain Bank Criteria for PD ¹⁷, and the DLB Consortium ¹⁸. PD patients were classified without cognitive impairment (noCI) or with dementia (PDD) according to the Movement Disorder Society (MDS) Task Force criteria ⁶, or using Montreal Cognitive Assessments (MoCA) taking into account the mean score, minimum score and at last examination ¹⁹⁻²². Patients with raw MoCA scores >21 but <26 were considered to have some degree of cognitive impairment, and were not used in stratified cognitive analyses. When quantitative scores were unavailable, a qualitative diagnosis of PDD was made on the basis of longitudinal evaluations and clinical impression (n=57, UK/ICL: PD Brain Bank).

Genetic screening

SNCA gene dosage was assessed by quantitative real-time PCR ²³. Short tandem repeat (STR) genotyping was performed using fluorescent-labeled primer PCR reaction with capillary electrophoresis on an ABI3730xl Genome Analyzer and analyzed with Genemapper software

(Life Technologies). All subjects were genotyped for APOE $\epsilon 2/\epsilon 3/\epsilon 4$ using a TaqMan SNP Genotyping Assay (Life Technologies, Carlsbad, CA, USA).

High Throughput Sequencing (HTS)

The entire 135kb *SNCA* genomic locus (chr4:90,635,215-90,769,364) was sequenced as part of a custom designed high-throughput sequencing (HTS) panel, capturing the exonic regions of candidate genes previously associated or linked to neurodegenerative disease. Pair-end sequencing was performed on a SOLiD 5500xl platform (Life Technologies, Carlsbad, CA, USA) as previously described ²⁴. Mapping, sequence alignment, duplicate removal, SNP calling and indel detection were performed by Lifescope v2.5.1 (Life Technologies). Annotation was performed with ANNOVAR ²⁵ using NCBI Build 37 (hg19) as the reference genome.

SNPs selection

Forty-four SNPs (Supplementary table 1) were selected for the *SNCA* locus using the TAGGER program as implemented in HaploView 4.1 ²⁶ with parameters of minor allele frequency (MAF) > 5% and pairwise r^2 threshold of 0.8.

Genotyping

SAMtools (version 0.1.18) ²⁷ was used to generate genotype calls from individual BAM files. Genotypes with depth of coverage less than 10 were set as missing. Additional genotyping of 44 SNPs was carried out by Sequenom MassArray iPLEX system (Sequenom, San Diego, CA). Sequenom primers were designed using MassARRAY Designer 4.0 software (Sequenom, CA). PCR amplification, shrimp alkaline phosphatase (SAP) treatment and single-base extension and

desalting were performed in 384-well microplates (Thermo Fisher Scientific, Fremont, CA) using Sequenom PCR reagents according to the manufacturers protocol. Reproducibility was assessed by comparing replicated samples both within and across platforms. A genotype call rate >95% and a $P > 0.01$ for test of deviation from Hardy-Weinberg equilibrium (HWE) were used as quality-control criteria. Samples with more than 5% missing genotypes were removed from the study.

Statistical analysis

Highly polymorphic genetic variability in candidate genes (MAF >0.2), beyond the SNCA locus, was used in a factor analysis to generate eigenvectors and correct for potential population stratification, as previously described (Steele, Guella et al. 2015). HWE was tested in PLINK ²⁸ and markers that deviated from expectation ($p < 0.001$) were excluded ²⁹. Association testing was performed using the logistic regression function in PLINK, using gender, age/age at death, site and APOE dosage as covariates. Pairwise linkage disequilibrium (LD) was calculated for all 44 SNCA SNPs. Twenty-one markers in disequilibrium ($r^2 > 0.8$) were excluded from the Bonferroni correction, and the significance level was set to $0.05/23 = 0.002$. Odd ratios (ORs) and 95% confidence intervals (CI) are calculated for the minor allele.

The linkage disequilibrium (LD) structure of the SNCA locus was assessed with the software package Haploview version 4.1 ²⁶. For each block only haplotypes with frequency >0.01 were considered. Logistic regression analysis implemented in PLINK was used to test the association, and multiple testing was adjusted for using the max(T) permutation procedure ($n=10,000$) ²⁸.

Results

The final study population consisted of 1492 PD, 922 DLB and 978 HC samples; demographic and clinical characteristics, including those with autopsy, are summarized in Table 2. All subjects were wild-type for *SNCA* multiplication and without known pathogenic mutations for PD. APOE genotypes and allele frequencies observed in the samples are reported in Supplementary table 2. There was no significant difference in allele or genotype distribution between patients with PD and HC. However, an overrepresentation of the APOE $\epsilon 4$ allele in both PDD (OR [95%CI]=1.28 [1.11-1.48], $p=0.09$) and DLB (OR [95%CI]=2.50 [2.29-2.70], $p<0.001$) groups was observed.

