A comprehensive screening of copy-number variability in dementia with Lewy bodies

Celia Kun-Rodrigues¹, Tatiana Orme¹, ², Susana Carmona¹, ², Dena G. Hernandez³, ⁴, Owen A. Ross⁵, John D. Eicher⁶, Claire Shepherd⁷, Laura Parkkinen⁸, Lee Darwent¹, ², Michael G. Heckman⁹, Sonja W. Scholz¹⁰, Juan C. Troncoso¹¹, Olga Pletnikova¹¹, Olaf Ansorge⁶, Jordi Clarimon¹², Alberto Lleo¹², Estrella Morenas-Rodriguez¹², Lorraine Clark¹³, Lawrence S Honig¹³, Karen Marder¹³, Afina Lemstra¹⁴, Ekaterina Rogaeva¹⁵, Peter St. George-Hyslop¹⁵, ¹⁶, Elisabet Londos¹⁷, Henrik Zetterberg¹⁸, Imelda Barber¹⁹, Anne Braae¹⁹, Kristelle Brown¹⁹, Kevin Morgan¹⁹, Claire Troakes²⁰, Safa Al-Sarraj²⁰, Tammaryn Lashley²¹, Janice Holton²¹, Yaroslau Compta²², Viviana Van Deerlin²³, Geidy E Serrano²⁴, Thomas G Beach²⁴, Suzanne Lesage²⁶, Douglas Galasko²⁶, Eliezer Masliah²⁷, Isabel Santana²⁸, Pau Pastor²⁹, Monica Diez-Fairen²⁹, Miquel Aguilar²⁹, Pentti J. Tienari³⁰, Liisa Myllykangas³¹, Minna Oinas³², Tamas Revesz²¹, Andrew Lees²¹, Brad F Boeve³³, Ronald C. Petersen³³, Tanis J Ferman³⁴, Valentina Escott-Price³⁵, Neill Graff-Radford³⁶, Nigel J. Cairns³⁷, John C. Morris³⁷, Stuart Pickering-Brown³⁸, David Mann³⁸, Glenda M. Halliday³⁹, ⁴⁰, John Hardy¹, John Q. Trojanowski²³, Dennis W. Dickson⁵, Andrew Singleton³, David J. Stone⁴¹, Rita Guerreiro¹, ², ⁴², Jose Bras¹, ², ⁴²,*

* Corresponding author. Email: j.bras@ucl.ac.uk

1 - Department of Molecular Neuroscience, UCL Institute of Neurology, London, UK
2 - UK Dementia Research Institute (UK DRI) at UCL, London, UK
3 - Laboratory of Neurogenetics, National Institutes on Aging, NIH, Bethesda, MD, USA
4 - German Center for Neurodegenerative Diseases (DZNE)-Tubingen
5 - Department of Neuroscience, Mayo Clinic, Jacksonville, FL, USA
6 - Genetics and Pharmacogenomics, Merck Research Laboratories, Boston, MA, USA
7 - Neuroscience Research Australia, Sydney, Australia and School of Medical Sciences, Faculty of Medicine, University of New South Wales, Sydney, Australia
8 - Nuffield Department of Clinical Neurosciences, Oxford Parkinsons Disease Centre, University of Oxford, Oxford, UK
9 - Division of Biomedical Statistics and Informatics, Mayo Clinic, Jacksonville, FL, USA
10 - Neurodegenerative Diseases Research Unit, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, Maryland, USA
11 - Department of Pathology (Neuropathology), Johns Hopkins University School of Medicine, Baltimore, MD, USA
12 - 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
13 - Taub Institute for Alzheimer Disease and the Aging Brain and Department of Pathology and Cell Biology, Columbia University, New York, NY, USA

14 - Department of Neurology and Alzheimer Center, Neuroscience Campus Amsterdam, VU University Medical Center, Amsterdam, the Netherlands

15 - Tanz Centre for Research in Neurodegenerative Diseases and department of Medicine, University of Toronto, Ontario, Canada

16 - Department of Clinical Neurosciences, Cambridge Institute for Medical Research, University of Cambridge, Cambridge, UK

17 - Clinical Memory Research Unit, Institution of Clinical Sciences Malmo, Lund University, Sweden

18 - UK Dementia Research Institute at UCL, London UK, Department of Neurodegenerative Disease, UCL Institute of Neurology, London, UK, and Clinical Neurochemistry Laboratory, Institute of Neuroscience and Physiology, Sahlgrenska Academy at the University of Gothenburg, Molndal, Sweden

