Genome-wide association analysis of Dementia with Lewy bodies reveals unique genetic architecture

Rita Guerreiro[1,2,3]*, Owen A Ross[4]*, Celia Kun-Rodrigues[1,2], Tatiana Orme[1,2], Claire Shepherd[5], Laura Parkkinen[6], Lee Darwent[1,2], Dena Hernandez[7], Sonja Scholz[8,9], Olaf Ansorge[6], Lorraine Clark[10], Lawrence S. Honig[10], Karen Marder[10], Jordi Clarimon[11], Afina Lemstra[12], Philippe Scheltens[12], Wiesje van der Flier[12], Eva Louwersheimer[12], Henne Holstege[12], Ekaterina Rogaeva[13], Peter St. George-Hyslop[14], Elisabet Londos[15], Henrik Zetterberg[16], Imelda Barber[17], Anne Braae[17], Kristelle Brown[17], Kevin Morgan[17], Walter Maetzler[18], Daniela Berg[18], Claire Troakes[19], Safa Al-Sarraj[19], Tammaryn Lashley[20], Janice Holton[20], Yaroslau Compta[21], Vivianna Van Deerlin[22], Geidy E Serrano[23], Thomas G Beach[23], Suzanne Lesage[24], Douglas Galasko[25], Eliezer Masliah[25], Isabel Santana[26], Pau Pastor[27], Pentti J Tienari[28], Liisa Myllykangas[29], Minna Oinas[30], Tamas Revesz[20], Andrew J Lees[20], Brad F Boeve[31], Ronald C Petersen[31], Tanis J Ferman[32], Valentina Escott-Price[33], Neill Graff-Radford[34], Nigel Cairns[35], John C Morris[35], David Stone[36], Stuart Pickering-Brown[37], David Mann[37], Glenda Halliday[5], Andrew Singleton[7], John Q Trojanowski[22], John Hardy[2], Dennis W Dickson[4], Jose Bras[1,2,3,†]

* - Denotes equally contributing authors

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

1- UK Dementia Research Institute 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 - Neuroscience Research Australia, Sydney, Australia and School of Medical Sciences, Faculty of Medicine, University of New South Wales, Sydney, Australia

6 - Nuffield Department of Clinical Neurosciences, Oxford Parkinson's Disease Centre, University of Oxford, Oxford, UK

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

8 - Neurodegenerative Diseases Research Unit, National Institute of Neurological Disorders and

Stroke, National Institutes of Health, Bethesda, Maryland, USA
9 - Department of Neurology, Johns Hopkins University Medical Center, Baltimore, Maryland, USA
10 - Taub Institute for Alzheimer Disease and the Aging Brain and Department of Pathology and Cell Biology, Columbia University, New York, NY, USA
11 - Memory Unit, Department of Neurology, IIB Sant Pau, Hospital de la Santa Creu i Sant Pau, Universitat Autònoma de Barcelona, Barcelona, Spain; Centro de Investigación Biomédica en Red en Enfermedades Neurodegenerativas (CIBERNED), Instituto de Salud Carlos III, Madrid, Spain
12 - Department of Neurology and Alzheimer Center, Neuroscience Campus Amsterdam, VU University Medical Center, Amsterdam, the Netherlands
13 - Department of Medicine, Tanz Centre for Research in Neurodegenerative Diseases, University of Toronto, Toronto, Ontario, Canada
14 - Department of Medicine, Tanz Centre for Research in Neurodegenerative Diseases,University of Toronto, Toronto, Ontario, Canada and Cambridge Institute for Medical Research, Wellcome Trust MRC Building, Addenbrookes Hospital, Hills Road, Cambrdge, UK CB2 0XY
15 - Clinical Memory Research Unit, Institution of Clinical Sciences Malmö, Lund University, Sweden
16 - 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, Mölndal, Sweden
17 - Translation Cell Sciences - Human Genetics, School of Life Sciences, Queens Medical Centre, University of Nottingham, Nottingham, UK
18 - Hertie Institute for Clinical Brain Research, Department of Neurodegeneration, Center of Neurology, University of Tuebingen, Tuebingen, and Department of Neurology, Christian-Albrechts University of Kiel, Germany
19 - Department of Basic and Clinical Neuroscience and Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK
20 - Queen Square Brain Bank, Department of Molecular Neuroscience, UCL Institute of Neurology, London, UK
21 - 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 Clínic, University of Barcelona, IDIBAPS, Barcelona,