Genotypes for 44 SNPs (MAF>5%) spanning the entire *SNCA* locus were obtained for the cohort. Sequencing of the *SNCA* locus was performed in 1366 PD, 122 DLB and 490 HC samples, selected as they had the most detailed history and sufficient DNA. Overall, 92% of the *SNCA* locus was sequenced with a minimum average depth >20 \times , across all the samples. Regions with no coverage were found to be in or near repetitive elements (Supplementary figure 1). An additional 126 PD, 800 DLB and 488 HC samples were genotyped for the 44 *SNCA* SNPs using Sequenom technology. All SNPs had a genotyping call rate >90%, MAF > 5% and were in HWE in control subjects. After LD pruning ($r^2>0.8$) 23 SNPs were selected for single SNP association analysis (Supplementary table 1). Logistic-regression analysis was used to test for association between the 23 tagging SNPs and disease status (PD, PD-noCI, PDD or DLB vs HC) with and without adjusting for the following covariates; age, gender, site and APOE $\epsilon 4$ dosage. Allele frequencies, ORs and p-values are reported in Supplementary table 3. Results are displayed in Figure 1. After correction for multiple testing, six SNPs (rs356220, rs356225, rs3857057,

rs10018362, rs2737029, rs7689942) showed a statistically significant association with PD, and three SNPs (rs62306323, rs974711, rs1348224) reached statistical significance in the DLB samples. All SNPs except rs62306323 remained significantly associated after adjusting for covariates. Cognitive assessments were available for 1067 (72%) of 1492 patients with PD; 572 patients were classified as PD-noCI and 198 as PDD. The remaining 297 patients were considered to have some degree of cognitive impairment, and were not included in subsequent analyses. After Bonferroni correction, rs356220 and rs10018362 reached the statistical significance in the PD-noCI samples, and rs10018362 remained significant after covariate adjustment, whereas in the PDD group three SNPs (rs10018362, rs7689942, rs1348224) showed a statistically significant association that remained significant after covariate adjustment for rs10018362 and rs7689942.

Further, haplotype-based association analysis was performed for SNPs within LD blocks (Figure 1 and Supplementary table 4). A highly significant risk haplotype was identified in both the PD and PD-noCI groups (frequency: 6.7% in PD, 8.2% in PD-noCI, 2.4% in HC; OR [95% CI]=3.24 [2.27-4.63], $p=9.40 \times 10^{-11}$, $p\text{-perm} < 1 \times 10^{-4}$, OR [95% CI]=3.88 [2.63-5.73], $p=9.21 \times 10^{-12}$, $p\text{-perm} < 1 \times 10^{-4}$ in PD and PD-noCI respectively). The risk haplotype, spanning approximately 74kb from intron 4 to the 3' end of *SNCA*, is tagged by rs356218-A, rs356189-T, rs356185-G, rs2737029-C. In DLB the same haplotype was found at an extremely low frequency and is inversely associated with disease (frequency: 0.6%, OR [95% CI]=0.23 [0.11-0.47], $p=6.89 \times 10^{-5}$, $p\text{-perm}=9.00 \times 10^{-4}$). Two alternative 11kb haplotypes in *SNCA* intron 4 were also significantly associated with increased risk of PDD (rs62306323-C rs7689942-T; frequency: 9.1% in PDD, 4.9% in HC; OR [95%CI]=2.01 [1.33-3.04], $p=9.18 \times 10^{-4}$, $p\text{-perm}=0.01$) or DLB (rs62306323-T,

rs7689942-C; frequency: 15.3% in DLB, 11.6% in HC; OR [95%CI]=1.37 [1.13-1.66], $p=0.001$, $p\text{-perm}=0.02$).

To identify the complete set of DNA variants in the 11kb associated haplotype in *SNCA* intron 4, all variants within this region were extracted for the samples that underwent HTS (1366 PD, 122 DLB and 490 HC). A total of 79 SNVs were identified of which 45 had a MAF < 0.01 and 15 were novel relative to dbSNP build 142. Of the 79 variants in the interval 11 are in complete LD ($r^2>0.98$) with the PDD associated SNP (rs7689942), whereas none is in LD ($r^2>0.5$) with the DLB associated SNP (rs62306323). Haplotypes reconstruction identified 21 distinct haplotypes of which 11 had a frequency > 0.01 (Figure 2). Remarkably, both the DLB and PDD risk haplotypes were uniquely tagged by SNPs included in the initial set of 44 SNPs. Haplotype association analysis confirmed the association of the rs62306323-C rs7689942-T haplotype with PDD (frequency: 9.7% in PDD, 5.0% in HC; OR [95%CI]=2.14 [1.33-3.43], $p=0.002$, $p\text{-perm}=0.01$, Supplementary table 5). The number of rare variants in the 11kb region was no different between PDD risk haplotype carriers and non-carriers.

Four repeated elements (AluJb, chr4:90716499-90716796; AluSx1, chr4:90717144-90717436; TTTC_n repeat, chr4:90723737-90723915; THE1D, chr4:90724732-90725112) within the 11 kb region were not covered by HTS (Figure 2). These regions were Sanger sequenced in homozygote subjects for each of the 6 common haplotypes (frequency > 0.5). Three repeated elements (AluJb, AluSx1, THE1D) were wild-type in all subjects. However, different haplotypes had a variable number of TTTC_n repeats (see supplementary table 6 for a detailed description of the repeat structure). Genotyping of the repeat in 540 individuals (239 PD, 281 HC) showed that each haplotype is associated with a specific repeat size (ranging from 289bp to 301bp), with the

exception of the PDD risk haplotype, which is associated with multiple repeat sizes (all >309bp) (Figure 2). Subsequent genotyping of all the PDD risk haplotype carriers (81 HC and 180 PD) revealed that the PDD risk allele is uniquely tagged by an expanded TTTC_n repeat (size ranging from 309 to 345 bp). Nevertheless, among carriers of the PDD haplotype the distribution of the repeat sizes was not different between diagnostic groups (Supplementary figure 2).