19 - Human Genetics, School of Life Sciences, Queens Medical Centre, University of Nottingham, Nottingham, UK

20 - Department of Basic and Clinical Neuroscience and Institute of Psychiatry, Psychology and Neuroscience, Kings College London, London, UK

21 - Queen Square Brain Bank, Department of Molecular Neuroscience, UCL Institute of Neurology, London, UK

22 - 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

23 - 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

24 - Banner Sun Health Research Institute, 10515 W Santa Fe Drive, Sun City, AZ 85351, USA

25 - Inserm U1127, CNRS UMR7225, Sorbonne Universites, UPMC Univ Paris 06, UMR and S1127, Institut du Cerveau et de la Moelle epiniere, Paris, France

26 - Department of Neurosciences, University of California, San Diego, La Jolla, CA, United States; Veterans Affairs San Diego Healthcare System, La Jolla, CA, United States

27 - 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

28 - Neurology Service, University of Coimbra Hospital, Coimbra, Portugal

29 - 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 Neuroporoticas (CIBERNED), Madrid, Spain

30 - Molecular Neurology, Research Programs Unit, University of Helsinki, Department of Neurology, Helsinki University Hospital, Helsinki, Finland

31 - Department of Pathology, Haartman Institute, University of Helsinki and HUSLAB

32 - Department of Neuropathology and Neurosurgery, Helsinki University Hospital and University of Helsinki, Helsinki, Finland
33 - Neurology Department, Mayo Clinic, Rochester, MN, USA
34 - Department of Psychiatry and Department of Psychology, Mayo Clinic, Jacksonville, FL, USA
35 - MRC Centre for Neuropsychiatric Genetics and Genomics, School of Medicine, Cardiff University, Cardiff, UK
36 - Department of Neurology, Mayo Clinic, Jacksonville, FL, USA
37 - Knight Alzheimer’s Disease Research Center, Department of Neurology, Washington University School of Medicine, Saint Louis, MO, USA
38 - Institute of Brain, Behaviour and Mental Health, Faculty of Medical and Human Sciences, University of Manchester, Manchester, UK
39 - Neuroscience Research Australia, Sydney, Australia and School of Medical Sciences, Faculty of Medicine, University of New South Wales, Sydney, Australia
40 - Brain and Mind Centre, Sydney Medical School, The University of Sydney, Sydney, Australia
41 - Genetics and Pharmacogenomics, Merck Research Laboratories, West Point, Pennsylvania, USA
42 - Department of Medical Sciences and Institute of Biomedicine, iBiMED, University of Aveiro, 3810-193 Aveiro, Portugal
Abstract

The role of genetic variability in dementia with Lewy bodies (DLB) is now indisputable, however data regarding copy number variation (CNV) in this disease has been lacking. Here, we used whole-genome genotyping of 1,454 DLB cases and 1,525 controls to assess copy number variability. We used two algorithms to confidently detect CNVs, performed a case-control association analysis, screened for candidate CNVs previously associated with DLB-related diseases, and performed a candidate gene approach to fully explore the data. We identified five CNV regions with a significant genome-wide association to DLB, two of these were only present in cases and absent from publicly available databases: one of the regions overlapped *LAPTM4B*, a known lysosomal protein; whilst the other overlapped the *NME1* locus and *SPAG9*. We also identified DLB cases presenting CNVs in genes previously associated with DLB or related neurodegenerative diseases, namely *SNCA* and *MAPT*. To our knowledge, this is the first study reporting genome-wide CNVs in a large DLB cohort.

Keywords

Dementia with Lewy bodies; Copy number variants; *MAPT*; *SNCA*; Genome-wide
1. Introduction

Dementia with Lewy bodies (DLB) is a common and complex form of dementia and its diagnosis can often be complicated by phenotypic similarities with Alzheimer’s disease (AD), Parkinson’s disease (PD) or even frontotemporal dementia (FTD) (Claassen et al., 2008; Heidebrink, 2002). A more accurate DLB diagnosis is usually obtained by integrating clinical and pathological data from brain autopsy (McKeith et al., 2017).

Genetic studies in DLB have been limited, certainly in comparison with studies on AD or PD, for a number of reasons, most notably because DLB has not been historically considered a genetic disease, given the lack of multiplex kindreds where the disease segregates. Additionally, large cohorts of patients are difficult to collect given the frequency of the disease and the rate of misdiagnosis. Despite this, recent studies have conclusively shown that there is a role for genetics in the etiology of DLB (Bras et al., 2014; Guerreiro et al., 2018, 2016; Nalls et al., 2013). Exome sequencing studies have been performed in small cohorts; as have case studies and Sanger sequencing of specific target genes (Clark et al., 2009; Keogh et al., 2016; Koide et al., 2002; Ohtake et al., 2004). Copy number variants (CNVs) have not been assessed thus far in DLB, particularly in an unbiased manner and at a genome-wide level.