Spain
22 - 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
23 - Banner Sun Health Research Institute, 10515 W Santa Fe Drive, Sun City, AZ 85351, USA
24 - Inserm U1127, CNRS UMR7225, Sorbonne Universites, UPMC Univ Paris 06, UMR and S1127, Institut du Cerveau et de la Moelle épinière, Paris, France
25 - Department of Neurosciences, University of California, San Diego, La Jolla, CA, United States; Veterans Affairs San Diego Healthcare System, La Jolla, CA, United States
26 - Neurology Service, University of Coimbra Hospital, Coimbra, Portugal
27 - Memory Unit, Department of Neurology, University Hospital Mutua de Terrassa, University of Barcelona, Terrassa, Barcelona, Spain and Centro de Investigación Biomédica en Red Enfermedades Neurodegenerativas (CIBERNED), Madrid, Spain
28 - Molecular Neurology, Research Programs Unit, University of Helsinki, Department of Neurology, Helsinki University Hospital, Helsinki, Finland
29 - Department of Pathology, Haartman Institute, University of Helsinki and HUSLAB
30 - Department of Neuropathology and Neurosurgery, Helsinki University Hospital and University of Helsinki, Helsinki, Finland
31 - Neurology Department, Mayo Clinic, Rochester, MN, USA
32 - Department of Psychiatry and Department of Psychology, Mayo Clinic, Jacksonville, FL, USA
33 - MRC Centre for Neuropsychiatric Genetics and Genomics, School of Medicine, Cardiff University, Cardiff, UK
34 - Department of Neurology, Mayo Clinic, Jacksonville, FL, USA
35 - Knight Alzheimer's Disease Research Center, Department of Neurology, Washington University School of Medicine, Saint Louis, MO, USA
36 - Genetics and Pharmacogenomics, Merck Research Laboratories, West Point, Pennsylvania, USA
37 - Institute of Brain, Behaviour and Mental Health, Faculty of Medical and Human Sciences, University of Manchester, Manchester, UK

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, due in part, to similarities between DLB, Parkinson's (PD) and Alzheimer's diseases (AD). This overlap complicates clinical care in that an accurate diagnosis is not always straightforward, and suggests that these diseases may share common aetiology. We have recently shown that loci implicated in susceptibility to PD and AD also play a role in DLB and that the proportion of genetic correlation between these diseases is very similar, when the major risk locus, *APOE*, is excluded. These results demonstrate that not only is DLB related to these more common diseases from a purely genetic perspective, but also, that DLB has a strong and quantifiable genetic component.

Methods: Here we have performed the first large-scale genome-wide association study of DLB in a combined cohort of 6,197 samples. We exploited the recently established Haplotype Reference Consortium panel as the basis for imputation to a total of 8.4 million high-quality imputed genotypes and performed independent replication and a meta-analysis of significant and suggestive results.

Findings: Results confirm previously reported associations (*APOE, SNCA, GBA*) and provide genome-wide significant signals for two novel loci (*BCL7C/STX1B* and *CNTN1*), in addition to several loci with suggestive levels of association. Additionally, using the genome-wide SNP data we estimate the heritable component of DLB to be approximately 36%.

Interpretation: These results allow us to start to characterize, for the first time, the role of common genomic variability in DLB. They show unequivocally that common genetic variability plays a role in this disease, that this variability is, to some extent, shared with PD and AD and suggest a unique genetic risk profile in this disease.

Funding

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

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, AD and Parkinson's disease (PD).