Discussion

In the present study, we have explored the contribution of *SNCA* genetic variability to PD, PD-noCI, PDD and DLB. These disorders share similarities in motor and cognitive dysfunction, and are characterized by Lewy body pathology. Results from an analysis of 44 tagging SNPs spanning the entire *SNCA* locus show two distinct association profiles for symptoms of parkinsonism and/or dementia respectively, towards the 3' or the 5' of the *SNCA* gene. In addition, we identify a specific haplotype in *SNCA* intron 4 that is directly associated with PDD. Collectively, our results suggest that PD, PDD and DLB, rather than a disease continuum have distinct genetic aetiologies albeit within one locus.

Cognitive decline is one of the most debilitating manifestations of disease progression in PD and it has an important influence on patient management and prognosis. The incidence rate of dementia is estimated to be at least fourfold higher among patients with PD than in the general population ³. However, only few genetic studies have been conducted in this area and the specific genetic contributions to cognitive impairment are still poorly understood. Herein, we have looked at *SNCA* variability in patients with PD at the two extremes of the cognitive spectrum. Genetic variability within the *SNCA* gene has been unequivocally associated with sporadic PD susceptibility ³⁰⁻³². Our results show a 74 kb haplotype, from intron 4 to the 3' end of *SNCA*, is significantly associated with increased risk of PD (OR [95% CI]=3.24 [2.27-4.63], $p=9.40 \times 10^{-11}$, $p\text{-perm} < 1 \times 10^{-4}$). These results are consistent with the literature showing that, variants towards the 3' end of the *SNCA* gene convey risk for PD ³³⁻³⁷. The association profile in PD without cognitive impairment overlaps the one observed for all patients with PD. Indeed in the PD-noCI group the 74kb haplotype exhibits a stronger effect and more significant

association (OR [95% CI]=3.88 [2.63-5.73], $p=9.21 \times 10^{-12}$, $p\text{-perm} < 1 \times 10^{-4}$), despite the smaller sample size, that suggests this haplotype may have a more specific role in cognitively preserved patients with PD.

In agreement with previous studies^{38,39}, our results show significant association of the *SNCA* locus with PD and DLB, and show that alleles conferring that risk are different in these two diseases. The top DLB associated SNP (rs1348224, OR [95% CI]=0.71 [0.61-0.83], $p=1.1 \times 10^{-5}$) is located 2.5 kb upstream the *SNCA* gene. This SNP is in almost complete LD ($r^2 > 0.95$, LD calculation based on genotypes extracted from the samples that underwent *SNCA* locus HTS) with rs894280, the top SNP recently reported by Bras et al.,³⁸. rs894280 and rs1348224 are 1063bp apart (chr4:90760883-90761946) and show comparable odds ratios in the same direction. However, they are frequent ($MAF_{HC}=0.50$) and tag a common ancestral haplotype. Higher-resolution nucleotide sequencing 5' of the *SNCA* gene is now warranted in additional patients with DLB to precisely define the functional variant(s).

Extending prior studies, our analysis reveals a significant association between an 11kb haplotype located more 5' in intron 4 in both PDD and DLB albeit *on alternate alleles*; the PDD risk haplotype (OR [95%CI]=2.01 [1.33-3.04], $p=9.18 \times 10^{-4}$, $p\text{-perm}=0.01$) is defined by rs62306323-C, rs7689942-T whereas the DLB risk haplotype (OR [95%CI]=1.37 [1.13-1.66], $p=0.001$, $p\text{-perm}=0.02$) is defined by rs62306323-T and rs7689942-C. With HTS we have ultimately resolved *SNCA* disease-associated haplotypes, at single nucleotide resolution, to identify all genetic variability and its potential contribution to disease. The DLB/PPD haplotype overlaps with strong H3K4Me1 and H3K27Ac marks, histone modifications that denote an active enhancer^{40 41, 42} (Figure 2), suggesting variants confer disease risk through cis-regulation

of *SNCA* expression. In addition, complementary analysis of repeated elements within the 11kb haplotype found a novel TTTC_n repeat to be expanded, and which uniquely tags PDD risk haplotype based on single nucleotide variability. Further studies are needed to identify the specific functions of these enhancer elements, the variants within them and their impact on gene expression and the distribution and burden of brain Lewy pathology.

Several studies have reported dysregulation of *SNCA* expression in sporadic PD brains ⁴³⁻⁴⁶, whereas variability in 5' and/or 3' regions of the gene have been shown to affect *SNCA* expression levels ^{35, 47-49}. Many efforts have been made in order to investigate possible regulatory elements. Sterling and colleagues ⁵⁰ analyzed conserved non-coding genomic regions across the *SNCA* locus and identified 12 cis-regulatory regions that exhibit either an increase or reduction of the expression of the reporter gene. The element with the highest fold change (2.5 fold) is located in intron 4 (l12:chr4: 90721509-90721763) within the PDD/DLB associated haplotype. Recently, an intronic CT-rich region in *SNCA* intron 4 was found to be associated with an increased risk to develop Lewy body pathology in Alzheimer's disease ⁵¹. The CT-rich region resides within the histone modification mark signal suggesting that it acts as an enhancer element of transcription regulation. The minor allele of rs2298728 uniquely tagged the associated haplotype. However, no association was detected in our study between PD or DLB and rs17016193 (in nearly complete LD ($r^2 > 0.96$) with rs2298728, LD calculation based on genotypes extracted from the samples that underwent *SNCA* locus HTS).

