Investigating the genetic architecture of Dementia with Lewy bodies: a genome-wide association study

Rita Guerreiro, PhD[1, 2, 3]*, Owen A. Ross, PhD[4]*, Celia Kun-Rodrigues, MSc[2], Dena G. Hernandez, PhD[5, 6], Tatiana Orme, BSc[2], John D. Eicher, PhD[7], Claire Shepherd, PhD[8], Laura Parkkinen, PhD[9], Lee Darwent, MSc[2], Michael G. Heckman, MS[10], Sonja W. Scholz, PhD[11], Juan C. Troncoso, MD[12], Olga Pletnikova, MD[12], Olaf Ansorge, MD[9], Jordi Clarimon, PhD[13], Alberto Lleo, PhD[13], Estrella Morenas-Rodriguez, MD[13], Lorraine Clark, PhD[14], Lawrence S. Honig, PhD[14], Karen Marder, MD[14], Afina Lemstra, PhD[15], Ekaterina Rogaeva, PhD[16], Peter St. George-Hyslop, MD[16, 17], Elisabet Londos, MD[18], Henrik Zetterberg, PhD[19], Imelda Barber, PhD[20], Anne Braae, PhD[20], Kristelle Brown, PhD[20], Kevin Morgan, PhD[20], Claire Troakes, PhD[21], Safa Al-Sarraj, FRCPath [21], Tammaryn Lashley, PhD[22], Janice Holton, PhD[22], Yaroslau Compta, PhD[23], Vivianna Van Deerlin, PhD[24], Geidy E. Serrano, PhD[25], Thomas G. Beach[25], Suzanne Lesage, PhD[26], Douglas Galasko, MD[27], Eliezer Masliah, MD[28], Isabel Santana, PhD[29], Pau Pastor, MD[30], Monica Diez-Fairen, BSc[30], Miquel Aguilar, MD[30], Pentti J. Tienari, PhD[31], Liisa Myllykangas, PhD[32], Minna Oinas, PhD[33], Tamas Revesz, PhD[22], Andrew Lees, MD[22], Brad F. Boeve, MD[34], Ronald C. Petersen, PhD[34], Tanis J. Ferman, PhD[35], Valentina Escott-Price, PhD[36], Neill Graff-Radford, MD[37], Nigel J. Cairns, PhD[38], John C. Morris, MD[38], Stuart Pickering-Brown, PhD[39], David Mann, PhD[39], Glenda M. Halliday, PhD[40, 41], John Hardy, PhD[2], John Q. Trojanowski, PhD[24], Dennis W. Dickson, MD[4], Andrew Singleton, PhD[5], David J. Stone, PhD[42], Jose Bras, PhD[1, 2, 3, ♯]

* - Denotes equally contributing authors

♯ - Corresponding author. Email: j.bras@ucl.ac.uk

1 - UK Dementia Research Institute (UK DRI) at UCL, London, UK

2 - Department of Molecular Neuroscience, UCL Institute of Neurology, London, UK

3 - Department of Medical Sciences and Institute of Biomedicine, iBiMED, University of Aveiro, 3810-193 Aveiro, Portugal

4 - Department of Neuroscience, Mayo Clinic, Jacksonville, FL, USA

5 - Laboratory of Neurogenetics, National Institutes on Aging, NIH, Bethesda, MD, USA

6 - German Center for Neurodegenerative Diseases (DZNE)-Tubingen

7 - Genetics and Pharmacogenomics, Merck Research Laboratories, Boston, MA, USA

8 - Neuroscience Research Australia, Sydney, Australia and School of Medical Sciences,

Faculty of Medicine, University of New South Wales, Sydney, Australia
9 - Nuffield Department of Clinical Neurosciences, Oxford Parkinsons Disease Centre, University of Oxford, Oxford, UK
10 - Division of Biomedical Statistics and Informatics, Mayo Clinic, Jacksonville, FL, USA
11 - Neurodegenerative Diseases Research Unit, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, Maryland, USA
12 - Department of Pathology (Neuropathology), Johns Hopkins University School of Medicine, Baltimore, MD, USA
13 - Memory Unit, Department of Neurology, IIB Sant Pau, Hospital de la Santa Creu i Sant Pau, Universitat Autonoma de Barcelona, Barcelona, Spain; Centro de Investigacion Biomedica en Red en Enfermedades Neurodegenerativas (CIBERNED), Instituto de Salud Carlos III, Madrid, Spain
14 - Taub Institute for Alzheimer Disease and the Aging Brain and Department of Pathology and Cell Biology, Columbia University, New York, NY, USA
15 - Department of Neurology and Alzheimer Center, Neuroscience Campus Amsterdam, VU University Medical Center, Amsterdam, the Netherlands
16 - Tanz Centre for Research in Neurodegenerative Diseases and department of Medicine, University of Toronto, Ontario, Canada
17 - Department of Clinical Neurosciences, Cambridge Institute for Medical Research, University of Cambridge, Cambridge, UK
18 - Clinical Memory Research Unit, Institution of Clinical Sciences Malmo, Lund University, Sweden
19 - UK Dementia Research Institute at UCL, London UK, Department of Molecular Neuroscience, UCL Institute of Neurology, London, UK and Clinical Neurochemistry Laboratory, Institute of Neuroscience and Physiology, Sahlgrenska Academy at the University of Gothenburg, Molndal, Sweden
20 - Human Genetics, School of Life Sciences, Queens Medical Centre, University of Nottingham, Nottingham, UK
21 - Department of Basic and Clinical Neuroscience and Institute of Psychiatry, Psychology and Neuroscience, Kings College London, London, UK
22 - Queen Square Brain Bank, Department of Molecular Neuroscience, UCL Institute of Neurology, London, UK
23 - Queen Square Brain Bank, Department of Molecular Neuroscience, UCL Institute of Neurology, London, UK and Movement Disorders Unit, Neurology Service, Clinical

Neuroscience Institute (ICN), Hospital Clinic, University of Barcelona, IDIBAPS, Barcelona, Spain
24 - Department of Pathology and Laboratory Medicine, Center for Neurodegenerative Disease Research, Perelman School of Medicine at the University of Pennsylvania, 3600 Spruce Street, Philadelphia, USA
25 - Banner Sun Health Research Institute, 10515 W Santa Fe Drive, Sun City, AZ 85351, USA
26 - Inserm U1127, CNRS UMR7225, Sorbonne Universites, UPMC Univ Paris 06, UMR and S1127, Institut du Cerveau et de la Moelle epiniere, Paris, France
27 - Department of Neurosciences, University of California, San Diego, La Jolla, CA, United States; Veterans Affairs San Diego Healthcare System, La Jolla, CA, United States
28 - Department of Neurosciences, University of California, San Diego, La Jolla, CA, United States; Department of Pathology, University of California, San Diego, La Jolla, CA, United States
29 - Neurology Service, University of Coimbra Hospital, Coimbra, Portugal
30 - Memory Unit, Department of Neurology, University Hospital Mutua de Terrassa, University of Barcelona, and Fundacio de Docencia I Recerca Mutua de Terrassa, Terrassa, Barcelona, Spain. Centro de Investigacion Biomedica en Red Enfermedades Neurdegenerativas (CIBERNED), Madrid, Spain
31 - Molecular Neurology, Research Programs Unit, University of Helsinki, Department of Neurology, Helsinki University Hospital, Helsinki, Finland
32 - Department of Pathology, Haartman Institute, University of Helsinki and HUSLAB
33 - Department of Neuropathology and Neurosurgery, Helsinki University Hospital and University of Helsinki, Helsinki, Finland
34 – Department of Neurology, Mayo Clinic, Rochester, MN, USA
35 - Department of Psychiatry and Department of Psychology, Mayo Clinic, Jacksonville, FL, USA
36 - MRC Centre for Neuropsychiatric Genetics and Genomics, School of Medicine, Cardiff University, Cardiff, UK
37 - Department of Neurology, Mayo Clinic, Jacksonville, FL, USA
38 - Knight Alzheimers Disease Research Center, Department of Neurology, Washington University School of Medicine, Saint Louis, MO, USA
39 - Institute of Brain, Behaviour and Mental Health, Faculty of Medical and Human Sciences, University of Manchester, Manchester, UK
40 - Neuroscience Research Australia, Sydney, Australia and School of Medical Sciences,