So far, no single high penetrant genetic variant has been identified and replicated as a specific cause of DLB, either in families or population-based series. Three major factors may have contributed to this: first, DLB, 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 cases including less than 100 patients. However, it is currently indisputable that DLB has a strong genetic component. The epsilon-4 allele of *APOE* [4,5] is recognized to be a strong risk factor, as are heterozygous mutations and common polymorphisms in the glucocerebrosidase gene (*GBA*)[6]. 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 shown that DLB has a significant heritable component [7].

It has been shown that there is no overlap in common genetic risk between PD and AD [8]. A fact that is not entirely surprising given the obvious 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 5,033 controls. The majority of cases (n=1,324) were neuropathologically assessed, greatly improving diagnostic accuracy. 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 (n=1,743) were diagnosed according to the consensus criteria for either clinical or pathological diagnosis of DLB [9]. The majority of cases were pathologically diagnosed (n=1,324). Cases were included only when the likelihood of a diagnosis of DLB was "Intermediate" or "High". Control subjects (n=5,033) 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)) and the Mayo Clinic Florida control database.

Discovery stage genotyping and quality control

Case subjects (n=1,687) were genotyped in either Illumina Omni2.5M or Illumina OmniExpress genotyping arrays (n=987 and n=700, respectively) (Table 1). Controls (n=4,370) 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).

Country of origin	N	N neuropathological diagnosis	M:F	Mean age at onset	Successfully Genotyped	N neuropathological diagnosis
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	**1687**	**1308**	**1.83**	**70.1**	**1216**	**974**

Table 1: Characteristics of the DLB discovery cohort of DLB. 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.

Population outliers were determined by principal components analysis (PCA), using SNPs passing the aforementioned quality-control filters. After LD-based pruning with version 1.9 of PLINK [10] 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 Principal Components Analysis (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.

Imputation

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[11,12]. Imputation was conducted using the Michigan Imputation Server[13]. Following imputation, variants passing a standard imputation quality threshold (R^2 >= 0.3) were kept for further analysis.

Association tests

We used logistic regression as implemented in PLINK2[10] to test for association of variants with the binary case-control phenotype. To control for population stratification, we used coordinates from the top twenty PCA dimensions as covariates in the regression model. 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. A meta-analysis of stage 1 and 2 was conducted with GWAMA[14] using estimates of the allelic odds ratio and 95% confidence intervals. Gene-wise burden tests were performed using all variants with an effect in protein sequence and a maximum MAF of 5%, using SKAT-O [15,16] as implemented in EPACTS [17].

Replication genotyping

A total of 527 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. 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

Country of origin	N	N neuropathological diagnosis	M:F	Mean age at onset
USA - cases	527	350	2.01	75.5
USA - controls	663	0	0.75	69[a]

Table 2: Characteristics of the replication cohort. [a] Denotes age at examination for controls.

gender.

Phenotypic variance explained

To estimate the phenotypic variance explained by the genotyped SNPs in this cohort we used GREML analysis as implemented in GCTA [18,19]. We used the first ten principal components as covariates and a disease prevalence of 0.1% [20]. 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.

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. After imputation and quality control, genotypes for 8,410,718 variants were available for downstream analyses. QQ plot and genomic inflation factor (λ=1) indicated good control of population stratification (Supplementary Figure 1).

Five regions were associated at genome-wide significance ($p<5x10^{-8}$) (Figure 1; Table 3).