A limitation of this cross-sectional study is that we analysed phenotype data related to PD and cognition at one point in time. Detailed longitudinal cognitive assessments are only available for a subset of subjects for the PD-MCI Consortium (PD-MCI, Table 1). Therefore,

some of the *de-novo* patients with PD with the PPMI Consortium may develop dementia later in the disease course. We accounted for predictors of cognitive function by including demographic characteristics (e.g., age, gender and APOE ϵ 4 dosage) in the regression models, age being the most prominent risk factor for PDD^{52, 53}. However, cognitive impairment in PD also correlates with the severity of motor disability⁵⁴ and cognitive tests scores should be corrected for education²¹, which were not available for most of the subjects. Although our initial sample size was large, groups with cognitive data are much smaller and power to detect novel associations of small to intermediate effect size may not have been identified.

Our results provide replication of *SNCA* association with DLB, and further extend the concept of allelic heterogeneity for the *SNCA* locus to PDD. The identification of genetic predictors of cognitive decline could be used to plan therapeutic trials that could slow cognitive progression, optimize stratification in clinical trials and advance clinical care (reviewed by⁵⁵). Longitudinal studies investigating *SNCA* variability and rate of progression of cognitive decline, the distribution and burden of Lewy pathology and *SNCA* gene and protein expression are now warranted in genotype-defined cases.

Acknowledgement:

We very much wish to thank all the individuals and families who generously participated in this research. We appreciate the technical contribution from F. T. Pishotta and C. Thompson. Data used in the preparation of this article were obtained from the Parkinson's Progression Markers Initiative (PPMI) database (<http://www.ppmi-info.org/data>). For up-to-date information on the study, visit www.ppmi-info.org. Biospecimens used in the analyses presented in this article were obtained from the Parkinson's Progression Markers Initiative (PPMI) (www.ppmi-info.org/specimens). As such, the investigators within PPMI contributed to the design and implementation of PPMI and/or provided data and collected biospecimens but did not participate in the analysis or writing of this report. The complete list of PPMI Investigators can be found at <http://www.ppmi-info.org/Authorslist>. PPMI – a public-private partnership – is funded by the Michael J. Fox Foundation for Parkinson's Research and funding partners, including [list the full names of all of the PPMI funding partners found at www.ppmi-info.org/fundingpartners].

This research was undertaken, in part, thanks to funding from the Canada Excellence Research Chairs program (MJF), Leading Edge Endowment Funds provided by the Province of British Columbia, LifeLabs, and Genome BC support the Dr. Donald Rix BC Leadership Chair (MJF), and the Cundill Foundation (MJF). We acknowledge the Oxford Brain Bank, supported by the Medical Research Council (MRC), Brains for Dementia Research (BDR) (Alzheimer Society and Alzheimer Research UK), Autistica UK and the NIHR Oxford Biomedical Research Centre.

Author Contributions:

IG - conception and design of the study; analysis and interpretation of data, drafting the manuscript. DME, EN, HH - acquisition and analysis of data, CST, SB, AISO - execution of the study. SAC, AJT, MM, JKS, JOA, AR, AOR, JGG, JDA, GJG, IL, OAR, DWD, LTM, JH, LP - samples and clinical data collection, revising manuscript. MJF - conception and design of the study, interpretation of data and revising manuscript.

Potential Conflicts of Interest

IG, DME, EN, CST, SB have nothing to disclose. MJF reports grants from Canadian Federal Government, the Cundill Foundation and BC Leading Edge Endowment during the conduct of this study. Personal fees from Bristol Myers Squibb, Genentech and Teva were paid outside the submitted work. In addition, MJF has a patent on genetic variability in LRRK2 and Parkinson's disease (US8409809, US8455243B2) and on related mouse models on which royalties have been paid. MJF is on the Editorial Boards of the Journal of Parkinson's disease

and Parkinsonism and related Disorders. He currently serves on the scientific advisory boards of the Michael J. Fox Foundation, Parkinson's Society Canada and Parkinson's UK.

JGG reports grants from NIH/NINDS, Michael J. Fox Foundation, Rush University during the conduct of the study; JGG is site-PI for clinical trial from Teva and Biotie, and Consultant for Pfizer and Teva outside the submitted work.

LP reports grants from Parkinson's UK and Michael J. Fox Foundation during the conduct of the study; LP is on the Editorial Board of Acta Neuropathologica Communications and serves on the Assessment Panel of the Parkinson's UK Tissue Bank.