Faculty of Medicine, University of New South Wales,Sydney, Australia
41 - Brain and Mind Centre, Sydney Medical School, The University of Sydney, Sydney, Australia
42 - Genetics and Pharmacogenomics, Merck Research Laboratories, West Point, Pennsylvania, USA

Abstract

Background: Dementia with Lewy Bodies (DLB) is the second most common form of dementia in the elderly but has been overshadowed in the research field, in part due to similarities between DLB, Parkinson's (PD) and Alzheimer's diseases (AD). So far, no large-scale genetic study of DLB has been performed. To better understand the genetic basis of DLB, we have performed a genome-wide association study, with the aim of identifying genetic risk factors for this disorder

Methods: Here we have performed the first genome-wide association study of DLB in a combined cohort of 1,743 DLB patients and 4,454 controls. To reduce genetic heterogeneity, all samples were of European ancestry. All cases were diagnosed according to established criteria for clinical or pathological DLB. In the discovery stage (comprising 1,216 cases and 3,791 controls, the latter were part of two publicly available dbGaP studies (phs000404.v1.p1 and phs000982.v1.p1)) we performed genotyping and exploited the recently established Haplotype Reference Consortium panel as the basis for imputation. Association analysis was performed in all cases as well as in only those with pathological diagnosis (974 cases). In the replication stage (comprising 527 DLB cases and 663 controls) we performed genotyping of significant and suggestive results. Lastly, we conducted a meta-analysis of both stages. Genotyping was undertaken at three locations: UCL, NIH and the Mayo Clinic.

Findings: Results confirm previously reported associations: *APOE* (rs429358; OR=2.4 [95%CI 2.14-2.70]; $p=1.05x10^{-48}$), *SNCA* (rs7681440; OR=0.73 [0.66- 0.81]; $p=6.39x10^{-10}$) and *GBA* (rs35749011; OR=2.55 [1.88-3.46]; $p=1.78x10^{-09}$). They also provide novel candidate loci, namely *CNTN1* (rs7314908; OR=1.53 [1.29-1.81]; $p=1.02x10^{-06}$); further replication of findings at these two loci will be important. Additionally, we estimate the heritable component of DLB to be approximately 36%.

Interpretation: Despite the relatively modest sample size for a GWAS, and acknowledging the potential biases from ascertaining samples from multiple locations, we present the most comprehensive and well-powered genetic study in DLB to date. These data unequivocally show that common genetic variability plays a role in this disease.

Funding

Funded by the Alzheimer's Society, and the Lewy Body Society.

Introduction

Dementia with Lewy Bodies (DLB) is the second most common form of dementia following Alzheimer's disease (AD) [1]. Despite this fact, very little attention has been devoted to understanding the pathogenesis of this disorder, particularly when compared with the other common neurodegenerative diseases such as AD and Parkinson's disease (PD).

So far, the only fully penetrant genetic variability that has been identified and replicated as a specific cause of DLB are *SNCA* point mutations and gene dosage. Three major factors may have contributed to this low number of causative mutations: first, DLB, often a disease of old age, is not commonly seen in multiplex kindreds, meaning that successful linkage studies have been rare [2]; second, the accurate clinical diagnosis of DLB is complex, with a relatively high rate of misdiagnosis [3]; and third, because even the largest cohorts of DLB samples have been generally small, in many instances including as little as 100 patients [4,5]. However, it is currently indisputable that DLB has a strong genetic component. The epsilon-4 allele of *APOE* [6,7] is recognized to be a strong risk factor, as are heterozygous mutations and common polymorphisms in the glucocerebrosidase gene (*GBA*)[8]. Both of these results have stemmed from candidate gene association studies; it was known that *APOE* was strongly associated with AD and *GBA* was a strong risk factor for PD/Lewy body disorders. In addition to these genetic associations with susceptibility, we have recently provided evidence that DLB has a heritable component [9].

It has been shown that there is no overlap in common genetic risk between PD and AD [10], a fact that is not entirely surprising given the differences in phenotype. However, it is reasonable to hypothesize that the overlaps and differences in clinical and pathological presentation between DLB with both PD and AD stem, at least in part, from aspects in their underlying genetic architecture and, consequently, disease pathobiology. Specific genes/loci associated with disease as well as strength of association are factors that can be expected to modulate these phenotypic overlaps and differences. However, despite these encouraging findings, large-scale unbiased genetic studies in DLB have not yet been performed, which is likely due to the difficulty in identifying large, homogeneous cohorts of cases.

To address the need for more powerful and comprehensive genetic studies of DLB, we performed the first large-scale genome-wide association study in this disease, using a total of 1,743 cases and 4,454 controls. The majority of cases were neuropathologically assessed, providing a greater level of diagnostic detail. Controls used were derived from two publicly available datasets and from the Mayo Clinic Florida control database. We performed imputation using the most recent imputation panel provided by the Haplotype Reference Consortium

enabling us to have a detailed overview of common and intermediate frequency genetic variability.

Methods

Participants

All case subjects were Caucasian of European ancestry and diagnosed according to the consensus criteria for either clinical or pathological diagnosis of DLB [11]. The majority of cases were pathologically diagnosed and these were included only when the likelihood of a diagnosis of DLB was "Intermediate" or "High" [11]. Caucasian control subjects in the discovery stage are part of the "General Research Use" controls from the two studies publicly available at dbGaP (The Genetic Architecture of Smoking and Smoking Cessation (phs000404.v1.p1) and Genetic Analysis of Psoriasis and Psoriatic Arthritis (phs000982.v1.p1)) (Table 1). Replication stage Caucasian controls are from the Mayo Clinic Florida control database (Table 2). Investigators at every site obtained written informed consent from patients and control individuals as well as approval from a local ethics committee.