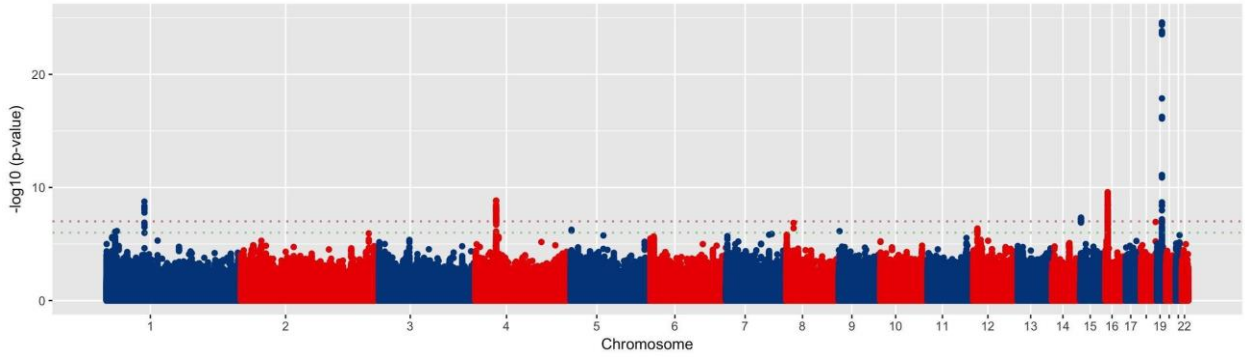

Figure 1: Manhattan plot showing genome-wide p-values of association. The p-values were obtained by logistic regression analysis using the first 20 principal components as covariates. The y-axis shows −log10 p-values of 8,410,718 SNPs, and the x-axis shows their chromosomal positions. The y-axis was 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.

These included the previously described loci *APOE* (rs429358, OR=2.4, $p=5.31x10^{-50}$), *SNCA* (rs7681440, OR=0.7412, $p=1.45x10^{-9}$) and *GBA* (rs35749011, OR=2.5, $p=1.77x10^{-9}$). Additionally, loci overlapping *BCL7C/STX1B* (rs897984, OR=0.72, $p=2.64x10^{-10}$) and *GABRB3*

(rs1426210, OR=1.32, p=4.63x10^{-8}) were also genome-wide significant. Two additional regions surpassed a suggestive threshold of significance (p=1x10^{-6}), the first overlapping the *SOX17* gene (rs139919032, OR=2.4, p=1.37x10^{-7}) and the second overlapping the *CNTN1* gene (rs79329964, OR=1.5, p=4.35x10^{-7}). Stage 2 of the GWAS design provided independent replication for 4 of the loci. The associations at *GABRB3*, *BCL7C*/*STX1B* and *SOX17* were not replicated. However, meta-analysis of both stages showed the *BCL7C*/*STX1B* association to survive genome-wide multiple test correction (p=1.19x10^{-8}; Table 3).

Named Region	CHR	Position	Variant	Discovery					Replication				Meta-Analysis					
				R2	OR	L95	U95	P-value	OR	L95	U95	P-value	OR	L95	U95	P-value	N	effects
APOE	19	45411941	rs429358	0.949	2.41	2.14	2.7	5.31E-50	2.74	2.15	3.49	4.00E-16	2.46	2.22	2.74	3.31E-64	6197	++
BCL7C/STX1B	16	30886643	rs897984	0.984	0.73	0.66	0.8	2.64E-10	0.98	0.81	1.19	8.30E-01	0.77	0.71	0.85	1.19E-08	6197	--
SNCA	4	90756550	rs7681440	0.996	0.74	0.67	0.82	1.45E-09	0.68	0.56	0.82	6.00E-05	0.73	0.67	0.79	9.22E-13	6197	--
GBA	1	155121143	rs35682329	0.957	2.43	1.81	3.27	4.33E-09	1.81	1.05	3.11	3.30E-02	2.27	1.75	2.95	6.57E-10	6197	++
GABRB3	15	26840998	rs1426210	0.982	1.32	1.2	1.46	4.63E-08	0.84	0.68	1.04	1.00E-01	1.22	1.11	1.33	2.05E-05	6197	+-
SOX17	8	55395693	rs144770207	0.937	2.44	1.73	3.44	4.02E-07	0.41	0.19	0.86	1.90E-02	1.81	1.32	2.48	2.23E-04	6197	+-
CNTN1	12	41179589	rs79329964	0.993	1.54	1.3	1.81	4.35E-07	1.54	1.04	2.28	3.30E-02	1.54	1.32	1.79	3.99E-08	6197	++

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. CHR: Chromosome. R2: Imputation R-squared of each specific variant. OR: Odds ratio. L95: Lower 95% interval. U95: Upper 95% confidence interval. N: Total number of samples. Effects: direction of association

