Genome-wide association study of REM sleep behavior disorder in Parkinson’s disease

Running head: Genetics of REM sleep behavior disorder

Yuri L. Sosero, MD,1,2 Karl Heilbron, PhD,3 Pierre Fontanillas, PhD,3 Lucy Norcliffe-Kaufmann, PhD,3 Eric Yu, MSc,1,2 Uladzislau Rudakou, MSc,1,2 Jennifer A. Ruskey, MSc,2 Kathryn Freeman, BSc,2 Farnaz Asayesh, MSc,2 Kajsa Brolin, MSc,4 Maria Swanberg, PhD,4 Huw R Morris, FRCP PhD,5,6 Lesley Wu, MSc,5,6 Raquel Real, MD, PhD,5,6 Lasse Pihlstrøm, MD, PhD,7 Manuela Tan, PhD,7 Thomas Gasser, MD, PhD,8 Kathrin Brockmann, MD, PhD,8 Hui Liu, PhD,8 Michele T.M. Hu, MBBS, FRCP, PhD,9,10 Donald G. Grosset, MD,11 Simon J.G. Lewis, MD,12 John B. Kwok, PhD,12 Pau Pastor, MD, PhD,13,14 Ignacio Alvarez, MSc,13,14 Matej Skorvanek, MD, PhD,15,16 Alexandra Lackova, PhD,15,16 Miriam Ostrozovicova, MD,15,16 Mie Rizig, MD, PhD,17 23andMe Research Team,3 The International Parkinson’s Disease Genomics Consortium, Lynne Krohn, PhD,1,2 Ziv Gan-Or, MD, PhD,1,2

Affiliations

1. Department of Human Genetics, McGill University, Montréal, QC, Canada.
2. The Neuro (Montreal Neurological Institute-Hospital), McGill University, Montréal, QC, Canada.
3. 23andMe, Inc., Sunnyvale, California, USA.
4. Lund University, Translational Neurogenetics Unit, Department of Experimental Medical Science, Lund, Sweden.
5. Department of Clinical and Movement Neurosciences, UCL Queen Square Institute of Neurology, University College London, UK
6. UCL Movement Disorders Centre, University College London, London, UK
7. Department of Neurology, Oslo University Hospital, Oslo, Norway
8. Department of Neurodegeneration at Hertie Institute for Clinical Brain Research, Tuebingen, Germany
11. Institute of Neurological Sciences, Queen Elizabeth University Hospital, Glasgow, United Kingdom
12. Parkinson’s Disease Research Clinic, Brain and Mind Centre, School of Medical Sciences, University of Sydney, Australia
13. Unit of Neurodegenerative diseases, Department of Neurology, University Hospital Germans Trias i Pujol and The Germans Trias i Pujol Research Institute (IGTP) Badalona, Barcelona, Spain
14. Department of Neurology, Hospital Universitari Mutua de Terrassa, Barcelona, Spain
15. Department of Neurology, Pavol Jozef Šafárik University in Košice, Slovakia

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
Corresponding author:

Ziv Gan-Or
Montreal Neurological Institute,
McGill University
1033 Pine Avenue, West,
Ludmer Pavilion, room 312,
Montréal, QC, H3A 1A1
Phone: +1-514-398-5845,
E-mail: ziv.gan-or@mcgill.ca

Keywords: REM sleep behavior disorder, Parkinson’s disease, genetics, SNCA, LRRK2, cognitive decline

Number of characters in the title and running head: 83 and 29

Number of words in the abstract, and the body of the manuscript: 233 and 3573

Number of figures: 2 main figures and 3 supplementary figures

Number of tables: 3 main tables and 7 supplementary tables
Abstract

Objective: REM sleep behavior disorder (RBD) is a prodromal synucleinopathy, reported in a subset of Parkinson’s disease (PD) patients, and associated with neuropsychiatric symptoms in PD. We aimed to compare the genetic background of PD patients with probable RBD (PD+RBD) and PD patients without probable RBD (PD-RBD). Furthermore, we examined genetic correlations and potential causal associations between multiple neuropsychiatric traits and PD+RBD.

Methods: We performed a genome-wide association study (GWAS) including 5,403 PD+RBD and 13,020 PD-RBD. To test for genetic correlations and potential causal associations between neuropsychiatric traits and PD+RBD, we used linkage disequilibrium score regression and Mendelian randomization.

Results: The SNCA locus was associated with PD+RBD compared to PD-RBD (rs10005233, OR=1.21, 95% CI=1.16-1.27, p=1.81e-15). Further examination of known genetic loci associated with PD from the most recent PD GWAS in Europeans and Asians identified additional variants associated with reduced risk for PD+RBD: two in the SNCA locus (rs5019538-G, OR=0.85, 95% CI=0.81-0.89, p=2.46E-10; rs356182-G, OR=0.89, 95% CI=0.84-0.95, p=0.0001), and one in the LRRK2 locus (rs34637584, p.G2019S, OR=0.41, 95% CI=0.28-0.61, p=1.04E-5). We found a potential genetic correlation between attention deficit hyperactivity disorder (ADHD) and PD+RBD, which was not statistically significant after correction for multiple comparisons. No causative association emerged between PD and neuropsychiatric traits.