Discovery stage genotyping and quality control

Case subjects were genotyped in either Illumina Omni2.5M or Illumina OmniExpress genotyping arrays (n=987 and n=700, respectively). Controls were genotyped in either Illumina Omni2.5M or Illumina Omni1M arrays (n=1,523 and n=2,847 respectively). Autosomal variants with GenTrain scores >0.7 were included in the QC stage. We removed SNPs with a call rate <95%, HWE p-value in controls $<1\times10^{-7}$, or a minor allele frequency (MAF) <0.01. Samples were removed if they had substantial non-European admixture, were duplicates or first- or second-degree relatives of other samples, had a genotype call rate <98% or had substantial cryptic relatedness scores (PI_HAT >0.1).

Population outliers were determined by principal components analysis (PCA), using SNPs passing the aforementioned quality-control filters. After linkage disequilibrium (LD)-based pruning with version 1.9 of PLINK [12] to quasi-independence (variance inflation factor =2), 130,715 SNPs remained in the dataset. Genotypes for these SNPs were combined with 1000Genomes phase 3 genotypes for samples from the YRI, CEU, JPT, and CHB reference populations, and subjected to PCA. Individuals lying farther than ¼ of the distance between CEU and JPT/CHB/YRI when plotted on the first two PCA axes were considered to have substantial non-European admixture and were excluded (Supplementary Figure 1).

Imputation

Since samples were genotyped in a variety of arrays, we selected only variants that intersected between all arrays to be included in the imputation stage. We performed imputation using the most recent reference panels provided by the Haplotype Reference Consortium (HRC v1.1 2016). Eagle v2.3 was used to pre-phase haplotypes based on genotype data[13,14]. Imputation was conducted using the Michigan Imputation Server[15]. Following imputation, variants passing a standard imputation quality threshold (R^2 >= 0.3) were kept for further analysis.

Statistical Analysis of discovery stage

We used logistic regression as implemented in PLINK1.9[12] to test for association of hard-call variants with the binary case-control phenotype using gender as a covariate. Variants were examined under an additive model (i.e. effect of each minor allele) and odds ratios (ORs) and 95% confidence intervals (CIs) were estimated. To control for population stratification, we used coordinates from the top six PC dimensions as additional covariates in the logistic regression models. We utilized QQ plots and the genomic inflation factor (λ) to test for residual effects of population stratification not fully controlled for by the inclusion of PCA and cohort covariates in the regression model. Additionally, we have performed a sub-analysis in the discovery stage including only pathologically diagnosed cases.

Moreover, to take into account the uncertainty of imputation, we have performed the same association in PLINK1.9 using dosage data. Results are identical to the best-guess calls and are present in Supplementary Table 3.

Gene-wise burden tests were performed using all variants with an effect in protein sequence and a maximum minor allele frequency of 5%, using SKAT-O [16,17] as implemented in EPACTS [18]. We used the top six principal components and gender as covariates in the burden test.

Replication genotyping

A total of 527 DLB cases and 663 controls from the Mayo Clinic were included in the replication stage (Table 2). Replication was attempted for top variants showing a p-value in discovery of less than $1x10^{-6}$. A total of 32 signals were tested for replication using a Sequenom MassARRAY iPLEX SNP panel (Supplementary Table 1). Power calculations for replication sample size selection were performed using the R package 'RPower'. An average statistical

power of 81% was estimated for the 32 signals, based on sample size, variant frequency and effect size in the discovery stage and a replication p-value threshold of 0.05. Association in replication was tested using logistic regression models adjusted for age (age at onset for the clinically diagnosed DLB patients, age at death for the high likelihood DLB patients, and age at study for controls) and gender.

A combined analysis of stage 1 and 2 was conducted with GWAMA[19] under a fixed-effects model, using estimates of the allelic odds ratio and 95% confidence intervals.

Phenotypic variance explained

To estimate the phenotypic variance explained by the genotyped SNPs in this cohort we used GREML analysis as implemented in GCTA [20,21]. We used the first ten principal components as covariates and a disease prevalence of 0.1% [22]. We have also estimated the partitioned heritability by chromosome, where a separate genetic relationship matrix was generated for each chromosome. Each matrix was then run in a separate REML analysis. Linear regression was applied to determine the relationship between heritability and chromosome length.

Role of the funding source

The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Results

Single variant analysis

Application of quality control filters to the dataset yielded high-quality genotypes at 448,155 SNPs for 1,216 cases and 3,791 controls. A total of 52 cases were excluded for cryptic relatedness, 20 cases for genetic ancestry, and the remaining 399 for low call rates/poor genotyping. After imputation and quality control, genotypes for 8,397,716 variants were available for downstream analyses. QQ plot and genomic inflation factor (λ=1.01) indicated good control of population stratification (Supplementary Figure 2).

Five regions were associated with DLB risk at genome-wide significance ($p<5x10^{-8}$) in the discovery stage (Figure 1; Table 3).

These included the previously described AD and PD loci *APOE* (rs429358; OR=2.40 [95%CI 2.14-2.70]; $p=1.05x10^{-48}$), *SNCA* (rs7681440; OR=0.73 [0.66- 0.81]; $p=6.39x10^{-10}$) and

GBA (rs35749011; OR=2.55 [1.88-3.46]; p=1.78x10^{-09}). Additionally, loci overlapping *BCL7C/STX1B* (rs897984; OR=0.74 [0.67-0.82]; p=3.30x10^{-09}) and *GABRB3* (rs1426210; OR=1.32 [1.21-1.48]; p=2.62x10^{-08}) were also genome-wide significant.

A sub-analysis including only pathologically diagnosed cases revealed that all but *GABRB3* maintained their genome-wide significance in that smaller dataset (Table 3; Supplementary Figure 6).

The replication stage of the GWAS design provided independent replication (p<0.05) for 3 of the loci (*APOE*, *SNCA*, and *GBA*), all of which were also genome-wide significant in the combined analysis of both stages (Table 3; Supplementary Table 1). Additionally, suggestive evidence of an association (p<1x10^{-6}) with DLB was observed for two loci (*SOX17* and *CNTN1*) in the discovery stage. The association at *SOX17* did not replicate (Supplementary Table 1). For *CNTN1*, the association with DLB improved slightly when performing the sub-analysis on the pathologically confirmed cases (rs74461734; OR=1.58 [1.32-1.88]; p=4.32x10^{-07}), and this candidate locus shows evidence of replication with very similar effect size (OR=1.54 [1.32-1.79]; p=0.03). Further replication will be important given the lack of a genome-wide significant association in the discovery stage, however this association appears promising.

A systematic assessment of genetic loci previously associated with AD or PD showed no evidence of other genome-wide significant associations in this DLB cohort (Supplementary Figures 7 to 66). These include the *TREM2* locus, where the p.Arg47His variant has been shown to have a strong effect in AD [23]. In our cohort this variant did not show genome-wide significant levels of association (OR=3.46 [1.54-7.77]; p=0.002), despite the overrepresentation in cases. Similarly, *MAPT*, which is strongly associated with PD and has been previously linked to DLB [24], shows no strong evidence of association in this study (rs17649553; OR=0.86 [0.76-0.96]; p=0.0126).

Gene burden analysis

Gene based burden analysis of all low frequency (MAF < 0.05) and rare variants changing the amino acid sequence, showed a single genome-wide significant result comprised of 6 variants at *GBA* (p.Asn409Ser, p.Thr408Met, p.Glu365Lys, p.Arg301His, p.Ile20Val and p.Lys13Arg), (p=1.29x10^{-13}). No other gene showed evidence of strong association with disease or overlap with single variant analysis (Table 4).