CNVs have been widely studied in a number of neurological conditions, particularly in developmental phenotypes such as schizophrenia (SCZ) and autism (Glessner et al., 2009; Marshall et al., 2017; McCarthy et al., 2009) where several microdeletions and microduplications have been found to be associated with both diseases (Bassett et al., 2017; Cook et al., 1997; McCarthy et al., 2009; Stefansson et al., 2008; Weiss et al., 2008). In these phenotypes, CNVs play a prominent role in the disease genetic architecture.

Several studies have analyzed CNVs in AD, where APP duplications have been unequivocally shown to cause disease (Delabar et al., 1987; Ghani et al., 2012; Swaminathan et al., 2012, 2011; Zheng et al., 2015, 2014). In PD, pathogenic CNVs are also known to occur in SNCA and PARK2 (Chartier-Harlin et al., 2004; Ibáñez et al., 2004; Lesage et al., 2008; Waters and Miller, 1994). Together, these data show that CNVs are an important mutational event in neurological conditions.
Here, we report the first genome-wide analysis of CNVs in DLB in a large cohort of patients, many of which with neuropathology diagnoses of DLB. We performed a case-control association study that was complemented by discovery stage analyses guided by candidate genes and CNVs previously reported as being associated with DLB-related neurodegenerative diseases.

2. Materials and Methods

2.1. Sample selection

A total of 1,454 patients diagnosed with DLB and of European ancestry were selected for this study. Diagnosis of DLB was made according to clinical or pathological criteria (McKeith et al., 2005). Briefly, these included 298 clinically diagnosed and 1,156 neuropathologically diagnosed cases. Detailed sample and processing information has been described previously (Guerreiro et al., 2018). Data from 1,525 control samples was obtained from The Genetic Architecture of Smoking and Smoking Cessation study (phs000404.v1.p1) publicly available at the database of Genotypes and Phenotypes (https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000404.v1.p1). Supplementary Figure 1 shows an overview of the study design, different QC steps and analyses performed.

2.2. Genotyping, quality control and CNV calling

Seven hundred and fifty four DLB samples were genotyped on HumanOmni2.5Exome-8 v1.0.B Illumina arrays, and 700 DLB samples were genotyped using Infinium OmniExpress-24 v1.2.A1 Illumina arrays (Illumina, Inc., CA, USA). Control samples were genotyped on HumanOmni2.5-4v1.D arrays (Illumina, Inc., CA, USA). Intensity files were analysed using GenomeStudio v2011.1 software (Illumina, Inc., CA, USA) along with the respective manufacturer’s cluster files. Quality control (QC) procedures were performed in GenomeStudio (GS) prior to CNV analysis as described by Jarick and colleagues (Jarick et al., 2014). In short, samples with call rates lower than 0.97 were filtered out. SNP statistics were re-calculated following visual inspection of B allele frequency (BAF) and Log R ratio (LRR) plots. SNPs with GenTrain Scores below 0.7 were excluded. Lastly, samples
with substantial cryptic relatedness scores (PI_HAT >0.1) were removed, as previously described (Guerreiro et al., 2018).

CNV calls were generated using two different algorithms: cnvPartition v2.3.0 (Illumina, Inc.) and PennCNV v1.0.4 (Wang et al., 2007). CNV calling based on cnvPartition was performed by GS with default parameters. For PennCNV, probe positions, LRR and BAF values for samples that passed QC procedures were exported from GS. Population frequency of the B allele (PFB) files were calculated for each array separately. All Smoking Cessation samples were used to generate CNVs in cnvPartition and PennCNV, but only a subset of the best performing 700 samples was used for the compilation of the PFB file in PennCNV to match the number of samples used for cases. PennCNV GC-model files were then created based on these PFBs. Lastly, CNVs were inferred by PennCNV using the hidden Markov model (HMM) and the GC-model for wave adjustment. Calls for the X chromosome were generated separately. Chromosomes Y, MT and XY were not analyzed.