Interpretation: Genetic variants contribute to the occurrence of RBD in PD, further distinguishing between the PD+RBD and PD-RBD subtypes. Understanding the mechanisms underlying these genetic associations could contribute to the development of subtype-specific treatments.
INTRODUCTION

Rapid-eye-movement (REM) sleep behavior disorder (RBD) is a parasomnia characterized by the absence of muscle atonia during REM sleep and dreams enactment.\(^1\) When no neurological conditions or other concomitant factors are identified, it is referred to as isolated/idiopathic RBD (iRBD).\(^2\) iRBD is typically considered a prodromal stage of synucleinopathies, as about 80%-90% of the cases convert to either Parkinson’s disease (PD), dementia with Lewy bodies (DLB) or, more rarely, multiple system atrophy (MSA).\(^3\), \(^4\) These disorders are all characterized by the accumulation of alpha-synuclein, encoded by the \(SNCA\) gene.\(^5\) RBD is therefore a key prodromal clinical marker of synucleinopathies, and its presence is also associated with a distinctive, more severe clinical presentation. In PD patients with RBD (approximately 25%-58% of cases\(^6\)), RBD is associated with a more malignant phenotype, characterized by faster progression\(^7\) and greater frequency and/or severity of neuropsychiatric manifestations, including cognitive decline, hallucinations, depression, anxiety and apathy.\(^8\)-\(^11\)

In recent years, it was shown that the genetic background of iRBD only partially overlaps with that of PD or DLB. Genes such as \(GBA1\), \(^12\) \(TMEM175\)\(^13\) and \(SNCA\)\(^14\) are important across all conditions,\(^15\), \(^16\) whereas other genes including \(LRRK2\), \(^17\) \(APOE\)\(^18\) and familial PD genes,\(^19\) seem to not have a major role in iRBD. A recent RBD genome-wide association study (GWAS) identified 5 risk loci associated with RBD,\(^20\) namely \(GBA1\), \(TMEM175\), \(INPPSF\), \(SNCA\) and \(SCARB2\). Notably, the variants associated with RBD in the \(SNCA\) and \(SCARB2\) regions were different and independent to those associated with PD,\(^15\), \(^20\) supporting RBD as a distinctive subtype, with specific genetic and clinical correlates.

In the current study, we aimed to examine whether there are differences in the genetics of PD with probable RBD (PD+RBD) compared to PD without RBD (PD-RBD). For this purpose, we performed a GWAS including 18,423 patients, composed of PD+RBD patients (N=5,403), and PD-RBD patients (N=13,020). To further explore the relationships between RBD and neuropsychiatric...
manifestations in PD, we performed genetic correlation and Mendelian randomization (MR) analyses using the GWAS summary statistics of the current study and multiple neuropsychiatric conditions.

METHODS

Population

The study population included 18,423 PD patients (detailed in Table 1), of whom 5,403 had probable RBD (PD+RBD) and were treated as cases, whereas 13,020 did not (PD-RBD) and were treated as controls. Probable RBD was defined using either the RBD single-question screen (RBD1Q)\textsuperscript{21} or the RBD screening questionnaire (RBDSQ),\textsuperscript{22} both of which show high sensitivity and specificity in PD patients.\textsuperscript{23} We refer to iRBD when RBD occurs prior to the neurodegeneration and to RBD for subjects with RBD regardless of the time of onset of neurodegeneration. PD was diagnosed by movement disorder specialists according to the UK Brain Bank\textsuperscript{24} or International Parkinson Disease and Movement Disorders Society criteria.\textsuperscript{25} The 23andMe cohort had self-reported a diagnosis of PD as well as RBD and/or dream re-enactment behaviors.

The participants were of European ancestry and their clinical and genetic data were collected from 15 different cohorts (Table 1), 11 of which are from the International Parkinson’s Disease Genomics Consortium (IPDGC), three cohorts are from the Accelerating Medicines Partnership Parkinson's disease (AMP-PD, https://amp-pd.org/) and one cohort was collected and analyzed by 23andMe Inc. (https://www.23andme.com/research/). The Central European Group on Genetics of Movement Disorders (CEGEMOD) contributed to the Kosice cohort.

Genetic analysis

23andMe
Participants provided informed consent and volunteered to participate in the research online, under a protocol approved by the external AAHRPP-accredited IRB, Ethical & Independent (E&I) Review Services. As of 2022, E&I Review Services is part of Salus IRB (https://www.versiticlinicaltrials.org/salusirb). DNA extraction and genotyping were performed on saliva samples by National Genetics Institute (NGI), a CLIA-licensed clinical laboratory and a subsidiary of Laboratory Corporation of America. Samples were genotyped on one of five genotyping platforms. The v1 and v2 platforms were variants of the Illumina HumanHap550 + BeadChip, including about 25,000 custom single nucleotide polymorphisms (SNPs) selected by 23andMe, with a total of about 560,000 SNPs. The v3 platform was based on the Illumina OmniExpress+ BeadChip, with custom content to improve the overlap with the 23andMe v2 array, with a total of about 950,000 SNPs. The v4 platform was a fully customized array, including a lower redundancy subset of v2 and v3 SNPs with additional coverage of lower-frequency coding variation, and about 570,000 SNPs. The v5 platform, in current use, is an Illumina Infinium Global Screening Array (~640,000 SNPs) supplemented with ~50,000 SNPs of custom content. This array was specifically designed to better capture global genetic diversity and to help standardize the platform for genetic research. Samples that failed to reach 98.5% call rate were re-analyzed. Individuals whose analyses failed repeatedly were re-contacted by 23andMe customer service to provide additional samples.

Participants were restricted to European ancestry through an analysis of local ancestry. A support vector machine (SVM) to classify individual haplotypes into one of 31 reference populations was used (https://www.23andme.com/ancestry-composition-guide/). The SVM classifications are then fed into a hidden Markov model (HMM) that accounts for switch errors and incorrect assignments, and gives probabilities for each reference population in each window. Finally, we used simulated admixed individuals to recalibrate the HMM probabilities so that the reported assignments are consistent with the simulated admixture proportions. A maximal set of unrelated individuals was chosen for each analysis using a segmental identity-by-descent (IBD) estimation algorithm.
We phased participant data using either an internally-developed tool, Finch (V1-V4 genotyping arrays) or Eagle2 (V5 genotyping array).\textsuperscript{28} Finch implements the Beagle haplotype graph-based phasing algorithm, modified to separate the haplotype graph construction and phasing steps.\textsuperscript{29} It extends the Beagle model to accommodate genotyping error and recombination, to handle cases where there are no consistent paths through the haplotype graph for the individual being phased. We constructed haplotype graphs for European and non-European samples on each 23andMe genotyping platform from a representative sample of genotyped individuals, and then performed out-of-sample phasing of all genotyped individuals against the appropriate graph. For the X-chromosome, we built separate haplotype graphs for the non-pseudoautosomal region and each pseudoautosomal region, and these regions were phased separately.