Estimation of heritability of DLB

Using the first ten principal components as covariates and a disease prevalence of 0.1%,

estimation of the phenotypic variance attributed to genetic variants showed a heritable component of DLB of 36% (± 0.03). Results for the chromosome-partitioned heritability are presented in Figure 3. As expected for a common complex disease, we found a strong correlation between chromosome length and heritability ($p = 6.88x10^{-5}$).

Interestingly, the heritability for DLB at chromosome 19 is much higher than what would be expected given chromosome size and likely reflects the role of *APOE*. It should also be noted that chromosomes 5, 6, 7 and 13 all have higher heritability for DLB than expected, while none of them have variants with genome-wide significant results.

Discussion

This is the first comprehensive, unbiased study of common and intermediate frequency genetic variability in DLB. We identified five genome-wide significant associations in the discovery stage (*APOE*, *BCL7C*/*STX1B, SNCA, GBA*, and, *GABRB3*), with the associations regarding *APOE*, *SNCA*, and *GBA* being confirmed in the replication stage.

The most significant association signal is observed at the *APOE* locus (*APOE E4*) which has been previously shown to be highly associated with DLB [6,7]. As described *APOE E4* is the major genetic risk locus for AD and has been implicated in cognitive impairment within PD although not with PD risk per se. It has also been observed to affect the levels of both β-amyloid and Lewy body pathology in brains of patients [25], and in a small Finnish dataset the E4 association with DLB was largely driven by the subgroup with concomitant AD pathology [26].

The second strongest association is observed at the *SNCA* locus. To examine whether this association is independent of that observed in PD, we conditioned our analysis on the top PD variant (rs356182), which showed only a negligible effect on the DLB association (conditioned OR=0.70 [0.63-0.78]; $p= 2.89x10^{-10}$), confirming the different association profile between DLB and PD that we had previously reported [7]. *SNC*A is the most significant common genetic risk factor for PD, with rs356182 having a meta-analysis p-value of $1.85x10^{-82}$ (OR:1.34 [1.30-1.38]) in PDGene. This variant is located 3' to the gene [27], while in DLB no association can be found in that region (Figure 4). Additionally, the most associated DLB SNP for the *SNCA* locus (rs7681440) has a PD meta-analysis p-value>0.05 in PDGene. Interestingly, when performing a conditional analysis on the top PD SNP (rs356182), Nalls and colleagues reported an independent association at the 5' region of the gene (rs7681154), and this variant is in strong LD (r^2=0.91) with the rs7681440 *SNCA* variant identified in our study. It is tempting to speculate that these differences may reflect pathobiological differences between the two diseases, perhaps mediated by differential regulation of gene expression.

To gain insight into potential regulatory effects of this distinct SNCA signal, we used eQTL data from GTEx and the Harvard Brain Bank Resource Center to determine whether rs7681154 and rs7681440 influence gene expression as eQTLs. In the GTEx data, the most associated SNP in DLB is a strong eQTL in the cerebellum for *RP11-67M1.1*, a known antisense gene located at the 5'-end of *SNCA*, with the alternative allele showing a reduction in expression of *RP11-67M1.1* (Figure 2a). These results are compatible with a model in which rs7681440 genotypes influence the expression levels of *SNCA* indirectly through the action of *RP11-67M1.1*. More specifically, the alternative allele associates with a lower expression of *RP11-67M1.1* and consequently less repression of *SNCA* transcription (higher *SNCA* expression), which is in accordance with a higher frequency of the alternative allele in cases when compared to controls. Additionally, rs7681154 was associated with *SNCA* expression in cerebellum using the Harvard Brain Bank Resource Center results ($p=2.87x10^{-11}$) (Figure 2b), with the alternate allele associated with increased *SNCA* expression. Such a relationship between this locus and *SNCA* expression is supported by the high expression of SNCA in brain and the association of rs7681440 with increased *SNCA* expression in whole blood ($p=2.13x10^{-38}$) [28,29]. However, further investigation of the identified significant eQTLs is needed as the effect was observed for only one brain region. This could plausibly result from low overall expression of *RP11-67M1.1* and higher cerebellum RNA quality when compared to other assayed brain regions in these datasets. Nonetheless, it is interesting to note that both eQTLs' effects fit with a model of increased *SNCA* expression in cases compared to controls.

The top hit at the *GBA* locus (rs35682329) is located 85,781bp downstream of the gene and is in high LD (D': 0.9; R^2: 0.8) with p.Glu365Lys (also reported in the literature as E365K, E326K, rs2230288), which has been suggested as a risk factor for DLB [8]. The top associated variant for PD at this locus is the rs71628662 (PDGene meta-analysis OR:0.52 [0.46-0.58] and p-value $6.86x10^{-28}$). This variant is also in high LD with the top SNP identified here (D': 0.9 and R^2:0.8). Interestingly in this study we show that *APOE* and *GBA* have similar effect sizes in DLB (ORs of 2.5 and 2.2, respectively). Gene burden based analysis showed *GBA* as the only genome-wide significant association with DLB risk. The inexistence of other associations should be interpreted with some caution. As we are not ascertaining the complete spectrum of genetic variability, it is possible that other genes will have a significant burden of genetic variants that were simply not captured in our study design, despite using the most recent imputation panel.

Although in our meta-analysis we observed a genome-wide significant association with DLB at the *BCL7C/STX1B* locus, as this was mostly driven by the discovery stage data (replication stage results were OR=0.98, P=0.83), further replication is needed. That being

acknowledged, an association at the *BCL7C/STX1B* locus has been previously reported for PD [27,30]. The top PD-associated variants at this locus were rs14235 (synonymous) and rs4889603 (intronic), located at *BCKDK* and *SETD1A*, respectively. The top SNP identified in DLB at this locus (rs897984) shows the same direction of association seen in PD (OR=0.93, 95%CI:0.90-0.96), a PD meta-analysis p-value of $1.34x10^{-5}$ (data from PDgene) and strong LD with both PD hits (r-squared 0.28-0.32; correlation p-values<0.0001). This is a gene-rich region of the genome (Figure 5) making it difficult to accurately nominate the gene driving the association. Mining data from the GTEx project showed that rs897984 is not an eQTL for any gene in the locus. Nonetheless, in both PD studies, the nominated gene at the locus was *STX1B* likely due to its function as a synaptic receptor [31]. In addition, *STX1B* has a distinctive pattern of expression across tissues, presenting the highest expression in the brain. In this tissue, when compared to the closest genes in the locus (*HSD3B7*, *BCL7C*, *ZNF668*, *MIR4519, CTF1*, *FBXL19*, *ORAI3*, *SETD1A*, *STX4*), *STX1B* also shows the highest levels of expression (Supplementary Figure 4). Mutations in *STX1B* have recently been shown to cause fever associated epileptic syndromes [32] and myoclonic astatic epilepsy [33].