References

1. Muslimovic D, Post B, Speelman JD, Schmand B. Cognitive profile of patients with newly diagnosed Parkinson disease. *Neurology*. 2005 Oct 25;65(8):1239-45.
2. Litvan I, Aarsland D, Adler CH, et al. MDS Task Force on mild cognitive impairment in Parkinson's disease: critical review of PD-MCI. *Movement disorders : official journal of the Movement Disorder Society*. 2011 Aug 15;26(10):1814-24.
3. Aarsland D, Andersen K, Larsen JP, Lolk A, Kragh-Sorensen P. Prevalence and characteristics of dementia in Parkinson disease: an 8-year prospective study. *Archives of neurology*. 2003 Mar;60(3):387-92.
4. Hely MA, Reid WG, Adena MA, Halliday GM, Morris JG. The Sydney multicenter study of Parkinson's disease: the inevitability of dementia at 20 years. *Movement disorders : official journal of the Movement Disorder Society*. 2008 Apr 30;23(6):837-44.
5. Rosenthal E, Brennan L, Xie S, et al. Association between cognition and function in patients with Parkinson disease with and without dementia. *Movement disorders : official journal of the Movement Disorder Society*. 2010 Jul 15;25(9):1170-6.
6. Emre M, Aarsland D, Brown R, et al. Clinical diagnostic criteria for dementia associated with Parkinson's disease. *Movement disorders : official journal of the Movement Disorder Society*. 2007 Sep 15;22(12):1689-707; quiz 837.
7. Lippa CF, Duda JE, Grossman M, et al. DLB and PDD boundary issues: diagnosis, treatment, molecular pathology, and biomarkers. *Neurology*. 2007 Mar 13;68(11):812-9.
8. McKeith IG. Consensus guidelines for the clinical and pathologic diagnosis of dementia with Lewy bodies (DLB): report of the Consortium on DLB International Workshop. *Journal of Alzheimer's disease : JAD*. 2006;9(3 Suppl):417-23.

9. Ballard C, Ziabreva I, Perry R, et al. Differences in neuropathologic characteristics across the Lewy body dementia spectrum. *Neurology*. 2006 Dec 12;67(11):1931-4.
10. Guo L, Itaya M, Takanashi M, Mizuno Y, Mori H. Relationship between Parkinson disease with dementia and dementia with Lewy bodies. *Parkinsonism & related disorders*. 2005 Aug;11(5):305-9.
11. Harding AJ, Halliday GM. Cortical Lewy body pathology in the diagnosis of dementia. *Acta neuropathologica*. 2001 Oct;102(4):355-63.
12. Jellinger KA. Morphological substrates of parkinsonism with and without dementia: a retrospective clinico-pathological study. *Journal of neural transmission Supplementum*. 2007(72):91-104.
13. Howlett DR, Whitfield D, Johnson M, et al. Regional Multiple Pathology Scores Are Associated with Cognitive Decline in Lewy Body Dementias. *Brain pathology*. 2015 Jul;25(4):401-8.
14. Ruffmann C, Calboli FC, Bravi I, et al. Cortical Lewy bodies and Abeta burden are associated with prevalence and timing of dementia in Lewy body diseases. *Neuropathol Appl Neurobiol*. 2015 Nov 3.
15. Pouloupoulos M, Levy OA, Alcalay RN. The neuropathology of genetic Parkinson's disease. *Movement disorders : official journal of the Movement Disorder Society*. 2012 Jun;27(7):831-42.
16. Rajput AH, Rozdilsky B, Rajput A. Accuracy of clinical diagnosis in parkinsonism--a prospective study. *The Canadian journal of neurological sciences Le journal canadien des sciences neurologiques*. 1991 Aug;18(3):275-8.
17. Hughes AJ, Daniel SE, Kilford L, Lees AJ. Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. *Journal of neurology, neurosurgery, and psychiatry*. 1992 Mar;55(3):181-4.
18. McKeith IG, Dickson DW, Lowe J, et al. Diagnosis and management of dementia with Lewy bodies: third report of the DLB Consortium. *Neurology*. 2005 Dec 27;65(12):1863-72.

19. Dalrymple-Alford JC, MacAskill MR, Nakas CT, et al. The MoCA: well-suited screen for cognitive impairment in Parkinson disease. *Neurology*. 2010 Nov 9;75(19):1717-25.
20. Hoops S, Nazem S, Siderowf AD, et al. Validity of the MoCA and MMSE in the detection of MCI and dementia in Parkinson disease. *Neurology*. 2009 Nov 24;73(21):1738-45.
21. Hu MT, Szewczyk-Krolikowski K, Tomlinson P, et al. Predictors of cognitive impairment in an early stage Parkinson's disease cohort. *Movement disorders : official journal of the Movement Disorder Society*. 2014 Mar;29(3):351-9.
22. Nazem S, Siderowf AD, Duda JE, et al. Montreal cognitive assessment performance in patients with Parkinson's disease with "normal" global cognition according to mini-mental state examination score. *Journal of the American Geriatrics Society*. 2009 Feb;57(2):304-8.
23. Johnson J, Hague SM, Hanson M, et al. SNCA multiplication is not a common cause of Parkinson disease or dementia with Lewy bodies. *Neurology*. 2004 Aug 10;63(3):554-6.
24. Steele JC, Guella I, Szu-Tu C, et al. Defining neurodegeneration on Guam by targeted genomic sequencing. *Annals of neurology*. 2015 Mar;77(3):458-68.
25. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic acids research*. 2010 Sep;38(16):e164.
26. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*. 2005 Jan 15;21(2):263-5.
27. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009 Aug 15;25(16):2078-9.
28. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *American journal of human genetics*. 2007 Sep;81(3):559-75.
29. Wigginton JE, Cutler DJ, Abecasis GR. A note on exact tests of Hardy-Weinberg equilibrium. *American journal of human genetics*. 2005 May;76(5):887-93.