The association observed at the *SNCA* locus represents an independent signal when compared to the top association reported for PD. In an attempt to dissect the differential association between the two diseases, we used GTEx data to try to determine if the top SNPs for each disease act as eQTLs. 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 2). These results are compatible with a model in which rs7681440 genotypes influence the expression levels of *SNCA* 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*

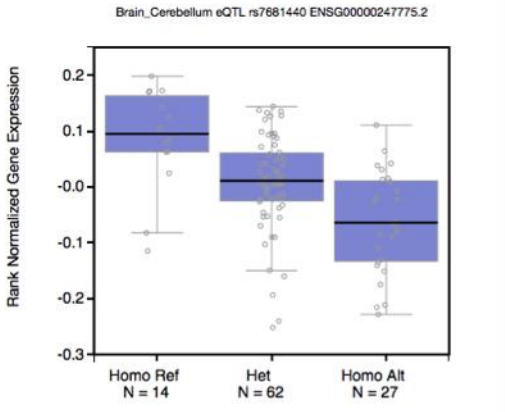

Figure 2: Boxplot showing the association between rs7681440 genotypes and RP11-67M1.1 expression in the cerebellum in 103 healthy post-mortem samples (p=2.00E-07) from the GTEx Consortium. Carriers of the GG genotype (alternative allele) show the lowest levels of expression of the gene. Medians, interquartile ranges and individual data points are indicated.

transcription (higher *SNCA* expression), which is in accordance with a higher frequency of the alternative allele in cases when compared to controls.

A systematic assessment of genetic loci previously associated with AD or PD showed no evidence of other significant associations in this DLB cohort. This includes the *TREM2* locus, where the p.R47H has been shown to have a strong effect in AD. In our cohort the p.R47H variant did not show genome-wide significant levels of association (p=0.002).

Gene burden analysis

Gene-wide burden based 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* ($p=1.29x10^{-13}$). No other gene showed evidence of strong association with disease or overlap with single variant analysis (Table 4).

Position_ID	GENE	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

Table 4: Gene burden results, showing the most significant genes. 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.

Estimation of heritability of DLB

Using the first ten principal components as covariates and a disease prevalence of 0.1%, the estimation of the phenotypic variance attributed to genetic variants showed a heritable component of DLB of 36% (± 0.03). We have also partitioned the heritability across the genome, using each chromosome as a unit. We applied linear regression to determine the relationship between heritability and chromosome length. The results are presented in Figure 3. We found a strong correlation between chromosome length and heritability (Pearson correlation r = 0.745, p-value = $6.87x10^{-5}$).

Interestingly the heritability 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 than expected, while none of them have variants with genome-wide significant results.

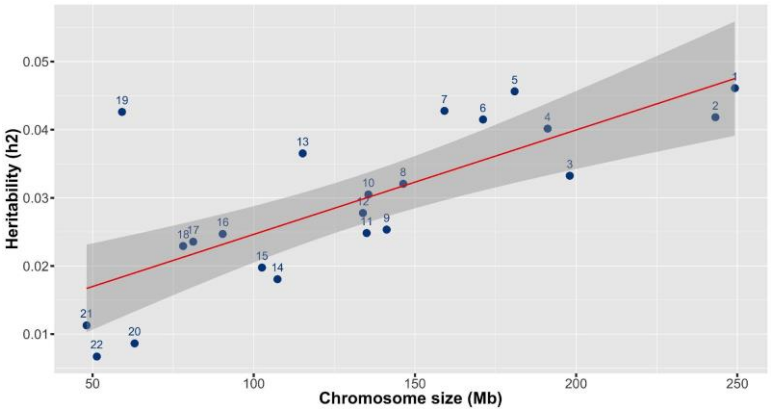

Figure 3: DLB heritability by chromosome. Heritability (y-axis) per chromosome is plotted against chromosome length (x-axis). The red line represents heritability regressed on chromosome length and the shaded grey area represents the 95% confidence interval of the regression model.