Although not quite genome-wide significant in the discovery stage, the association between *CNTN1* and DLB risk replicated with a very similar association OR as the discovery stage. Interestingly, the locus has been previously associated with PD in a genome-wide study of IBD segments in an Ashkenazi cohort [34], and with cerebral amyloid deposition, assessed with PET imaging in *APOE E4* non-carriers [35]. This locus was also shown to be sub-significantly associated with clinico-pathologic AD dementia [36]. The Contactin 1 protein encoded by *CNTN1* is a glycosylphosphatidylinositol (GPI)-anchored neuronal membrane protein that functions as a cell adhesion molecule with important roles in axonal function [37,38]. Mutations in *CNTN1* were found to cause a familial form of lethal congenital myopathy [39]. Contactin 1 drives Notch signalling activation and modulates neuroinflammation events, possibly participating in the pathogenesis of Multiple Sclerosis and other inflammatory disorders [40]. A functional protein association network analysis of *CNTN1* using STRING shows it is in the same network as *PSEN2* (Supplementary Figure 5), supporting its potential role in neurodegeneration. It is also worth noting that *LRRK2* is located less than 500kb away from the most associated SNP at this locus, which could suggest that the association might be driven by variation at the *LRRK2* locus. We assessed LD across the region and that analysis revealed that rs79329964 is in equilibrium with both p.Gly2019Ser (R^2: 0.000043) as well as with the PD hit at this locus rs76904798 (R^2: 0.003), suggesting it to be an independent association from the PD risk. Although samples were not screened for p.Gly2019Ser directly, the variant was well imputed (R^2=0.94). The exclusion

of all samples that carried the p.Gly2019Ser variant showed no significant effect on the association at the *CNTN1* locus. It is worth noting that the p.Gly2019Ser variant showed a higher minor allele frequency in cases when compared to controls (0.0021 and 0.0003 respectively). Further validation of the involvement of *CNTN1* variation in modifying risk of DLB will be important.

In addition to performing a GWAS with clinico-pathologic AD dementia, Beecham and colleagues [36] also analysed commonly comorbid neuropathologic features observed in older individuals with dementia, including Lewy body disease (LBD). In this latter analysis, only the *APOE* locus was found to achieve genome-wide significance. However, when testing known common AD risk variants with coincident neuropathologic features, the authors identified hits at *SORL1* and *MEF2C* as nominally associated. In our cohort of DLB cases we found no genome-wide significant associations between these variants and disease. Similarly, we had previously reported an association at the *SCARB2* locus with DLB [7]. In the larger dataset studied herein, the association remained at the suggestive level and did not reach genome-wide significance (top SNP in the current study rs13141895: p-value=9.58×10^{-4}). No other variant previously reported to be significantly associated with AD or PD in recent GWAS meta-analyses showed a genome-wide significant association with DLB. The top AD or PD variants at the following loci showed nominal ($p<0.05$) association levels: *MAPT*, *BIN1*, *GAK*, *HLA-DBQ*, *CD2AP*, *INPP5D*, *ECHDC3* and *SCIMP*. Additionally, variants previously suggested to be associated with Lewy-related pathology in a Finnish cohort, did not show evidence of association in this study (Supplementary Table 2).

There are noted limitations to this study. The control population, because it was derived from publicly available data, is not perfectly matched to the case cohort. To address this, we have used all available information (both clinical and genetic) to create a cohort that is as comparable as possible. Additionally, despite using the same diagnostic criteria for all included cases, these were collected in a variety of locations, suggesting that diagnostic accuracy might have been variable with “contamination” from PD or AD cases. Of note, we do not see an overrepresentation of genetic risk factors from those diseases in our results (*MAPT*, *CLU* or *CR1* for example) suggesting minimal inclusion. Similarly, since samples were collected in various countries, population stratification could bias the results. Here, we have used standard methodology to correct for any such bias and, as a consequence, our results show no evidence of population stratification as evidenced by the QQ plot as defined by the acquired unbiased genotype data. Additionally, cases were genotyped at three locations and controls were all

derived from publicly available datasets, using a mixture of genotyping arrays, which could provide a source of genotyping bias. However, our approach was to select variants that were at the intersection of all used arrays prior to imputation, which makes use, effectively, of the same genotyping probes for all samples. This approach has been shown to remove any bias from this type of results and any effects of using different array scanners is negligible for high quality variants [41].

This is the first large-scale genome-wide association study performed in DLB. We estimate the heritability of DLB to be approximately 36%, which is similar to what is known to occur in PD [42]. This shows that, despite not having multiple causative genes identified so far, genetics plays a relevant role in the common forms of DLB. Additionally, we provide evidence suggesting that novel DLB loci are likely to be found at chromosomes 5, 6, 7 and 13 given the high heritability estimates at these chromosomes. A significant majority of our case cohort in the current study was comprised of cases with neuropathological diagnoses, which provides a greater level of information for diagnostic accuracy. These results provide us with the first glimpse into the molecular pathogenesis of DLB; they reveal that this disorder has a strong genetic component and suggest a unique genetic risk profile. From a molecular perspective, DLB does not simply sit between PD and AD; instead, the combination of risk alleles is unique, with loci that are established risk factors for those diseases having no clear role in DLB (e.g. *MCCC1*, *STK39*, *CLU*, *CR1* or *PICALM*). Further increases in the size of DLB cohorts will likely reveal additional common genetic risk loci, and these will, in turn, improve our understanding of this disease, its commonalities and differences with other neurodegenerative conditions, ultimately allowing us to identify disease-specific targets for future therapeutic approaches.

Contributors

JB, RG, JH, DJS and AS designed the study. JB, AS, DS, and OAR obtained funding for the study. JB, RG, OAR, CKR, LD, SWS and DH performed data acquisition. JB, RG, OAR, and CKR analysed and interpreted the data. CS, LP, SWS, OA, JC, LC, LH, KM, AL, EL, ER, PGH, EL, HZ, IB, AB, KB, KM, DB, CT, SAS, TL, JH, YC, VVD, JQT, GES, TGB, SL, DG, EM, IS, PP, PJT, LM, MO, TR, AJL, BFB, RCP, TJF, VEP, NGR, NC, JCM, DS, SPB, DM, DWD, GH collected and characterised samples. JB, RG, OAR, CKR, and TO wrote the first draft of the paper. All other co-authors participated in preparation of the paper by reading and commenting on drafts before submission.

Declaration of interests

We declare that we have no conflicts of interest.