30. Nalls MA, Pankratz N, Lill CM, et al. Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. *Nature genetics*. 2014 Sep;46(9):989-93.
31. Satake W, Nakabayashi Y, Mizuta I, et al. Genome-wide association study identifies common variants at four loci as genetic risk factors for Parkinson's disease. *Nature genetics*. 2009 Dec;41(12):1303-7.
32. Sharma M, Ioannidis JP, Aasly JO, et al. Large-scale replication and heterogeneity in Parkinson disease genetic loci. *Neurology*. 2012 Aug 14;79(7):659-67.
33. Elbaz A, Ross OA, Ioannidis JP, et al. Independent and joint effects of the MAPT and SNCA genes in Parkinson disease. *Annals of neurology*. 2011 May;69(5):778-92.
34. Mata IF, Shi M, Agarwal P, et al. SNCA variant associated with Parkinson disease and plasma alpha-synuclein level. *Archives of neurology*. 2010 Nov;67(11):1350-6.
35. Sotiriou S, Gibney G, Baxevas AD, Nussbaum RL. A single nucleotide polymorphism in the 3'UTR of the SNCA gene encoding alpha-synuclein is a new potential susceptibility locus for Parkinson disease. *Neuroscience letters*. 2009 Sep 18;461(2):196-201.
36. Trotta L, Guella I, Solda G, et al. SNCA and MAPT genes: Independent and joint effects in Parkinson disease in the Italian population. *Parkinsonism & related disorders*. 2012 Mar;18(3):257-62.
37. Winkler S, Hagenah J, Lincoln S, et al. alpha-Synuclein and Parkinson disease susceptibility. *Neurology*. 2007 Oct 30;69(18):1745-50.
38. Bras J, Guerreiro R, Darwent L, et al. Genetic analysis implicates APOE, SNCA and suggests lysosomal dysfunction in the etiology of dementia with Lewy bodies. *Hum Mol Genet*. 2014 Dec 1;23(23):6139-46.
39. Nishioka K, Wider C, Vilarino-Guell C, et al. Association of alpha-, beta-, and gamma-Synuclein with diffuse lewy body disease. *Archives of neurology*. 2010 Aug;67(8):970-5.

40. Corradin O, Scacheri PC. Enhancer variants: evaluating functions in common disease. *Genome medicine*. 2014;6(10):85.
41. Heintzman ND, Hon GC, Hawkins RD, et al. Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature*. 2009 May 7;459(7243):108-12.
42. Kimura H. Histone modifications for human epigenome analysis. *Journal of human genetics*. 2013 Jul;58(7):439-45.
43. Beyer K, Ispierto L, Latorre P, Tolosa E, Ariza A. Alpha- and beta-synuclein expression in Parkinson disease with and without dementia. *Journal of the neurological sciences*. 2011 Nov 15;310(1-2):112-7.
44. Chiba-Falek O, Lopez GJ, Nussbaum RL. Levels of alpha-synuclein mRNA in sporadic Parkinson disease patients. *Movement disorders : official journal of the Movement Disorder Society*. 2006 Oct;21(10):1703-8.
45. Kingsbury AE, Daniel SE, Sangha H, Eisen S, Lees AJ, Foster OJ. Alteration in alpha-synuclein mRNA expression in Parkinson's disease. *Movement disorders : official journal of the Movement Disorder Society*. 2004 Feb;19(2):162-70.
46. Neystat M, Lynch T, Przedborski S, Kholodilov N, Rzhetskaya M, Burke RE. Alpha-synuclein expression in substantia nigra and cortex in Parkinson's disease. *Movement disorders : official journal of the Movement Disorder Society*. 1999 May;14(3):417-22.
47. Linnertz C, Saucier L, Ge D, et al. Genetic regulation of alpha-synuclein mRNA expression in various human brain tissues. *PloS one*. 2009;4(10):e7480.
48. McCarthy JJ, Linnertz C, Saucier L, et al. The effect of SNCA 3' region on the levels of SNCA-112 splicing variant. *Neurogenetics*. 2011 Feb;12(1):59-64.

49. Fuchs J, Tichopad A, Golub Y, et al. Genetic variability in the SNCA gene influences alpha-synuclein levels in the blood and brain. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*. 2008 May;22(5):1327-34.
50. Sterling L, Walter M, Ting D, Schule B. Discovery of functional non-coding conserved regions in the alpha-synuclein gene locus. *F1000Research*. 2014;3:259.
51. Lutz MW, Saul R, Linnertz C, Glenn OC, Roses AD, Chiba-Falek O. A cytosine-thymine (CT)-rich haplotype in intron 4 of SNCA confers risk for Lewy body pathology in Alzheimer's disease and affects SNCA expression. *Alzheimer's & dementia : the journal of the Alzheimer's Association*. 2015 Jun 13.
52. Kempster PA, O'Sullivan SS, Holton JL, Revesz T, Lees AJ. Relationships between age and late progression of Parkinson's disease: a clinico-pathological study. *Brain : a journal of neurology*. 2010 Jun;133:1755-62.
53. Williams-Gray CH, Evans JR, Goris A, et al. The distinct cognitive syndromes of Parkinson's disease: 5 year follow-up of the CamPaIGN cohort. *Brain : a journal of neurology*. 2009 Nov;132(Pt 11):2958-69.
54. Levy G, Tang MX, Cote LJ, et al. Motor impairment in PD: relationship to incident dementia and age. *Neurology*. 2000 Aug 22;55(4):539-44.
55. Chen-Plotkin AS. Unbiased approaches to biomarker discovery in neurodegenerative diseases. *Neuron*. 2014 Nov 5;84(3):594-607.