Imputation panels created by combining multiple smaller panels have been shown to give better imputation performance than the individual constituent panels alone.\textsuperscript{30} To that end, we combined the May 2015 release of the 1000 Genomes Phase 3 haplotypes with the UK10K imputation reference panel to create a single unified imputation reference panel.\textsuperscript{31, 32} Multiallelic sites with \( N \) alternate alleles were split into \( N \) separate biallelic sites. We then removed any site whose minor allele appeared in only one sample. For each chromosome, we used Minimac3 to impute the reference panels against each other, reporting the best-guess genotype at each site.\textsuperscript{33} This gave us calls for all samples over a single unified set of variants. We then joined these together to get, for each chromosome, a single VCF with phased calls at every site for 6,285 samples.

In preparation for imputation, we split each chromosome of the reference panel into chunks of no more than 300,000 variants, with overlaps of 10,000 variants on each side. We used a single batch of 10,000 individuals to estimate Minimac3 imputation model parameters for each chunk.\textsuperscript{33} We imputed phased participant data against the chunked merged reference panel using Minimac3, treating males as homozygous pseudo-diploids for the non-pseudoautosomal region. Throughout, we treated structural variants and small indels the same as SNPs.
We excluded SNPs that: 1) had a call rate < 90%, 2) had a Hardy-Weinberg p < 10–20 in people with predominantly European ancestry, 3) were only genotyped on the V1 and/or V2 platforms, 4) were found on the mitochondrial chromosome or the Y-chromosome, 5) failed a test for parent-offspring transmission (specifically, we regressed the child’s allele count against the mean parental allele count and excluded SNPs with fitted < 0.6 and p < 10–20 for a test of < 1), 6) had an association with genotype date (p < 10–50 by ANOVA of SNP genotypes against a factor dividing genotyping date into 20 roughly equal-sized buckets), 7) had a large sex effect (ANOVA of SNP genotypes, r² > 0.1), or 8) had probes matching multiple genomic positions in the reference genome.

We excluded SNPs with imputed r² < 0.3, as well as SNPs that had strong evidence of a platform batch effect. For each SNP we identified the largest sub-set of the data passing other quality control criteria based on their original genotyping platform – either v2+v3+v4+v5, v4+v5, v4, or v5 only – and computed association test results for the largest passing set. The batch effect test is an F test from an ANOVA of the SNP dosages against a factor representing the V4 or V5 platform; we excluded results with p < 10–50.

Across both genotyped and imputed GWAS results, we excluded SNPs that had sample size of less than 20% of the total GWAS sample size. We also removed SNPs that did not converge during logistic regression, as identified by abs (effect)>10 or stderr > 10 on the log-odds scale. If SNPs were both genotyped and imputed, and they passed QC for both, we used results from the imputed analysis. After quality control, we had analysed 904,040 genotyped SNPs and 25,208,208 imputed SNPs.

GWAS was performed using logistic regression adjusted for age, sex, top five principal components as well as the genotype platform to account for genotype batch effects. The significance threshold was set at p<5x10E-8.

Other cohorts
Genotyping in the different centers was performed using the OmniExpress, NeuroX or Global Screening (GSA) GWAS array according to the manufacturer’s instructions (Illumina Inc.). Parkinson's Families Project (PFP) was genotyped with NeuroChip, Parkinson's repository of biosamples and networked datasets (PRoBaND) with HumanCoreExome array, and Oxford Parkinson's Disease Centre (OPDC) with either HumanCoreExome-12 v.1.1 or Infinium HumanCoreExome-24v.1.1 arrays. Quality control was performed following standard pipelines (detailed in https://github.com/neurogenetics/GWAS-pipeline) using plink 1.9. In brief, we filtered out heterozygosity outliers using an F-statistic cut-off of <0.15 or >0.15. Individuals with a variant call rate <95% and sex mismatch were excluded. Variants missing in >5% of the participants, with disparate missingness between cases and controls (p<1E-04), or significantly deviating from the Hardy-Weinberg equilibrium in controls (p<1E-04) were also removed. We used GCTA to check for relatedness closer than first cousins between participants (pihat>0.125). We performed imputation using the Michigan imputation server (https://imputationserver.sph.umich.edu/index.html#) with the Haplotype Reference Consortium reference panel r1.1 2016 under default settings. Ancestry outliers were detected using HapMap3 principal component analysis (PCA) data in R version 4.0.1. After imputation, we selected variants with R²>0.8 and a minor allele frequency (MAF)>0.01, while retaining variants that have strong pathogenic implications in PD (i.e., the LRRK2 p.G2019S variant and the GBA1 p.N370S, p.E326K and p.T369M variants).

To test for genetic associations to RBD in PD, we performed GWAS using logistic regression comparing PD+RBD and PD-RBD adjusted for age at RBD questionnaire administration, sex and principal components. The significance threshold was set at p<5x10E-8. The analyses were performed separately in each cohort and the results were then meta-analyzed with a fixed-effect model using METAL (https://genome.sph.umich.edu/wiki/METAL_Documentation). To identify any possible secondary associations hidden by the principal signals of the GWAS, we also performed Conditional and Joint - Genome-wide Complex Trait Analysis (COJO-GCTA), a method that
harnesses a conditional stepwise regression approach to identify independent associations (https://yanglab.westlake.edu.cn/software/gcta/#Overview).  

**Genetic correlation**

To investigate the potential genetic correlation between the presence of RBD in PD and known neuropsychiatric conditions we used linkage-disequilibrium score regression (LDSC) on LDHub (http://ldsc.broadinstitute.org/ldhub/). The neuropsychiatric traits we analyzed include epilepsy, headache, amyotrophic lateral sclerosis, cognitive decline, Alzheimer's disease, Parkinson's disease, dementia with Lewy bodies, alcohol dependence, cannabis dependence, attention deficit hyperactivity disorder, Tourette syndrome, anorexia nervosa, post-traumatic syndrome, schizophrenia, bipolar disorder, obsessive-compulsive disorder, autism spectrum disorder and major depressive disorder. Summary statistics for the compared traits were accessed through the LDHub platform or downloaded from publicly available sources, then formatted and analyzed using LDHub python v2.7 scripts (https://github.com/bulik/ldsc/wiki). Positive correlation indicates association with PD+RBD, and negative correlation indicates association with PD-RBD.