2.3. **CNV quality control and analysis**

To improve the quality of CNVs, only calls generated by both algorithms were kept, while calls made by a single algorithm or calls of opposing type (for example, assigned as a deletion by one algorithm and as duplication by the other) were discarded. Adjacent CNVs were merged if the length of the sequence between them was smaller than 50% of the length of the larger CNV. CNVs were excluded if they were overlapping telomeres, centromeres, known segmental duplications, the immunoglobulin or the T cell receptor loci. Samples having LRR SD > 0.28, BAF drift > 0.002, Waviness Factor (WF) > 0.04, or having more CNV calls than 3*SD + median were excluded (Marshall et al., 2017; Need et al., 2009).

To identify potentially pathogenic CNVs, we analyzed CNVs spanning known genes. We used the Database of Genomic Variants (DGV) (http://dgv.tcag.ca/dgv/app/home, accessed November 2017) to determine the population frequency of CNVs (MacDonald et al., 2014). This information was complemented with the frequency from clinical samples available in DECIPHER v9.18 (https://decipher.sanger.ac.uk/, accessed November 2017).
2.4 Case-control association analysis

Case-control association analysis was implemented using ParseCNV (Glessner et al., 2013). Standard ParseCNV quality metrics were used to filter out low quality results. CNVs that were genome-wide significant (p-value < 5x10^{-4} as suggested by (Glessner et al., 2013)), had a minimum length of 50 kb, and passed visual inspection in GS were selected for further analyses.

2.5. Candidate CNVs approach

CNVs previously described in AD (Ghani et al., 2012; Heinzen et al., 2010; Swaminathan et al., 2012, 2011; Zheng et al., 2015, 2014), PD (Bademci et al., 2010; Liu et al., 2013; Mok et al., 2016; Pankratz et al., 2011) and FTD (Gijselinck et al., 2008) were specifically investigated in these data (Supplementary Table 1). This analysis was performed in the complete set of CNV results after QC, disregarding the filters used for the case-control association analysis performed by ParseCNV.

2.6. Candidate genes approach

CNVs located in known AD, PD, FTD and DLB genes were also assessed (Brás et al., 2015; Guerreiro et al., 2018, 2015, 2013; Jansen et al., 2015; Keogh et al., 2016; Koide et al., 2002; Ohtake et al., 2004; Saitoh et al., 1995). Supplementary Table 2 lists all genes studied using this approach.

3. Results

3.1. CNV calling and QC steps

After QC steps at the GS level, a total of 2,819 samples (1,294 cases and 1,525 controls) remained for further analyses. From the 754 DLB samples genotyped with HumanOmni2.5 arrays, 616 (81.7%) samples were kept and from the 2,567,845 probes in this array, 2,496,600 (97.2%) passed quality control. Six hundred and seventy-eight (96.9%) samples of the 700 samples genotyped with OmniExpress arrays passed quality control and from the 713,599 probes available in this chip, 698,680 (97.9%) probes remained. All controls from the Smoking Cessation database had good...
quality genotypes (call rate > 0.97) and, consequently, no samples were excluded, and 2,390,384 (97.8%) of the 2,443,177 probes were kept.

After combining the results obtained by the two CNV calling algorithms (cnvPartition and PennCNV), excluding samples due to their relatedness and performing the PennCNV QC steps on the LRR and BAF values and number of calls, a final number of 2,615 individuals (1,187 DLB cases and 1,428 control samples) and 80,416 CNVs were analysed.

3.2. Case-control association analysis

Of the 494 CNV regions (CNVR) resulting from the ParseCNV analysis, only 5 passed QC checks and were statistically significant (Table 1). Of these, two of the regions were not present in our control population or in public databases: a deletion overlapping LAPTM4B (p=6.29x10^{-7}) and a CNVR overlapping SPAG9-NME1-NME2 (p=2.72x10^{-4}) (Figure 1).

3.3. Candidate CNVs approach

Five CNVs previously associated with DLB-related neurodegenerative diseases were found in DLB patients (Table 1). Two of these were present in the control group with a higher frequency than in the patients’ group, and the remaining three are described in public databases and are detailed next. The duplication identified on chromosome 12 overlapping DDX11 and OVOS2 has a frequency tenfold higher in DGV than in the DLB cohort. The 16p11.2 microduplication found in one DLB patient, has a frequency of 1.69x10^{-4} in the DECIPHER database but does not occur in any control samples or in the DGV database. One of the CNVs that was previously significantly associated with AD locates at chr8:2792874-4852328 and overlaps CSMD1 (Swaminathan et al., 2011). At this locus, we identified over 100 CNVs in cases and controls. Therefore, in Table 1, we only report the shorter interval that showed suggestive significance in the case-control association analysis.