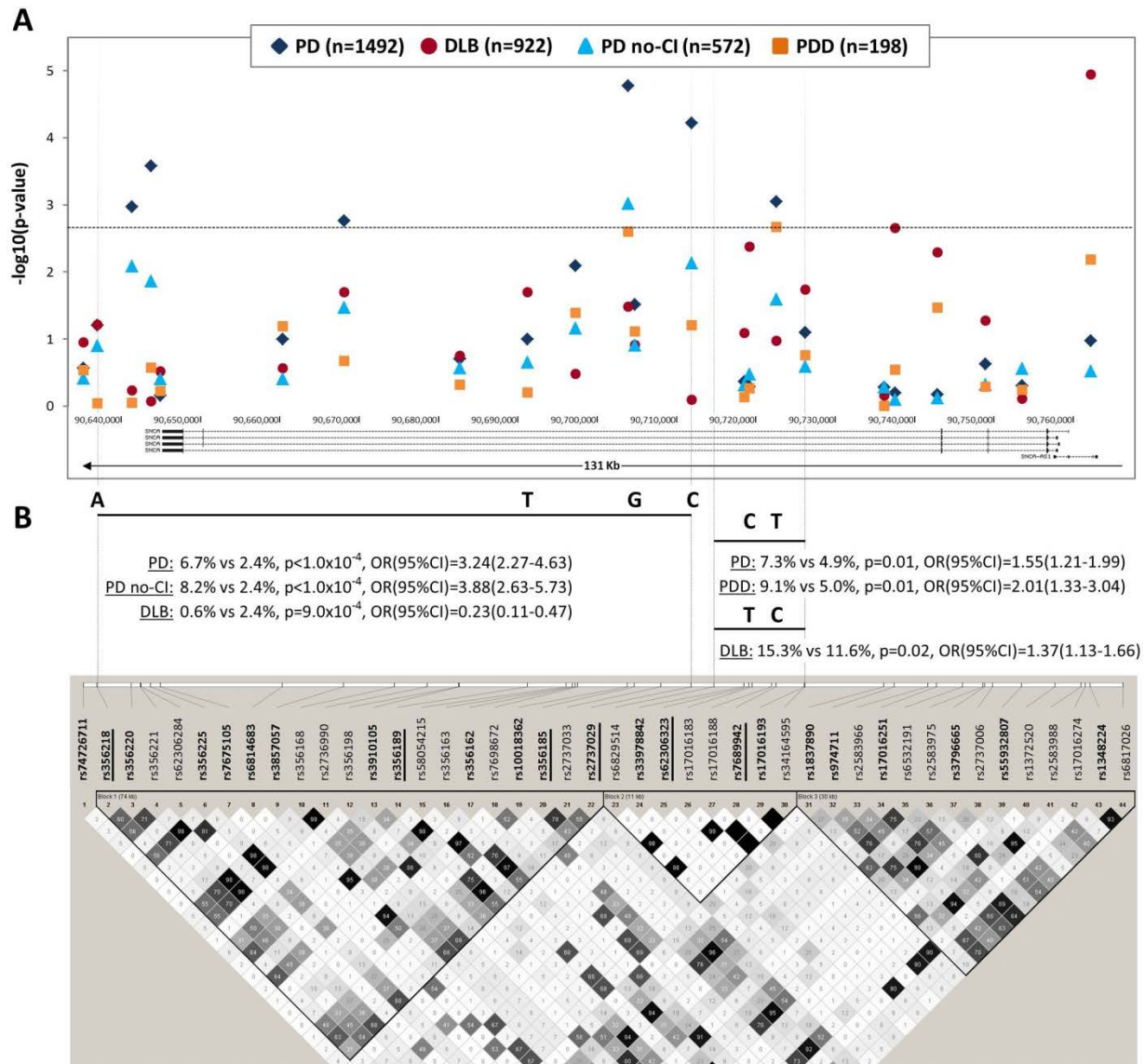


Figure 1: Regional association plot and linkage disequilibrium structure for the *SNCA* locus

A) Logistic regression, adjusting for age, gender, site and APOE dosage, was performed for each group (PD, PD-noCI, PDD, DLB) versus HC (n=978). P values for the 23 tagging SNPs are plotted (as $-\log_{10} P$) against their physical position on chromosome 4 (NCBI Build 37). The locations of known genes in the region are also shown. The black dotted line represents Bonferroni correction threshold of 0.002. **B)** Disease-associate haplotypes are indicated by black lines. Frequencies, ORs and p-values (after 10000 permutations) are also shown. SNPs defining the haplotypes are underlined and the corresponding alleles are indicated with capital letters. LD map based on r^2 values in the associated regions using genotyping results for all the 44 SNPs in HC, as derived by Haploview software (darker shades of black represent greater r^2 values).