**Mendelian randomization**

To assess any possible causal association between neuropsychiatric disorders and the presence of RBD in PD we performed Mendelian randomization (MR). In brief, this method harnesses summary statistics from an exposure (the neuropsychiatric traits, in this case) and an outcome (the presence of RBD in PD) and uses the statistically significant variants from the former as instrumental variables (IVs) to infer a potential causative association with the latter. This approach mimics randomized control trials, since genetics is randomly assigned at conception and unaffected by the environment. The neuropsychiatric traits for this analysis were selected based on relevance to RBD comorbidities, known neuropsychiatric manifestations in PD or with clinical relevance to PD. They include Alzheimer’s disease, dementia with Lewy Bodies, schizophrenia, major depressive disorder and
bipolar disorder. We used the TwoSampleMR R package (https://mrcieu.github.io/TwoSampleMR/) to perform MR analyses, including sensitivity analyses, tests assessing pleiotropy and heterogeneity between IVs, in R version 4.0.1 according to protocols previously established. Sensitivity analyses included MR Egger, inverse variance weighted (IVW), weighted median, simple mode and weighted mode. Steiger filtering was also performed to check for reverse causality. Summary statistics were downloaded by the MRBase GWAS catalog (http://www.mrbase.org/) and the Psychiatric Genomics Consortium (https://pgc.unc.edu/) publicly available database. To calculate the power to detect an odds ratio=1.2 we used an online Mendelian Randomization power calculation tool (https://sb452.shinyapps.io/power).42

RESULTS

Genome-wide association study identifies the SNCA and LRRK2 loci as modifiers of risk for RBD in PD

To assess whether genetics can affect the risk of RBD in PD we performed GWAS between PD+RBD (N=5,403) and PD-RBD (N=13,020). We evaluated the genomic inflation using quantile-quantile plots (Q-Q plots) and the lambda factor, showing no inflation (lambda=0.994, lambda1000=0.999), (Supplementary Fig 1).

We found that rs10005233, in the 5' region of the SNCA locus, was associated with PD+RBD (OR=1.21, 95% CI=1.16-1.27, p=1.81E-15, Fig 1). No secondary signal was detected in the GCTA-COJO analysis at a GWAS significance level. We also examined the 92 variants associated with PD in the most recently published GWAS in Europeans15 and Asians43 (Table 2, Supplementary Table 1). Using Bonferroni correction based on the number of these variants (α/number of variants=0.00054), we identified three associations. Two were variants in the SNCA locus, independent of each other, whose minor alleles were associated with decreased risk for PD+RBD
(rs5019538-G, OR=0.85, 95% CI=0.81-0.89, p=2.46E-10 and rs356182-G, OR=0.89, 95% CI=0.84-0.95, p=0.0001), and one was the LRRK2 p.G2019S variant, also associated with a reduced risk for PD+RBD (rs34637584, OR=0.41, 95% CI=0.28-0.61, p=1.04E-5, the carrier frequency for this variant across the different cohorts is detailed in Table 3). These three variants were associated with increased risk for PD in the most recent GWAS.\textsuperscript{15, 43} GBA1 variants did not show significant associations with PD+RBD (Supplementary Table 2). Additional potential associations in the \textit{SETD1A}, \textit{SPPL2B}, \textit{CRHR1} and \textit{LINC00693} loci should be further studied (Table 2).

\textbf{Genetic correlation and causative associations between PD with RBD and neuropsychiatric disorders}

To examine potential genetic correlations between the risk of RBD in PD and multiple neuropsychiatric conditions, we performed LDSC (Fig 2, Supplementary Table 3). We found that the PD+RBD trait was mildly correlated with attention deficit hyperactivity disorder (ADHD, rg=0.30, SE=0.14, p=0.04). The most recently published European PD GWAS was genetically correlated with PD-RBD (rg=-0.38, SE=0.15, p=0.01). However, these correlations were not statistically significant after Bonferroni correction (\(\alpha=0.0025\)).

To assess possible causative associations between neuropsychiatric conditions and PD+RBD we performed MR using neuropsychiatric disorders as exposures and PD+RBD as the outcome (Supplementary Figs 2 and 3, Supplementary Tables 4-7). No test showed a statistically significant causative association between neuropsychiatric traits and PD+RBD. However, our power for this analysis was suboptimal (35.7%), therefore there could be associations that we could not detect. We were not able to conduct reverse MR using PD+RBD as the exposure since only one locus passed GWAS significance, preventing us from performing appropriate sensitivity analyses.