3.4. Candidate genes approach
We investigated CNVs in genes known to be associated with diseases that are related to DLB (Supplementary Table 2) and identified a total of 8 CNVs (Table 1). These included one duplication in APP occurring in a clinically diagnosed case. This large duplication is not present in the databases or in the control cohort. Two samples were found to carry duplications spanning MAPT, and one neuropathologically diagnosed patient was found to carry a SNCA duplication (Figure 2). PARK2 was found to have many copy number losses and gains in controls (n=28) and cases (n=13) but none as homozygous. A duplication including CHCHD10 was also identified in a neuropathologically diagnosed DLB patient.

4. Discussion
We performed a systematic analysis of CNVs in a large cohort of DLB patients using three main approaches. The first of these approaches was a case-control association analysis which resulted in five significant CNVRs that have not been previously described as associated with the disease. The most significant result from this analysis was a deletion spanning a lysosome-associated transmembrane protein, LAPTM4B. Intraneuronal alpha-synuclein clearance likely occurs through a variety of mechanisms in order to maintain protein homeostasis. However, recent data has highlighted the importance of lysosomal pathways for degradation of this protein (Webb et al., 2003). While we cannot directly link this CNV to the development of DLB in these cases, it is interesting that a lysosomal enzyme is the top hit in our association analysis, given the prominent role of the lysosome in Lewy body diseases. Interestingly, a member of the same protein family, LAPTM5, was one of the top hits for incidental DLB in a recent network analysis study (Santpere et al., 2017).

Also associated with DLB and absent from publicly available databases was a deletion overlapping the NME1 locus. NME1 is involved in purine metabolism, which has been reported to be disrupted in AD, PD and Creutzfeldt–Jakob disease. NME1 mRNA was also found to be reduced in these diseases (Ansoleaga et al., 2016, 2015; Garcia-Esparcia et al., 2015). Here, we identified a deletion
at the 3’-end NME1 which could be consistent with a reduced expression of the gene in DLB, although this was not tested in the present study.

Using a candidate gene approach where we analyzed genes known to have a role in DLB and DLB-related diseases we identified several CNVRs of interest. The hallmark of DLB at autopsy is the accumulation of alpha-synuclein protein within neurons and their processes, termed Lewy bodies and Lewy neurites (Spillantini et al., 1997). Variants in the SNCA gene, which encodes alpha-synuclein, have been previously associated with the risk of developing DLB (Bras et al., 2014; Guerreiro et al., 2018). In addition to point mutations, CNVs including SNCA are known to cause PD, and over the past years evidence suggested that this gene may also be duplicated in DLB. Nishioka and colleagues identified a PD family with a duplication spanning all of SNCA and MMRN1 where the proband was later neuropathologically diagnosed as DLB (Nishioka et al., 2006b; Obi et al., 2008). Four neuropathologically diagnosed DLB cases presented a large duplication from DSPP to PDLIM5 including SNCA, three of these were heterozygous and one was homozygous (Ikeuchi et al., 2008). A duplication in SNCA was also described in a probable DLB patient in a study with 99 cases (Meeus et al., 2012). Here, we add to this body of evidence, by identifying another patient neuropathologically diagnosed with DLB carrying a SNCA duplication. In our DLB cohort this duplication shows a similar frequency to that provided by DGV. This frequency in DGV results from two entries in the database. When looking in more detail at these entries, they are associated with the same Human Genome Diversity Project (HGDP) sample from Cambodia (HGDP00721). Given that information for each HGDP samples is limited to sex of the individual, population and geographic origin, it is possible this sample originated from a PD or DLB patient; or from an asymptomatic carrier, as these have previously been reported, with SNCA multiplications having particular low penetrance levels in Asian populations (Ahn et al., 2008; Nishioka et al., 2006a). It is also possible the duplication reported is an artifact caused by the creation or passage of the lymphoblast cell lines used to extract DNA (Simon-Sanchez et al., 2007).
We also identified one heterozygous duplication encompassing CHCHD10, a gene previously shown to cause FTD. However, given that CHCHD10 FTD-associated mutations are loss-of-function (Perrone et al., 2017), it is unlikely that a duplication of the gene would be pathogenic.

GABRB3 is a gamma-aminobutyric acid (GABA) receptor that was reported as associated with DLB in a recent GWAS, but that did not survive independent replication (Guerreiro et al., 2018). However, loss of GABA receptors could underlie the typical visual hallucinations in DLB (Khundakar et al., 2016), and, because of this, we specifically looked at CNVs in GABRB3, and identified a duplication in one case. It is tempting to speculate that, given the GWAS discovery results, the fact that GABA receptors neurotransmission is altered in DLB (Santpere et al., 2017), and the CNV detected here, genetic variability in GABA receptors may in fact modulate risk for DLB.