Acknowledgments

The authors would like to thank Professor Ian McKeith for his continued support and encouragement. This work was supported in part by the National Institutes of Neurological Disease and Stroke. Jose Bras and Rita Guerreiro's work is funded by research fellowships from the Alzheimer's Society. Tatiana Orme is supported by a scholarship from the Lewy Body Society. For the neuropathologically confirmed samples from Australia, tissues were received from the Sydney Brain Bank, which is supported by Neuroscience Research Australia and the University of New South Wales, and Dr Halliday is funded by an NHMRC senior principal research fellowship. We would like to thank the South West Dementia Brain Bank (SWDBB) for providing brain tissue for this study. The SWDBB is supported by BRACE (Bristol Research into Alzheimer's and Care of the Elderly), Brains for Dementia Research and the Medical Research Council. 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. The brain samples and/or bio samples were obtained from The Netherlands Brain Bank, Netherlands Institute for Neuroscience, Amsterdam (open access: www.brainbank.nl). All Material has been collected from donors for or from whom a written informed consent for a brain autopsy and the use of the material and clinical information for research purposes had been obtained by the NBB. This study was also partially funded by the Wellcome Trust, Medical Research Council and Canadian Institutes of Health Research (Dr. St. George-Hyslop). Work from Dr. Compta was supported by the CERCA Programme / Generalitat de Catalunya, Barcelona, Catalonia, Spain. The Nottingham Genetics Group is supported by ARUK and The Big Lottery Fund. The effort from Columbia University was supported by the Taub Institute, the Panasci Fund, the Parkinson's Disease Foundation, and NIH grants NS060113 (Dr Clark), P50AG008702 (P.I. Scott Small), P50NS038370 (P.I.R. Burke), and UL1TR000040 (P.I.H. Ginsberg). Dr Ross is supported by the Michael J. Fox Foundation for Parkinson's Research, NINDS R01# NS078086. The Mayo Clinic Jacksonville is a Morris K. Udall Parkinson's Disease Research Center of Excellence (NINDS P50 #NS072187) and is supported by The Little Family Foundation and by the Mangurian Foundation Program for Lewy Body Dementia research and the Alzheimer Disease Research Center (P50 AG016547). The work from the Mayo Clinic Rochester is supported by the National Institute on Aging (P50 AG016574 and U01 AG006786). This work has received support from The Queen Square Brain Bank at the UCL Institute of Neurology; where Dr Lashley is funded by

an ARUK senior fellowship. Some of the tissue samples studied were provided by the MRC London Neurodegenerative Diseases Brain Bank and the Brains for Dementia Research project (funded by Alzheimer's Society and ARUK). This research was supported in part by both the NIHR UCLH Biomedical Research Centre and the Queen Square Dementia Biomedical Research Unit. This work was supported in part by the Intramural Research Program of the National Institute on Aging, National Institutes of Health, Department of Health and Human Services; project AG000951-12. The University of Pennsylvania case collection is funded by the Penn Alzheimer's Disease Core Center (AG10124) and the Penn Morris K. Udall Parkinson's Disease Research Center (NS053488). Tissue samples from UCSD are supported by NIH grant AG05131. The authors thank the brain bank GIE NeuroCEB, the French program “Investissements d'avenir” (ANR-10-IAIHU-06). Dr Tienari and Dr Myllykangas are supported by the Helsinki University Central Hospital, the Folkhälsan Research Foundation and the Finnish Academy. This work was in part supported by the Canadian Consortium on Neurodegeneration in Aging (ER). The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. The data used for the analyses described in this manuscript were obtained from the GTEx Portal on 04/01/17. The authors acknowledge the contribution of data from Genetic Architecture of Smoking and Smoking Cessation accessed through dbGAP. Funding support for genotyping, which was performed at the Center for Inherited Disease Research (CIDR), was provided by 1 X01 HG005274-01. CIDR is fully funded through a federal contract from the National Institutes of Health to The Johns Hopkins University, contract number HHSN268200782096C. Assistance with genotype cleaning, as well as with general study coordination, was provided by the Gene Environment Association Studies (GENEVA) Coordinating Center (U01 HG004446). Funding support for collection of datasets and samples was provided by the Collaborative Genetic Study of Nicotine Dependence (COGEND; P01 CA089392) and the University of Wisconsin Transdisciplinary Tobacco Use Research Center (P50 DA019706, P50 CA084724). The data used for the analyses described in this paper were obtained from the database of Genotypes and Phenotypes (dbGaP), at http://www.ncbi.nlm.nih.gov/gap. Genotype and phenotype data for the Genetic Analysis of Psoriasis and Psoriatic Arthritis study were provided by Dr. James T. Elder, University of Michigan, with collaborators Dr. Dafna Gladman, University of Toronto and Dr. Proton Rahman, Memorial University of Newfoundland, providing samples. This work was supported in part by the Intramural Research Program of the National Institutes of Health (National Institute of Neurological Disorders and Stroke; project ZIA NS003154). Tissue samples for genotyping were

provided by the Johns Hopkins Morris K. Udall Center of Excellence for Parkinson's Disease Research (NIH P50 NS38377) and the Johns Hopkins Alzheimer Disease Research Center (NIH P50 AG05146). This study was supported by grants from the National Institutes of Health, the Canadian Institute for Health Research, and the Krembil Foundation. Additional support was provided by the Babcock Memorial Trust and by the Barbara and Neal Henschel Charitable Foundation. JTE is supported by the Ann Arbor Veterans Affairs Hospital. The authors would like to thank the Genome Aggregation Database (gnomAD) and the groups that provided exome and genome variant data to this resource. A full list of contributing groups can be found at http://gnomad.broadinstitute.org/about.

References

1 Rahkonen T, Eloniemi-Sulkava U, Rissanen S, Vatanen A, Viramo P, Sulkava R. Dementia with Lewy bodies according to the consensus criteria in a general population aged 75 years or older. *J Neurol Neurosurg Psychiatry* 2003; **74**: 720–4.

2 Bogaerts V, Engelborghs S, Kumar-Singh S, *et al.* A novel locus for dementia with Lewy bodies: a clinically and genetically heterogeneous disorder. *Brain* 2007; **130**: 2277–91.

3 Walker Z, Possin KL, Boeve BF, Aarsland D. Lewy body dementias. *Lancet* 2015; **386**: 1683–97.

4 Keogh MJ, Kurzawa-Akanbi M, Griffin H, *et al.* Exome sequencing in dementia with Lewy bodies. *Transl Psychiatry* 2016; **6**: e728.

5 Geiger JT, Ding J, Crain B, *et al.* Next-generation sequencing reveals substantial genetic contribution to dementia with Lewy bodies. *Neurobiol Dis* 2016; **94**: 55–62.

6 Tsuang D, Leverenz JB, Lopez OL, *et al.* APOE ε4 increases risk for dementia in pure synucleinopathies. *JAMA Neurol* 2013; **70**: 223–8.

7 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; **23**: 6139–46.

8 Nalls MA, Duran R, Lopez G, *et al.* A multicenter study of glucocerebrosidase mutations in dementia with Lewy bodies. *JAMA Neurol* 2013; **70**: 727–35.

9 Guerreiro R, Escott-Price V, Darwent L, *et al.* Genome-wide analysis of genetic correlation in dementia with Lewy bodies, Parkinson's and Alzheimer's diseases. *Neurobiol Aging* 2016; **38**: 214.e7–10.

10 Moskvina V, Harold D, Russo G, *et al.* Analysis of genome-wide association studies of Alzheimer disease and of Parkinson disease to determine if these 2 diseases share a common genetic risk. *JAMA Neurol* 2013; **70**: 1268–76.

11 McKeith IG, Dickson DW, Lowe J, *et al.* Diagnosis and management of dementia with Lewy bodies: third report of the DLB Consortium. *Neurology* 2005; **65**: 1863–72.

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

13 McCarthy S, Das S, Kretzschmar W, *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet* 2016; **48**: 1279–83.

14 Loh P-R, Danecek P, Palamara PF, *et al.* Reference-based phasing using the Haplotype Reference Consortium panel. *Nat Genet* 2016; **48**: 1443–8.