\textbf{DISCUSSION}
In the current GWAS, we found that variants in the \textit{SNCA} and \textit{LRRK2} loci may modify the risk of RBD in PD. Additional loci (\textit{SETD1A}, \textit{SPPL2B}, \textit{CRHR1} and \textit{LINC00693}) require further studies to examine whether they have a role in PD+RBD. The top variant in the \textit{SNCA} locus, rs10005233, was previously reported to be associated with iRBD in a candidate gene study.\textsuperscript{14} Another study, using the Oslo and Parkinson’s Progression Marker Initiative cohorts, found another variant in the \textit{SNCA} locus associated with PD+RBD (rs3756063), which is in strong linkage disequilibrium (LD) with rs10005233 (D’=0.97, r\textsuperscript{2}=0.91).\textsuperscript{44} Furthermore, rs10005233 is in LD with other 5’ region \textit{SNCA} variants associated with synucleinopathies, including rs7681440 (D’=0.99, r\textsuperscript{2}=0.94), associated with DLB,\textsuperscript{45} rs763443 (D’=0.89, r\textsuperscript{2}=0.78), a secondary PD GWAS signal,\textsuperscript{14, \textit{15}, 46} as well as rs2583988 (D’=0.99, r\textsuperscript{2}=0.40), a variant located in the \textit{SNCA-AS1} region (discussed below) and associated with Lewy body variant of Alzheimer’s disease (ADLBV).\textsuperscript{47} It is still unclear whether it is a specific variant in the \textit{SNCA} locus or the presence of a specific \textit{SNCA} haplotype that drives these associations with cognitive phenotypes across synucleinopathies.\textsuperscript{14} The rs10005233 variant is also in LD (D’=0.97, r\textsuperscript{2}=0.91) with the top signal of a recently published RBD GWAS, rs3756059,\textsuperscript{20} which was associated with reduced expression of \textit{SNCA-AS1}, an antisense RNA molecule that could potentially reduce the translation of alpha-synuclein when it is overexpressed or increase the translation of alpha-synuclein when it is down-regulated. Notably, this reduced expression of \textit{SNCA-AS1} is mainly in cortical areas,\textsuperscript{20} thus potentially increasing alpha-synuclein levels and exposing the cerebral cortex to a greater risk of neurodegeneration in carriers of this RBD-associated variant. The latter hypothesis should be tested in relevant animal models. Altogether, these data suggest that, depending on possible region-specific effects, different \textit{SNCA} variants might play different roles in synucleinopathies.

Similar to PD+RBD, other variants could be involved in the development of PD-RBD. We found that three of the 92 PD GWAS signals associated with increased PD risk in Europeans and Asians,\textsuperscript{15, 43, 48} the \textit{LRRK2} variant p.G2019S and the \textit{SNCA} variants rs5019538 and rs356182, were all associated with PD-RBD, compared with PD+RBD. The association between p.G2019S and PD-
RBD is in line with a previously reported reduced frequency of RBD in PD patients carrying this variant, and with lack of p.G2019S carriers in about 1,000 iRBD patients in another study. In addition to a reduced occurrence of RBD, carriers of the p.G2019S LRRK2 variant also present an overall more benign phenotype, including less frequent and milder cognitive decline. These findings, in addition to the nominal correlation between PD+RBD and the most recent PD GWAS in Europeans suggest that such GWAS might explain the genetic background of PD-RBD more than it does for PD+RBD.

These findings further support a pathophysiological relationship between the manifestation of RBD in PD and cognitive decline, which is in line with the comorbidity of these two clinical entities. It was hypothesized that this clinical and pathophysiological correlation could reflect the two alternative directions of alpha-synuclein spreading, body-first or brain-first. In body-first PD, alpha-synuclein pathology may start in the enteric nervous system, whereas in brain-first PD it may arise in the amygdala, entorhinal cortex and substantia nigra. These different neuropathological patterns correspond to two different subgroups of clinical progression. In body-first PD, RBD may manifest before the motor PD symptoms, and cognitive decline occurs faster, whereas in brain-first PD, RBD may occur after the onset of motor PD symptoms, if at all, and cognitive impairment develop more slowly. We can therefore speculate that the SNCA rs10005233 variant associated with PD+RBD might also be associated with the body-first subtype of PD, whereas the LRRK2 p.G2019S variant might be associated with the brain-first PD subtype, with less frequent RBD and milder cognitive decline. Since we cannot determine in our data which patients had RBD prior to PD diagnosis and which had it after PD diagnosis, this hypothesis should be studied in future genetic analyses of brain-first vs. body-first PD.

Similar to previous reports in iRBD and PD+RBD, in this study we did not observe any involvement of APOE variants in PD+RBD, suggesting that this gene does not affect RBD risk in PD patients. The rs117615688 variant (chromosomal position 17:43798308) in the CRHRI gene, located
in the *MAPT* locus, was nominally associated with RBD (OR=1.19, 95% CI=1.05-1.36, \( p=0.008 \)) with an opposite direction of effect to that seen in PD (OR=0.79, 95% CI=0.75-0.84, \( p=6.71\times10^{-16} \)) (Table 2).

There are several limitations in this study. All participants were Europeans, therefore our results might not fully apply to other ancestries. In addition, although we included a large number of patients with PD, insufficient power in our analysis might explain the lack of causative associations between PD+RBD and neuropsychiatric traits as well as of genome-wide significance of the LRRK2 p.G2019S and *GBA1* variants. It is possible that *GBA1* variants are strongly implicated also in the PD subtype without RBD, thus counterbalancing their previously reported contribution to RBD risk.\(^{12,20}\) Another limitation is represented by the inclusion of patients who developed RBD both before and after PD, as they may represent body-first vs. brain-first subtypes of PD as discussed above. Future research with larger sample sizes could investigate possible genetic and biological differences between them and specifically differentiate brain-first and body-first PD in that sense.

In conclusion, in this study we demonstrated that the risk of PD+RBD may be modified by variants in the *SNCA* and *LRRK2* loci, and potentially other loci. These genetic associations may explain why cognitive decline is more frequently observed in PD+RBD compared to PD-RBD, with possible implications for therapeutic management of PD patients. Future research will need to further explore the relationship between genetics, biology and clinical comorbidities to define PD subtypes and implement a precision medicine guided by early markers.

**Relevant conflicts of interest**

ZGO has received consulting fees from Lysosomal Therapeutics Inc., Idorsia, Prevail Therapeutics, Denali, Ono Therapeutics, Neuron23, Handl Therapeutics, UBC, Bial Biotech Inc., Bial, Deerfield, Guidepoint, Lighthouse and Inception Sciences (now Ventus). None of these companies were
involved in any parts of preparing, drafting and publishing this study. KH, PF, LNK, and the 23andMe Research Team are employed by and hold stock or stock options in 23andMe. HM is employed by UCL. In the last 12 months he reports paid consultancy from Roche, Aprinoia and Amylyx; lecture fees/honoraria - BMJ, Kyowa Kirin, Movement Disorders Society. Research Grants from Parkinson’s UK, Cure Parkinson’s Trust, PSP Association, Medical Research Council, Michael J Fox Foundation. Dr Morris is a co-applicant on a patent application related to C9ORF72 - Method for diagnosing a neurodegenerative disease (PCT/GB2012/052140)

Remaining authors declare no competing interests.