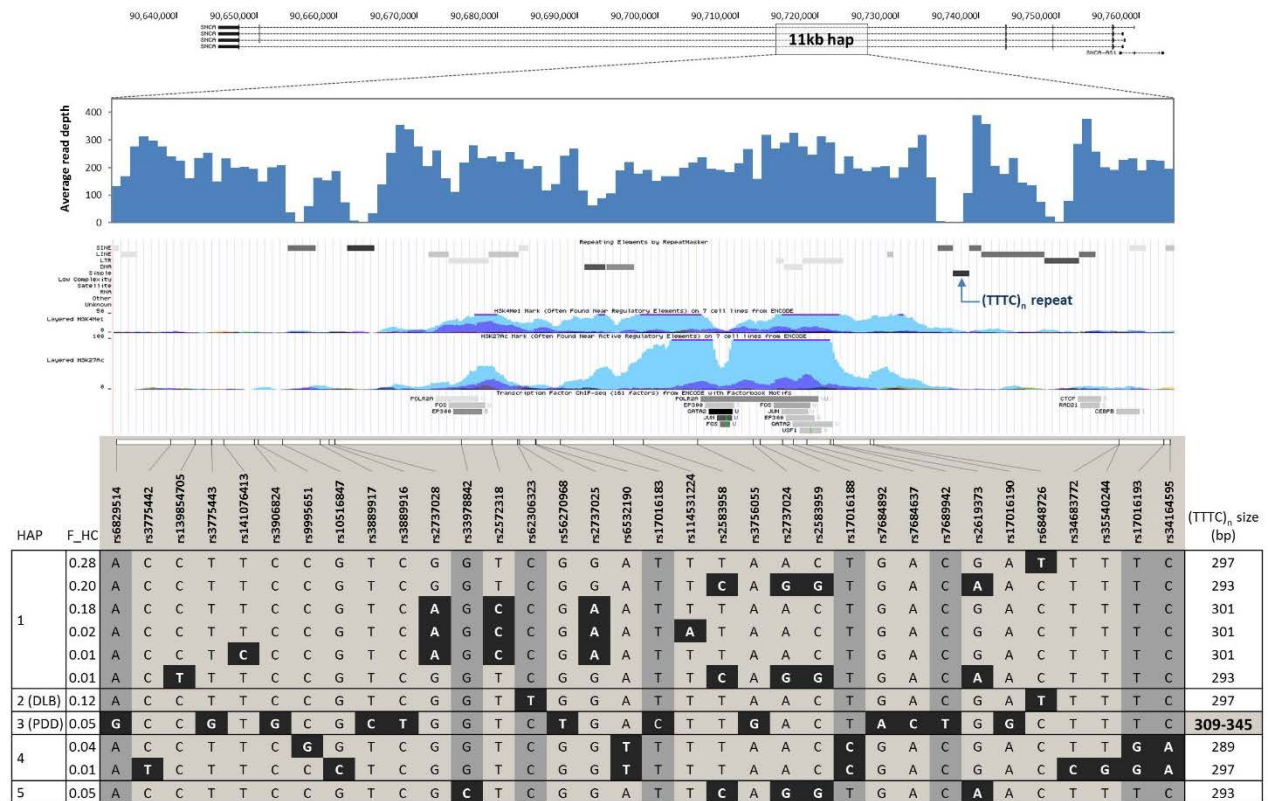


Figure 2: Analysis of the 11kb DLB/PDD associated haplotype

Schematic representation of the *SNCA* gene showing the relative position of the 11kb haplotype (grey box). The average read depth across the interval in 100bp bins is shown. Repeated elements in the region (RepeatMasker), and ENCODE regulatory tracks (including H3K4Me1, H3K27Ac and transcription factor binding sites) are annotated. The position of SNPs with MAF>0.1 is indicated. Haplotypes (frequency>0.1) are displayed and numbered (1-5) accordingly to Supplementary Table 4. The frequency in controls (F_HC) is given next to each haplotype. Alternative alleles are highlighted in black, and SNPs originally included in the 44 SNPs set are shaded in darker grey. The TTTC_n repeat length is also shown for each haplotype, a short repeat expansion (size 309bp to 345bp, in bold and shaded) is associated with the PDD risk haplotype.

Table 1: Cohort description

	PD (n)	DLB (n)	HC (n)
UBC	421	67	556
PPMI ⁵⁶	465		209
Holland: PD-MCI ⁵⁷	110		7
USA: PD-MCI ⁵⁷	75		
New Zealand: PD-MCI ⁵⁷	143		55
Mayo	84	798	91
UK/ICL: PD Brain Bank	194	5	
UK/OPDC: Oxford Brain Bank		52	60
TOTAL	1492	922	978

HC, Healthy controls; PD, Parkinson's disease; DLB, Dementia with Lewy body.

UBC, University of British Columbia; ICL, Imperial College London; OPDC Oxford Parkinson's Disease Centre.

56. Parkinson Progression Marker I. The Parkinson Progression Marker Initiative (PPMI). Progress in neurobiology. 2011 Dec;95(4):629-35.

57. Geurtsen GJ, Hoogland J, Goldman JG, et al. Parkinson's disease mild cognitive impairment: application and validation of the criteria. Journal of Parkinson's disease. 2014;4(2):131-7.

Table 2: Sample Demographics

Diagnosis	n	Neuropath diagnosis	Gender (Male %)	Age/ Age at Death	Age at Onset	Disease duration
PD	1492	314	65.3	71.5 ± 11.3	60.3 ± 10.2	9.6 ± 6.0
• PD-noCI	572	7	60.7	67.4 ± 10.4	58.5 ± 9.6	8.3 ± 4.8
• PDD	198	57	68.2	76.6 ± 7.8	63.2 ± 9.8	13.3 ± 6.8
DLB	922	518	63.7	81.5 ± 9.2	73.2 ± 8.1*	10.8 ± 4.9*
HC	978	115	53.4	72.0 ± 12.6	-	-

mean ± SD; * data only available for 468 subjects

HC, Healthy controls; PD, Parkinson's disease; PD-noCI, PD with no cognitive impairment; PDD, PD dementia; DLB, Dementia with Lewy body.