We identified two clinically diagnosed DLB samples with MAPT duplications (Figure 3). MAPT was not found to be significantly associated with DLB in recent GWAS (Bras et al., 2014; Guerreiro et al., 2018), but the MAPT H1 haplotype was previously described as a possible risk factor for DLB (Cervera-Carles et al., 2016; Labbé et al., 2016). Previous studies of small cohorts of FTD patients have not revealed causative MAPT duplications (Lladó et al., 2007; Skoglund et al., 2009) but the screening of French FTD patients including multiplex families led to the identification of an heterozygous partial deletion of MAPT (Rovelet-Lecrux et al., 2009) and of a 17q21.31 microduplication in an atypical FTD case (Rovelet-Lecrux et al., 2010). More recently, MAPT duplications were shown to increase expression of MAPT mRNA and were found to cause tangle pathology without Aβ deposition in probable AD patients (Le Guennec et al., 2017).

PARK2 homozygous CNVs are the most common copy-number cause of PD, accounting for more than 50% of all pathogenic mutations in the gene and more frequently affecting the region between exons 2 and 7 (Hedrich et al., 2004; Kim et al., 2012). Our results showed no significant differences between DLB cases and controls, similar to the findings by Kay and colleagues in PD (Kay et al., 2010). Additionally, we did not find any homozygous PARK2 CNVs suggesting that CNVs in this gene do not play a causative role in DLB.
APP duplications are known to cause AD (Delabar et al., 1987; Ghani et al., 2012; Swaminathan et al., 2012, 2011; Zheng et al., 2015, 2014). The sample carrying an APP duplication in our cohort has a clinical diagnosis of DLB without neuropathological confirmation - it is therefore possible this is an AD case misdiagnosed as DLB. However, there have been reports in the literature of DLB cases associated with APP duplications. For example, in a French family presenting with a diverse phenotype, APP duplication was associated with DLB confirmed by neuropathological findings (Guyant-Marechal et al., 2008). Similarly, one case with Lewy body-variant AD was reported in a multigenerational dementia family from the Netherlands (Sleegers et al., 2006). DLB cases frequently present Aβ pathology at autopsy (Hepp et al., 2016), and it has been suggested that Aβ accumulation can trigger Lewy body disease (Masliah et al., 2001).

There are two main limitations in this study: first, this is a relatively small-sized cohort, which means we cannot confidently assess associations of CNVs with low effect sizes on disease; second, we did not perform independent replication of these findings, which precludes us from establishing definite associations or causes of disease. Despite these limitations we report on the first systematic analysis of CNVs in a large cohort of DLB patients, using well-established analytical practices, as well as a high proportion of autopsy-confirmed cases. We identify potential disease causing CNVs as well as potential novel candidate genes for DLB.

Disclosure statement

The authors have no actual or potential conflicts of interest.

Acknowledgements

Rita Guerreiro and Jose Bras’ work is funded by fellowships from the Alzheimer’s Society. Tatiana Orme’s work is funded 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. Uddall 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 Folkhälsoan Research Foundation and the Finnish Academy. This work was in part supported by the Canadian Consortium on Neurodegeneration in Aging (ER). 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). 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.
Figures captions

Figure 1. Schematic representation of CNVs.

Schematic representation of statistically significant CNVs resulting from the case-control analysis that were not found in controls or publicly available databases. CNVs overlapping \textit{LAPTM4B} in A; and \textit{SPAG9/NME1} in B. Known transcripts are represented at the bottom of each panel. Passing QC SNPs used for the CNV calling are depicted by colour according to the genotyping array (red: OmniExpress, blue: Omni2.5M, green: Omni2.5M for controls). CNVs (duplications) are depicted as bars above the genes. The grey shadow area represents the associated region that is genome-wide significant.

Figure 2. Duplication identified at the \textit{SNCA} locus.

Log R ratio and B allele frequency plots of the CNV identified in the \textit{SNCA} locus. Each point represents a SNP according to location in chromosome 4 (position on X axis). The genomic duplication is indicated by an increase in log R ratio and B allele frequency clusters outside the expected values of 1 (B/B), 0.5 (A/B), and 0 (A/A). Genes are represented at the bottom as black bars. SNPs inside the CNV region are represented in red, SNPs outside the CNV region are represented in blue, and SNPs in \textit{SNCA} are represented in green.