Data availability

The PD with and without RBD GWAS summary statistics without the 23andMe cohort is publicly available on GWAS catalog (https://www.ebi.ac.uk/gwas/). To access the full summary statistics with 23andMe an application to 23andMe is required. The full GWAS summary statistics for the 23andMe dataset will be made available through 23andMe to qualified researchers under an agreement with 23andMe that protects the privacy of the 23andMe participants. Please visit https://research.23andme.com/collaborate/#dataset-access/ for more information and to apply to access the data. The top 10,000 SNPs GWAS summary statistics including 23andMe is available on https://github.com/gan-orlab. The GWAS summary statistics for the neuropsychiatric traits used in the study are available on the GWAS catalog (https://www.ebi.ac.uk/gwas/) and Psychiatric Genomics Consortium (https://pgc.unc.edu/).

Code availability

The codes used for the analyses are available on https://github.com/daskrohn/RBD_GWAS and https://github.com/gan-orlab.

Acknowledgement
We wholeheartedly thank the participants in this study. We would like to thank the research participants and all members of IPDGC for making this work possible. The AMP-PD cohort data used in this study included the Fox Investigation for New Discovery of Biomarkers (BioFIND), the Harvard Biomarker Study (HBS) and the Parkinson’s Disease Biomarkers Program (PDBP) cohorts. For up-to-date information on the study, visit https://www.amp-pd.org. AMP PD – a public-private partnership – is managed by the FNIH and funded by Celgene, GSK, the Michael J. Fox Foundation for Parkinson’s Research, the National Institute of Neurological Disorders and Stroke, Pfizer, Sanofi, and Verily. BioFIND is sponsored by The Michael J. Fox Foundation for Parkinson’s Research (MJFF) with support from the National Institute for Neurological Disorders and Stroke (NINDS). The BioFIND Investigators have not participated in reviewing the data analysis or content of the manuscript. For up-to-date information on the study, visit https://www.michaeljfox.org/news/biofind. The Harvard Biomarker Study (HBS) is a collaboration of HBS investigators [full list of HBS investigators found at https://www.bwhparkinsoncenter.org/biobank/ and funded through philanthropy and NIH and Non-NIH funding sources. The HBS Investigators have not participated in reviewing the data analysis or content of the manuscript. The Parkinson’s Disease Biomarker Program (PDBP) consortium is supported by the National Institute of Neurological Disorders and Stroke (NINDS) at the National Institutes of Health. A full list of PDBP investigators can be found at https://pdbp.ninds.nih.gov/policy. The PDBP investigators have not participated in reviewing the data analysis or content of the manuscript.

The PRoBaND cohort is primarily funded and supported by Parkinson’s UK (https://www.parkinsons.org.uk/) and supported by the National Institute for Health Research (NIHR) Dementias and Neurodegenerative Diseases Research Network (DeNDRoN) and by NHS Greater Glasgow and Clyde. PRoBaND has multicentre research ethics approval from the West of Scotland Research Ethics Committee: IRAS 70980, MREC 11/AL/0163. The Oxford Parkinson’s Disease
Centre (OPDC) Discovery cohort is also primarily funded by Parkinson’s UK and additionally supported by the NIHR-DeNDRoN and by the NIHR Oxford Biomedical Research Centre, based at the Oxford University Hospitals NHS Trust, and the University of Oxford. OPDC has multicentre research ethics approval from the South Central Oxford A Research Ethics Committee: 16/SC/0108. The Parkinson’s Family Project (PFP) cohort has received funding from Parkinson’s UK and the Janet Owens Bequest Fund. PFP has multicentre research ethics approval from the London – Camden and Kings Cross Research Ethics Committee: 5/LO/0097. RR is funded by Aligning Science Across Parkinson’s (grant number ASAP-000478) through the Michael J. Fox Foundation for Parkinson’s Research (MJFF). This work was financially supported by the Michael J. Fox Foundation, Parkinson’s Society Canada, the Canadian Consortium on Neurodegeneration in Aging (CCNA), and the Canada First Research Excellence Fund (CFREF), awarded to McGill University for the Healthy Brains for Healthy Lives (HBHL) program. ZGO is supported by the Fonds de recherche du Québec - Santé (FRQS) Chercheurs-boursiers award, and is a William Dawson Scholar.

We would also like to thank the research participants and employees of 23andMe for making this work possible.

The following members of the 23andMe Research Team contributed to this study:

Suyash Shringarpure, Qiaojuan Jane Su, Susana A. Tat, Vinh Tran, Joyce Y. Tung, Xin Wang, Wei Wang, Catherine H. Weldon, Peter Wilton, Corinna D. Wong.

K.H., P.F., L.K., and 23andMe Research Team are employed by and hold stock or stock options in 23andMe, Inc.

**Author contributions**

YLS, ZGO, LK: conception and design of the study

YLS, LK, KH, PF, LNK, EY, UR, JAR, KF, FA, KB, MS, HM, LW, RR, LP, MT, TG, KB, HL, MTMH, DGG, SJGL, JBK, PP, IA, MS, AL, MO, MR, ZGO: data acquisition/analysis

YLS, ZGO: manuscript drafting
REFERENCES


Fig 1 - Manhattan plot of PD+RBD vs. PD-RBD

Manhattan plot showing the results of the GWAS meta-analysis, comparing PD+RBD and PD-RBD, highlighting the SNCA and LRRK2 loci. The Y axis represents the negative logarithm of p-value, the X axis represents the chromosomal position of the variants and each dot on the figure represents a SNP. The red line represents the genome-wide Bonferroni-corrected statistical significance threshold (5x10⁻⁸), whereas the blue line is the false-discovery rate-corrected significance threshold (5x10⁻⁵).