15 Das S, Forer L, Schönherr S, *et al.* Next-generation genotype imputation service and methods. *Nat Genet* 2016; **48**: 1284–7.

16 Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X. Rare-variant association testing for sequencing data with the sequence kernel association test. *Am J Hum Genet* 2011; **89**: 82–93.

17 Lee S, Emond MJ, Bamshad MJ, *et al.* Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. *Am J Hum Genet* 2012; **91**: 224–37.

18 Kang HM. EPACTS: efficient and parallelizable association container toolbox. 2014.

19 Mägi R, Morris AP. GWAMA: software for genome-wide association meta-analysis. *BMC Bioinformatics* 2010; **11**: 288.

20 Yang J, Benyamin B, McEvoy BP, *et al.* Common SNPs explain a large proportion of the heritability for human height. *Nat Genet* 2010; **42**: 565–9.

21 Lee SH, Wray NR, Goddard ME, Visscher PM. Estimating missing heritability for disease from genome-wide association studies. *Am J Hum Genet* 2011; **88**: 294–305.

22 Zaccai J, McCracken C, Brayne C. A systematic review of prevalence and incidence studies of dementia with Lewy bodies. *Age Ageing* 2005; **34**: 561–6.

23 Guerreiro R, Wojtas A, Bras J, *et al.* TREM2 variants in Alzheimer's disease. *N Engl J Med* 2013; **368**: 117–27.

24 Labbé C, Heckman MG, Lorenzo-Betancor O, *et al.* MAPT haplotype H1G is associated with increased risk of dementia with Lewy bodies. *Alzheimers Dement* 2016; published online June 7. DOI:10.1016/j.jalz.2016.05.002.

25 Tsuang D, Leverenz JB, Lopez OL, *et al.* APOE ε4 increases risk for dementia in pure synucleinopathies. *JAMA Neurol* 2013; **70**: 223–8.

26 Peuralinna T, Myllykangas L, Oinas M, *et al.* Genome-wide association study of neocortical Lewy-related pathology. *Ann Clin Transl Neurol* 2015; **2**: 920–31.

27 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. *Nat Genet* 2014; **46**: 989–93.

28 GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* 2015; **348**: 648–60.

29 Westra H-J, Peters MJ, Esko T, *et al.* Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet* 2013; **45**: 1238–43.

30 International Parkinson's Disease Genomics Consortium (IPDGC), Wellcome Trust Case Control Consortium 2 (WTCCC2). A two-stage meta-analysis identifies several new loci for Parkinson's disease. *PLoS Genet* 2011; **7**: e1002142.

31 Smirnova T, Stinnakre J, Mallet J. Characterization of a presynaptic glutamate receptor. *Science* 1993; **262**: 430–3.

32 Schubert J, Siekierska A, Langlois M, *et al.* Mutations in STX1B, encoding a presynaptic protein, cause fever-associated epilepsy syndromes. *Nat Genet* 2014; **46**: 1327–32.

33 Vlaskamp DRM, Rump P, Callenbach PMC, *et al.* Haploinsufficiency of the STX1B gene is associated with myoclonic astatic epilepsy. *Eur J Paediatr Neurol* 2016; **20**: 489–92.

34 Vacic V, Ozelius LJ, Clark LN, *et al.* Genome-wide mapping of IBD segments in an Ashkenazi PD cohort identifies associated haplotypes. *Hum Mol Genet* 2014; **23**: 4693–702.

35 Li QS, Parrado AR, Samtani MN, Narayan VA, Alzheimer's Disease Neuroimaging Initiative. Variations in the FRA10AC1 Fragile Site and 15q21 Are Associated with Cerebrospinal Fluid Aβ1-42 Level. *PLoS One* 2015; **10**: e0134000.

36 Beecham GW, Hamilton K, Naj AC, *et al.* Genome-wide association meta-analysis of neuropathologic features of Alzheimer's disease and related dementias. *PLoS Genet* 2014; **10**: e1004606.

37 Berglund E, Stigbrand T, Carlsson SR. Isolation and characterization of a membrane glycoprotein from human brain with sequence similarities to cell adhesion proteins from chicken and mouse. *Eur J Biochem* 1991; **197**: 549–54.

38 Gennarini G, Bizzoca A, Picocci S, Puzzo D, Corsi P, Furley AJW. The role of Gpi-anchored axonal glycoproteins in neural development and neurological disorders. *Mol Cell Neurosci* 2016; published online Nov 18. DOI:10.1016/j.mcn.2016.11.006.

39 Compton AG, Albrecht DE, Seto JT, *et al.* Mutations in contactin-1, a neural adhesion and neuromuscular junction protein, cause a familial form of lethal congenital myopathy. *Am J Hum Genet* 2008; **83**: 714–24.

40 Derfuss T, Parikh K, Velhin S, *et al.* Contactin-2/TAG-1-directed autoimmunity is identified in multiple sclerosis patients and mediates gray matter pathology in animals. *Proc Natl Acad Sci U S A* 2009; **106**: 8302–7.

41 Johnson EO, Hancock DB, Levy JL, *et al.* Imputation across genotyping arrays for genome-wide association studies: assessment of bias and a correction strategy. *Hum Genet* 2013; **132**: 509–22.

42 Keller MF, Saad M, Bras J, *et al.* Using genome-wide complex trait analysis to quantify 'missing heritability' in Parkinson's disease. *Hum Mol Genet* 2012; **21**: 4996–5009.

43 Lek M, Karczewski KJ, Minikel EV, *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 2016; **536**: 285–91.

44 Zhang B, Gaiteri C, Bodea L-G, *et al.* Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease. *Cell* 2013; **153**: 707–20.

Tables

Table 1: Characteristics of the discovery cohort.

Country of origin	N	Neuropathological diagnosis	M:F	Mean age at onset	Passed quality controls	
					Total	Neuropathological
Australia	79	79	1.93	65	72	72
Canada	29	15	2.22	67.5	6	3
Finland	34	34	0.94	94.3 *	24	24
France	18	18	3.5	64.8	16	16
Germany	58	0	2.41	67.8	0	0
The Netherlands	133	133	1.71	78.7 *	132	132
Portugal	13	0	0.63	NA	11	0
Spain	133	16	0.94	73.2	132	15
UK	404	308	2.12	69.7	284	245
USA	786	705	1.93	71.9	539	467
Total Cases	**1687**	**1308**	**1.83**	**70.1**	**1216**	**974**
USA – Controls PSA	2847	0	0.88	NA	2832	0
USA – Controls SC	1523	0	0.78	38	959	
Total Controls	**4370**	**0**	**0.83**	**NA**	**3791**	**0**

N: number of samples; M:F: ratio of males to females. * Represents age at death, which was available for these cohorts. These values were not used for calculation of the complete mean age at onset.

Table 2: Characteristics of the replication cohort.

Country of origin	N	N neuropathological diagnosis	M:F	Mean age at onset
USA - cases	527	350	2.01	76.3
USA - controls	663	0	0.75	67.8 a

[a] Denotes age at examination for controls. For cases the age reflects age at onset for the clinical cases and age at death for the path-diagnosed cases.