Figure 3. Duplications identified at the \textit{MAPT} locus.

Log R ratio and B allele frequency plots of the CNVs identified at the \textit{MAPT} locus in two clinically diagnosed DLB cases. Each point represents a SNP according to location in chromosome 17 (position on X axis). The genomic duplication is indicated by an increase in log R ratio and B allele frequency clusters outside the expected values of 1 (B/B), 0.5 (A/B), and 0 (A/A). Genes are represented at the bottom as black bars. SNPs inside the CNV regions are represented in red, SNPs outside the CNV regions are represented in blue, and SNPs in \textit{MAPT} are represented in green.
### Table 1 Results from the three types of analyses performed

<table>
<thead>
<tr>
<th>Genes</th>
<th>Location</th>
<th>CNV</th>
<th>p-value</th>
<th>Cases (Neuro*)</th>
<th>Controls</th>
<th>Cases frequency</th>
<th>Controls frequency</th>
<th>DGV frequency</th>
<th>DECIPHER frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Case-control association analysis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADGRG7,TFG</td>
<td>chr3:100357671-100439759</td>
<td>Gain</td>
<td>8.93x10⁻⁵</td>
<td>21 (18)</td>
<td>13</td>
<td>1.77x10⁻²</td>
<td>9.10x10⁻³</td>
<td>1.43x10⁻²</td>
<td>2.09x10⁻²</td>
</tr>
<tr>
<td>PDZD2</td>
<td>chr5:32101400-32106628</td>
<td>Gain</td>
<td>2.94x10⁻⁶</td>
<td>14 (12)</td>
<td>0</td>
<td>1.18x10⁻²</td>
<td>0</td>
<td>3.00x10⁻⁴</td>
<td>1.69x10⁻⁴</td>
</tr>
<tr>
<td>LAPTM4B</td>
<td>chr8:98755434-98800334</td>
<td>Loss</td>
<td>6.29x10⁻⁷</td>
<td>12 (6)</td>
<td>0</td>
<td>1.01x10⁻²</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>MSR1</td>
<td>chr8:15948235-16021468</td>
<td>Loss</td>
<td>1.20x10⁻⁴</td>
<td>13 (7)</td>
<td>4</td>
<td>1.10x10⁻²</td>
<td>2.80x10⁻³</td>
<td>4.30x10⁻³</td>
<td>7.10x10⁻³</td>
</tr>
<tr>
<td>NME1,NME1-NME2,SPAG9</td>
<td>chr17:49177096-49231786</td>
<td>Loss</td>
<td>2.72x10⁻⁴</td>
<td>9 (4)</td>
<td>0</td>
<td>7.58x10⁻³</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Candidate CNVs approach</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CSMD1</td>
<td>chr8:4,033,908-4,126,540</td>
<td>Loss</td>
<td>2.89x10⁻²⁹</td>
<td>3 (2)</td>
<td>0</td>
<td>2.53x10⁻³</td>
<td>0</td>
<td>3.00x10⁻⁴</td>
<td>0</td>
</tr>
<tr>
<td>DDX11,OVOS2</td>
<td>chr12:31,249,834-31,407,303</td>
<td>Gain</td>
<td>1.41x10⁻²⁹</td>
<td>4 (2)</td>
<td>0</td>
<td>3.37x10⁻³</td>
<td>0</td>
<td>4.61x10⁻³</td>
<td>0</td>
</tr>
<tr>
<td>CYFIP1,GOLGA8,NIPA1,NIPA2,TUBGCP5,WHAMML1</td>
<td>chr15:22,750,305-23,272,733</td>
<td>Gain</td>
<td>na</td>
<td>5 (5)</td>
<td>8</td>
<td>4.21x10⁻³</td>
<td>5.60x10⁻³</td>
<td>2.50x10⁻³</td>
<td>1.86x10⁻³</td>
</tr>
<tr>
<td>CHRNA7,OTUD7A</td>
<td>chr15:31,932,865-32,515,849</td>
<td>Loss or Gain</td>
<td>na</td>
<td>6 (5)</td>
<td>13</td>