Chr: chromosome; PD: Parkinson’s disease; RBD: REM sleep behavior disorder.

Fig 2 - Genetic correlation between PD+RBD and neuropsychiatric traits

The bar plot shows the genetic correlations between PD+RBD and neuropsychiatric traits. The correlation coefficient is illustrated on the X axis. Green bars represent positive correlations whereas red bars negative ones (i.e., a positive correlation of the neuropsychiatric trait with PD-RBD). The asterisks highlight the nominally significant correlations.

ALS: amyotrophic lateral sclerosis; CD: cognitive decline; AD: Alzheimer's disease; PD: Parkinson's disease; DLB: dementia with Lewy bodies; Alcohol dep: alcohol dependence; cannabis dep: cannabis dependence; ADHD: attention deficit hyperactivity disorder; OCD: obsessive-compulsive disorder; ASD: autism spectrum disorder; TS: Tourette syndrome; AN: anorexia nervosa; PTS: post-traumatic syndrome; scz: schizophrenia; BD: bipolar disorder; MDD: major depressive disorder.

Supplementary Fig 1 – Q-Q plot of the PD with and without RBD GWAS

The Q-Q plot illustrates the negative log-adjusted p-values from the GWAS sorted into ascending order against the expected quantiles if the null hypothesis is true for all tests.

Supplementary Fig 2 – Leave-one-out analysis in the Mendelian Randomization
The plot shows the Mendelian Randomization coefficient (black dot) and 95% confidence interval (bars) when each variant is excluded from the analysis. The excluded variant is indicated on the Y axis. The red dot and line represent the results if all the variants are included.


**Supplementary Fig 3 – Single variant analysis in the Mendelian Randomization**

The plot shows the Mendelian Randomization coefficient (black dot) and 95% confidence interval (bars) for each individual variant included in the analysis. The variant analyzed is indicated on the Y axis. The red dot and line represent the results if all the variants are included.

Table 1

Demographic characteristics of PD patients in the individual cohorts

<table>
<thead>
<tr>
<th>Center</th>
<th>PD+RBD, n</th>
<th>Age PD+RBD (SD)</th>
<th>%Fem PD+RBD</th>
<th>PD-RBD, n</th>
<th>Age PD-RBD (SD)</th>
<th>%Fem PD-RBD</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oslo</td>
<td>130</td>
<td>65.6 (8.8)</td>
<td>25%</td>
<td>180</td>
<td>66.1 (10)</td>
<td>44%</td>
<td>310</td>
</tr>
<tr>
<td>Lund</td>
<td>365</td>
<td>71.4 (7)</td>
<td>39%</td>
<td>555</td>
<td>70.8 (9.2)</td>
<td>33%</td>
<td>920</td>
</tr>
<tr>
<td>McGill</td>
<td>285</td>
<td>67.4 (9.2)</td>
<td>31%</td>
<td>217</td>
<td>67 (8.8)</td>
<td>49%</td>
<td>502</td>
</tr>
<tr>
<td>AMP-PD</td>
<td>111</td>
<td>66.4 (9.4)</td>
<td>69%</td>
<td>269</td>
<td>65.5 (10.3)</td>
<td>62%</td>
<td>380</td>
</tr>
<tr>
<td>Sydney</td>
<td>105</td>
<td>59 (10.8)</td>
<td>32%</td>
<td>125</td>
<td>60 (10.7)</td>
<td>38%</td>
<td>230</td>
</tr>
<tr>
<td>Tuebingen</td>
<td>453</td>
<td>68.8 (9.1)</td>
<td>33%</td>
<td>659</td>
<td>67.7 (10.2)</td>
<td>38%</td>
<td>1,112</td>
</tr>
<tr>
<td>Barcelona</td>
<td>133</td>
<td>69.5 (9.8)</td>
<td>44%</td>
<td>71</td>
<td>74.3 (10)</td>
<td>66%</td>
<td>204</td>
</tr>
<tr>
<td>PRoBaND</td>
<td>585</td>
<td>67.3 (8.9)</td>
<td>30%</td>
<td>1,134</td>
<td>67.7 (9.3)</td>
<td>38%</td>
<td>1,719</td>
</tr>
<tr>
<td>PFP</td>
<td>257</td>
<td>62.4 (12.5)</td>
<td>37%</td>
<td>339</td>
<td>61.4 (13)</td>
<td>47%</td>
<td>596</td>
</tr>
<tr>
<td>OPDC</td>
<td>274</td>
<td>67.3 (9.3)</td>
<td>27%</td>
<td>539</td>
<td>67.4 (9.6)</td>
<td>38%</td>
<td>813</td>
</tr>
<tr>
<td>Kosice</td>
<td>102</td>
<td>71.4 (8.3)</td>
<td>33%</td>
<td>225</td>
<td>69.2 (10.7)</td>
<td>42%</td>
<td>327</td>
</tr>
<tr>
<td></td>
<td>PD+RBD, n: Number of Parkinson’s disease patients with REM sleep behavior disorder</td>
<td>PD-RBD, n: Number of Parkinson’s disease patients without REM sleep behavior disorder</td>
<td>Age: mean age in the group</td>
<td>SD: standard deviation</td>
<td>%Fem: percentage of females</td>
<td>Tot: total number of PD patients in the cohort</td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>----------------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------------</td>
<td>--------------------------</td>
<td>------------------------</td>
<td>----------------------------</td>
<td>----------------------------------</td>
<td></td>
</tr>
<tr>
<td>23andMe</td>
<td>2,603</td>
<td>NA*</td>
<td>38%</td>
<td>8,707</td>
<td>NA*</td>
<td>45%</td>
<td>11,310</td>
</tr>
<tr>
<td>TOTAL</td>
<td>5,403</td>
<td>/</td>
<td>/</td>
<td>13,020</td>
<td>/</td>
<td>/</td>
<td>18,423</td>
</tr>
</tbody>
</table>