Discussion

This is the first comprehensive unbiased study of common genetic variability in DLB. We identified five genome-wide significant associations (APOE, *BCL7C*/*STX1B, SNCA, GBA*, and, *CNTN1*), and two loci with suggestive levels of association in the discovery phase, that did not replicate in the smaller second stage (*GABRB3* and *SOX17*). In both of these cases the regions showed effect heterogeneity, suggesting that it is possible that a larger replication series would lead to a different outcome.

The most significant association signal is observed at the *APOE* locus (*APOE E4*) which has been previously shown to be highly associated with DLB [4,5]. 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 [21].

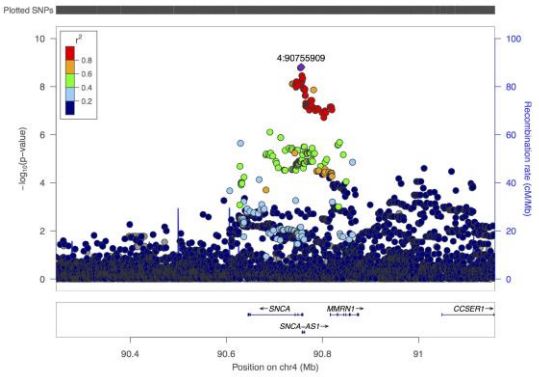

Figure 4: Regional association plot for the SNCA locus. Purple represents rs1372517, which is the most associated SNP at the locus also present in the 1000Genomes dataset. The variant rs1372517 is in complete LD with rs7681440. Colours represent LD derived from 1000Genomes between each variant and the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

The second strongest association is observed at the *SNCA* locus and we were able to confirm the different association profile between DLB and PD that we had previously reported [5]. *SNC*A is the most significant common genetic risk factor for PD, with rs356182 having a meta-analysis p-value of 1.85×10^{-82} (OR:1.34 [1.30-1.38]) in PDGene. This variant is located 3' to the gene [22], while in DLB no association can be found in that region (Figure 4). Additionally, the

most associated SNP reported here for the *SNCA* locus (rs7681440) has a 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, OR:0.841, p-value=$7.09x10^{-19}$). It is tempting to speculate that these differences may reflect pathobiological differences between both diseases, perhaps mediated by differential regulation of gene expression. We show that the top DLB hit is an eQTL in the cerebellum for a *SNCA* antisense gene, however, further investigation of the identified significant eQTL is needed: the effect was observed for only one brain region, even though other regions are present in the GTEx dataset, many with similar sample sizes, and including regions preferentially affected by Lewy body pathology (substantia nigra, frontal cortex, caudate). Nonetheless, it is interesting to note that the effect fits 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 [6]. 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).

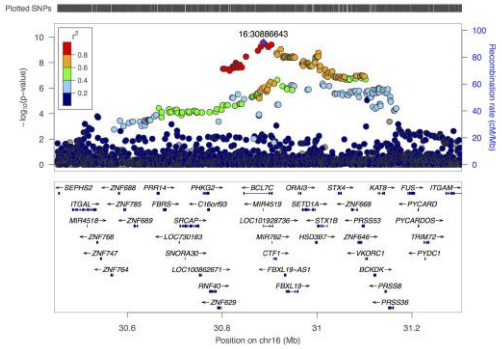

Figure 5: Regional association plot for the BCL7C/STX1B locus. Purple represents the most associated SNP. Colours represent LD derived from 1000Genomes between each variant and the most associated SNP.

An association at the *BCL7C/STX1B* locus has been previously reported for PD [22,23]. 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) and a meta-analysis p-value of $1.34x10^{-5}$ (data from PDgene). 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 [24]. 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 3). Mutations in *STX1B* have recently been shown to cause fever associated epileptic syndromes [25] and myoclonic astatic epilepsy [26].