Table 3: Top signals of association at each locus that passed genome-wide or suggestive thresholds for significance and their replication and meta-analysis p-values.

							Discovery						Replication							Meta-Analysis			
Named Region	**CHR**	**Position**	**Variant**	**R2**	**Eur_AF**	**MA**	**MAF_A**	**MAF_U**	**OR**	**L95**	**U95**	**P-value**	**Power**	**MAF_A**	**MAF_U**	**OR**	**L95**	**U95**	**P-value**	**OR**	**L95**	**U95**	**P-value**
APOE	19	45411941	rs429358	0.949	0.149	C	0.283	0.140	2.40	2.14	2.7	1.05E-48	1	0.282	0.148	2.74	2.15	3.49	4.00E-16	2.46	2.22	2.74	3.31E-64
BCL7C/STX1B	16	30886643	rs897984*	0.984	0.609	T	0.334	0.405	0.74	0.67	0.82	3.30E-09	0.96	0.368	0.388	0.98	0.81	1.19	0.83	0.77	0.71	0.85	1.19E-08
SNCA	4	90756550	rs7681440*	0.996	0.52	C	0.411	0.483	0.73	0.66	0.81	6.39E-10	0.95	0.38	0.47	0.68	0.56	0.82	6.00E-05	0.73	0.67	0.79	9.22E-13
GBA	1	155135036	rs35749011	0.957	0.014	G	0.033	0.014	2.55	1.88	3.46	1.78E-09	0.83	0.044	0.022	1.81	1.05	3.11	0.033	2.27	1.75	2.95	6.57E-10
GABRB3	15	26840998	rs1426210	0.982	0.315	G	0.348	0.293	1.34	1.21	1.48	2.62E-08	0.9	0.281	0.307	0.84	0.68	1.04	0.1	1.22	1.11	1.33	2.05E-05
Neuropathologically diagnosed cases																							
APOE	19	45411941	rs429358	0.949	0.149	C	0.292	0.140	2.52	2.23	2.85	2.77E-49	-	-	-	-	-	-	-	-	-	-	-
BCL7C/STX1B	16	30886643	rs897984*	0.984	0.609	T	0.332	0.405	0.73	0.65	0.81	4.32E-09	-	-	-	-	-	-	-	-	-	-	-
SNCA	4	90756550	rs7681440*	0.996	0.52	C	0.409	0.483	0.73	0.66	0.81	2.82E-09	-	-	-	-	-	-	-	-	-	-	-
GBA	1	155135036	rs35749011	0.957	0.014	G	0.037	0.014	2.87	2.10	3.90	2.67E-11	-	-	-	-	-	-	-	-	-	-	-
GABRB3	15	26840998	rs1426210	0.982	0.315	G	0.350	0.293	1.34	1.20	1.45	1.21E-07	-	-	-	-	-	-	-	-	-	-	-

CHR: Chromosome. R2: Imputation R-squared of each specific variant from HRC. OR: Odds ratio. L95: Lower 95% interval. U95: Upper 95% confidence interval. Eur_AF is the alternate allele frequency derived from the European population of gnomAD[43]. * Represents variants for which the gnomAD allele frequency corresponds to the alternate allele and not the effect allele. Power refers to the calculated statistical power to replicate the discovery signal, taking into account the replication sample size, effect and frequency in discovery and an association threshold of p<0.05.

Table 4: Top gene burden results

CHR	BEGIN	END	ID	NS	FRAC_WITH_RARE	NUM_ALL_VARS	NUM_PASS_VARS	NUM_SING_VARS	PVALUE
1	155204797	155210498	GBA	5016	0.05622	8	6	1	1.29E-13
22	39262224	39267761	CBX6	5016	0.010965	6	3	0	1.66E-05
11	130058428	130079477	ST14	5016	0.076754	20	11	2	4.29E-05
10	129347767	129350889	NPS	5016	0.076555	5	3	1	6.74E-05
4	40428010	40434855	RBM47	5016	0.0099681	3	2	1	0.00011289
11	18047141	18057637	TPH1	5016	0.0091707	8	5	0	0.00022217
6	31237124	31239829	HLA-C	5016	0.1262	32	10	1	0.00028923
19	45971941	45976122	FOSB	5016	0.00079745	4	2	1	0.00036517
1	44435905	44438171	DPH2	5016	0.004386	9	6	2	0.00043723

2	238785923	238820379	RAMP1	5016	0.00079745	2	2	1	0.00049746

CHR: Chromosome. NS: Number of samples with non-missing genotypes. FRAC_WITH_RARE : Fraction of individual carrying rare variants below the allele frequency threshold (0.05). NUM_ALL_VARS : Number of all variants defining the gene group. NUM_PASS_VARS : Number of variants passing the frequency and call-rate thresholds. NUM_SING_VARS : Number of singletons among variants in NUM_PASS_VARS.

Legends to Figures

Figure 1:
Manhattan plot showing genome-wide p-values of association. The p-values were obtained by logistic regression analysis using the first six principal components and gender as covariates. The y-axis shows −log10 p-values of 8,397,716 SNPs, and the x-axis shows their chromosomal positions. The y-axis was truncated at p-value of $1x10^{-25}$. Horizontal red and green dotted lines represent the thresholds of $p= 5x10^{-8}$ for Bonferroni significance and $p=1x10^{-6}$ for selecting SNPs for replication, respectively.

Figure 2:
a) Boxplot showing the association between rs7681440 genotypes and *RP11-67M1.1* expression in the cerebellum in 103 healthy post-mortem samples ($p=2.00x10^{-07}$) from the GTEx Consortium. Carriers of the GG genotype (alternative allele) show the lowest levels of expression of the gene. Medians, interquartile ranges and individual data points are indicated. See the GTEx website for details on methods.
b) Boxplot showing the association between rs7681154 and SNCA expression ($p=2.87x10^{-11}$) in brain cerebellum in 468 healthy post-mortem subjects from the Harvard Brain Bank Resource Center (www.brainbank.mclean.org) [44]. Individuals with the alternate allele (C) had increased SNCA expression in the cerebellum, on average, compared to those with the reference allele (G) Sample size for each genotype group is denoted in parentheses. Details on the subjects, experiments, and analytical methods of the eQTL study of the Harvard Brain Bank Resource Center samples are described in Zhang et al. 2013 and www.brainbank.mclean.org. Abbreviations: Homo Ref, homozygous for reference allele; Het, heterozygous; Homo Alt, homozygous for the alternative allele.

Figure 3:
DLB heritability by chromosome. Heritability (y-axis) per chromosome is plotted against chromosome length (x-axis). The red line represents heritability regressed on chromosome length and the shaded grey area represents the 95% confidence interval of the regression model.

Figure 4:
Regional association plot for the *SNCA* locus. Purple represents rs1372517, which is the most associated SNP at the locus also present in the 1000Genomes dataset. The variant rs1372517 is in complete LD with rs7681440. Colours represent LD derived from 1000Genomes between each variant and the most associated SNP.

Figure 5:
Regional association plot for the *BCL7C/STX1B* locus. Purple represents the most associated SNP. Colours represent LD derived from 1000Genomes between each variant and the most associated SNP.