*23andMe only provides age ranges (i.e., in cases: 5 individuals <30 years of age, 49 individuals 30-45, 346 individuals 45-60, 2203 individuals >60; in controls: 11 individuals <30, 146 individuals 30-45, 1302 individuals 45-60, 7248 individuals >60)
Table 2

Association of variants from previous PD GWAS with PD+RBD in the current study

<table>
<thead>
<tr>
<th>Variant</th>
<th>Nearest gene</th>
<th>OR (95% CI)</th>
<th>P-value</th>
<th>OR (95% CI)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>4:90636630</td>
<td>SNCA</td>
<td>0.85 (0.81-0.89)</td>
<td>2.46E-10*</td>
<td>1.17 (1.14-1.2)</td>
<td>1.13E-36</td>
</tr>
<tr>
<td>12:40734202</td>
<td>LRRK2</td>
<td>0.41 (0.28-0.61)</td>
<td>1.04E-05*</td>
<td>11.35 (9.44-13.63)</td>
<td>3.61E-148</td>
</tr>
<tr>
<td>4:90626111</td>
<td>SNCA</td>
<td>0.89 (0.84-0.95)</td>
<td>0.001365*</td>
<td>1.32 (1.29-1.35)</td>
<td>3.89E-154</td>
</tr>
<tr>
<td>16:30977799</td>
<td>SETD1A</td>
<td>0.92 (0.88-0.97)</td>
<td>0.001952</td>
<td>1.09 (1.07-1.12)</td>
<td>5.12E-20</td>
</tr>
<tr>
<td>19:2341047</td>
<td>SPPL2B</td>
<td>1.09 (1.03-1.15)</td>
<td>0.002506</td>
<td>0.93 (0.91-0.95)</td>
<td>4.18E-10</td>
</tr>
<tr>
<td>17:43798308</td>
<td>CRHRI</td>
<td>1.19 (1.05-1.36)</td>
<td>0.008267</td>
<td>0.79 (0.75-0.84)</td>
<td>6.71E-16</td>
</tr>
<tr>
<td>3:28705690</td>
<td>LINC00693</td>
<td>0.95 (0.9-0.99)</td>
<td>0.02552</td>
<td>1.07 (1.05-1.09)</td>
<td>8.09E-12</td>
</tr>
</tbody>
</table>

PD: Parkinson’s disease; RBD: REM sleep behavior disorder; OR (95% CI): odds ratio with relative 95% confidence interval.

*variant statistically significant after Bonferroni correction (α/ number of variants=0.00054).
<table>
<thead>
<tr>
<th>Cohort</th>
<th>Non-carriers PD-RBD, N</th>
<th>Carriers PD-RBD, N (%)</th>
<th>Non-carriers PD+RBD, N</th>
<th>Carriers PD+RBD, N (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oslo</td>
<td>156</td>
<td>1 (0.64%)</td>
<td>122</td>
<td>1 (0.81%)</td>
</tr>
<tr>
<td>Lund</td>
<td>552</td>
<td>3 (0.54%)</td>
<td>364</td>
<td>1 (0.27%)</td>
</tr>
<tr>
<td>McGill</td>
<td>212</td>
<td>5 (2.30%)</td>
<td>282</td>
<td>3 (1.05%)</td>
</tr>
<tr>
<td>AMP-PD</td>
<td>258</td>
<td>9 (3.37%)</td>
<td>110</td>
<td>1 (0.90%)</td>
</tr>
<tr>
<td>Sydney</td>
<td>115</td>
<td>1 (0.86%)</td>
<td>98</td>
<td>1 (1.01%)</td>
</tr>
<tr>
<td>Tuebingen</td>
<td>365</td>
<td>0 (0.00%)</td>
<td>641</td>
<td>0 (0.00%)</td>
</tr>
<tr>
<td>Barcelona</td>
<td>45</td>
<td>0 (0.00%)</td>
<td>100</td>
<td>0 (0.00%)</td>
</tr>
<tr>
<td>PRoBaND</td>
<td>1132</td>
<td>2 (0.18%)</td>
<td>582</td>
<td>3 (0.51%)</td>
</tr>
<tr>
<td>PFP</td>
<td>328</td>
<td>9 (2.67%)</td>
<td>246</td>
<td>6 (2.38%)</td>
</tr>
<tr>
<td>OPDC</td>
<td>537</td>
<td>2 (0.37%)</td>
<td>274</td>
<td>0 (0.00%)</td>
</tr>
<tr>
<td>Kosice</td>
<td>216</td>
<td>1 (0.46%)</td>
<td>101</td>
<td>0 (0.00%)</td>
</tr>
<tr>
<td>23andMe</td>
<td>8459</td>
<td>248 (2.85%)</td>
<td>2584</td>
<td>19 (0.73%)</td>
</tr>
<tr>
<td>TOTAL*</td>
<td>12010</td>
<td>281 (2.29%)</td>
<td>4863</td>
<td>35 (0.71%)</td>
</tr>
</tbody>
</table>
PD-RBD: participants without REM sleep behavior disorder (RBD); PD+RBD: participants with RBD; N: number of participants; %: percentage of participants; Oslo: Oslo University Hospital; Lund: Lund University; McGill: McGill University; AMP-PD: Accelerating Medicines Partnership Parkinson's disease, including the New Discovery of Biomarkers (BioFIND), the Harvard Biomarker Study (HBS) and the Parkinson’s Disease Biomarkers Program (PDBP) cohorts; Sydney: University of Sydney; Tuebingen: University of Tuebingen; Barcelona: Hospital Universitari Mutua de Terrassa; PRoBaND: Parkinson's repository of biosamples and networked datasets; PFP: Parkinson's Families Project; OPDC: Oxford Parkinson's Disease Centre; Kosice: Pavol Jozef Šafárik University in Kosice.

*The total excludes individuals with unknown carrier status for p.G2019S
PD with/without RBD - neuropsychiatric traits

Trait

HEADACHE
EPILEPSY
ALS
CD
AD
PD
DLB
ALCOL DEP
CANNABIS DEP
ADHD
OCD
ASD
AN
PTS
BD
MDD

Genetic correlation coefficient