The *CNTN1* locus has been previously associated with PD in a genome-wide study of IBD segments in an Ashkenazi cohort [27], and with cerebral amyloid deposition, assessed with PET imaging in *APOE E4* non-carriers [28]. This locus was also shown to be sub-significantly associated with clinico-pathologic AD dementia [29]. 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 [30,31]. Mutations in *CNTN1* were found to cause a familial form of lethal congenital myopathy [32]. Contactin 1 drives Notch signalling activation and modulates neuroinflammation events, possibly participating in the pathogenesis of Multiple Sclerosis and other inflammatory disorders [33]. A functional protein association network analysis of *CNTN1* using STRING shows it is in the same network as *PSEN2* (Supplementary Figure 4), supporting its potential role in neurodegeneration. However it is 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.G2019S (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.G2019S directly, the variant was well imputed (R^2=0.94) and showed a higher frequency in cases when compared to controls (0.004 and 0.0005 respectively).

In addition to performing a GWAS with clinico-pathologic AD dementia, Beecham and colleagues [29] 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 a study-wide association at the *SCARB2* locus with DLB [5]. 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 the recent GWAS meta-analyses showed a genome-wide significant association with DLB. Variants at the following loci showed nominal association levels: *MAPT*, *BIN1*, *GAK*, *HLA-DBQ*, *CD2AP*, *INPP5D*, *ECHDC3* and *SCIMP* (Supplementary Table 1).

Two loci (*SOX17* and *GABRB3*) were significantly associated with DLB in the discovery phase, but these associations did not survive replication when analysing an independent, albeit smaller, cohort. However, it is interesting to note that mutations in *GABRB3* have been associated with a broad phenotypic spectrum of epilepsies [34], similar to what has been reported for *STX1B* [25,26]. Genetic variability at *GABRB3* has also been reported to be significantly associated with hallucinations and delusions in cohorts of schizophrenia spectrum disorders [35,36], traits that could be seen as having parallels in DLB. Even though this is the largest cohort of DLB cases studied so far in a genome-wide manner, it is only a moderately sized dataset, suggesting that it is plausible that the study of larger numbers of samples would provide enough statistical power to confidently identify variants such as these.

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 [37]. This shows that, despite not having any causative genes identified so far, genetics plays a relevant role in the common forms of DLB. Additionally, we provide evidence suggesting that novel associated 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. This greatly reduces the chances of misdiagnosis, and consequently improves power to detect associations, particularly when compared with clinical cohorts. 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 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 strong 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 and AS designed the study. JB, AS, DS, and OAR obtained funding for the study.

JB, RG, OAR, CKR, LD and DH performed data acquisition. JB, RG, OAR, and CKR analysed and interpreted the data. CS, LP, SS, OA, JC, LC, LH, KM, AL, PS, WvdF, EL, HH, ER, PGH, EL, HZ, IB, AB, KB, KM, WM, 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

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. 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. 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, Canadian Institutes of Health Research, and Ontario Research Fund. 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). The authors would like to thank the Exome Aggregation Consortium and the groups that provided exome variant data for comparison. A full list of contributing groups can be found at http://exac.broadinstitute.org/about. 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 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.

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

5 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.

6 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.

7 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.

8 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.

9 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.

10 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.

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

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

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

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

15 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.

16 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.

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

18 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.

19 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.

20 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.

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

22 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.

23 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.

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

25 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.

26 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.

27 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.

28 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.

29 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.

30 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.

31 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.

32 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.

33 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.

34 Møller RS, Wuttke TV, Helbig I, *et al.* Mutations in GABRB3: From febrile seizures to epileptic encephalopathies. *Neurology* 2017; **88**: 483–92.

35 Sun J, Jayathilake K, Zhao Z, Meltzer HY. Investigating association of four gene regions (GABRB3, MAOB, PAH, and SLC6A4) with five symptoms in schizophrenia. *Psychiatry Res* 2012; **198**: 202–6.

36 Bergen SE, Fanous AH, Walsh D, O'Neill FA, Kendler KS. Polymorphisms in SLC6A4, PAH, GABRB3, and MAOB and modification of psychotic disorder features. *Schizophr Res* 2009; **109**: 94–7.

37 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.