<td>5.05x10⁻³</td>
<td>9.10x10⁻³</td>
<td>6.90x10⁻³</td>
<td>8.79x10⁻³</td>
</tr>
<tr>
<td>ASPHD1,BOLA2,C16orf54,CDIPT,CDIPT-AS1,KIF22,MAZ,MVP,PAGR1,PQRT2,QPRT,SEZ6L2,SPN,ZG16</td>
<td>chr16:29,595,483-29,912,902</td>
<td>Gain</td>
<td>na</td>
<td>1 (1)</td>
<td>0</td>
<td>8.42x10⁻⁴</td>
<td>0</td>
<td>0</td>
<td>1.69x10⁻⁴</td>
</tr>
<tr>
<td><strong>Candidate genes approach</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DNAJC6,LEPR,LEPROT</td>
<td>chr1:65,854,556-65,955,725</td>
<td>Gain</td>
<td>na</td>
<td>1 (0)</td>
<td>1</td>
<td>8.42x10⁻⁴</td>
<td>7.00x10⁻⁴</td>
<td>1.72x10⁻⁴</td>
<td>2.53x10⁻³</td>
</tr>
<tr>
<td>SNCA,SNCA-AS1,GPRIN3,MMRN1,CCSER1</td>
<td>chr4:90,035,549-91,420,358</td>
<td>Gain</td>
<td>na</td>
<td>1 (1)</td>
<td>0</td>
<td>8.42x10⁻⁴</td>
<td>0</td>
<td>6.42x10⁻⁴</td>
<td>0</td>
</tr>
<tr>
<td>PARK2</td>
<td>chr6:161,601,162-163,259,260</td>
<td>Loss or Gain</td>
<td>na</td>
<td>13 (8)</td>
<td>28</td>
<td>1.10x10⁻²</td>
<td>1.96x10⁻²</td>
<td>2.30x10⁻³</td>
<td>1.69x10⁻¹</td>
</tr>
<tr>
<td>GABRB3,GABRA5,GABRG3</td>
<td>chr15:26,996,126-27,220,713</td>
<td>Gain</td>
<td>na</td>
<td>1 (1)</td>
<td>0</td>
<td>8.42x10⁻⁴</td>
<td>0</td>
<td>0</td>
<td>1.18x10⁻³</td>
</tr>
<tr>
<td>MAPT,CRHR1,KANSL1,KANSL1-AS1,MAPT-AS1,MAPT-IT1,SPPL2C,STH</td>
<td>chr17:43,661,362-44,345,063</td>
<td>Gain</td>
<td>na</td>
<td>2 (0)</td>
<td>0</td>
<td>1.68x10⁻³</td>
<td>0</td>
<td>1.03x10⁻⁴</td>
<td>1.18x10⁻³</td>
</tr>
<tr>
<td>APP,ADAMTS1,ADAMTS5,ATP5J,CYRR1,CYRR1-AS1,GABPA,JAM2,MRP39</td>
<td>chr21:25,063,840-28,522,487</td>
<td>Gain</td>
<td>na</td>
<td>1 (0)</td>
<td>0</td>
<td>8.42x10⁻⁴</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>CHCHD10,ADORA2A,ADORA2A-AS1,C22orf15,CABIN1,CESSAP1,DDT,DDTL,DERL3,DRICH1,GGT,GGT5,GSTT1,GSTT1-AS1,GSTT2,GSTT2B,GSTTP1,GSTTP2,GUCD1,GUSBP1,IGL1,LRR7C5B,MIF,MIF-AS1,MMP11,POM121L9P,RGL4,SLC2A11,SMARCB1,SNRPD3,SPECCL1,SUSD2,UPB1,VPREB3,ZDHHC3P1,ZNF70</td>
<td>chr22:23,690,325-25,011,417</td>
<td>Gain</td>
<td>na</td>
<td>1 (1)</td>
<td>0</td>
<td>8.42x10⁻⁴</td>
<td>0</td>
<td>6.00x10⁻⁴</td>
<td>1.18x10⁻³</td>
</tr>
<tr>
<td>CYP2D6,NDUFA6-AS1</td>
<td>chr22:42,522,613-42,531,210</td>
<td>Loss or Gain</td>
<td>na</td>
<td>1 (1)</td>
<td>4</td>
<td>8.42x10⁻⁴</td>
<td>2.80x10⁻³</td>
<td>2.20x10⁻¹</td>
<td>3.00x10⁻²</td>
</tr>
</tbody>
</table>
CNVs identified as significantly associated with DLB in the case-control association analyses and CNVs identified using the candidate CNVs and genes approaches.

All genomic coordinates are for the genome assembly hg19. * Neuropathological diagnosis; na – not applicable; p-value - these were calculated including all cases (not only neuropathologically confirmed cases) # – p-values resulting from the case-control association analysis.
5. References


Genet. 16, 1–14.


