



Published in final edited form as:

*Ann Neurol.* 2020 April ; 87(4): 584–598. doi:10.1002/ana.25687.

## Fine-mapping of *SNCA* in REM sleep behavior disorder and overt synucleinopathies

A full list of authors and affiliations appears at the end of the article.

### Abstract

**Objective:** REM-sleep behavior disorder (RBD) is a prodromal synucleinopathy, as >80% will eventually convert to overt synucleinopathy. We performed an in-depth analysis of the *SNCA* locus to identify RBD-specific risk variants.

**Methods:** Full sequencing and genotyping of *SNCA* was performed in isolated/idiopathic RBD (iRBD, n=1,076), Parkinson's disease (PD, n=1,013), and dementia with Lewy bodies (DLB, n=415), and in control subjects (n=6,155). iRBD cases were diagnosed with RBD prior to neurodegeneration, although some have since converted. A replication cohort from 23andMe of PD patients with probable RBD (pRBD) was also analyzed (cases n=1,782, controls n=131,250). Adjusted logistic regression models and meta-analyses were performed. Effects on conversion rate were analyzed in 432 RBD patients with available data using Kaplan-Meier survival analysis.

**Results:** A 5'-region *SNCA* variant (rs10005233) was associated with iRBD (OR=1.43,  $p=1.1E-08$ ), which was replicated in pRBD. This variant is in linkage disequilibrium (LD) with other 5' risk variants across the different synucleinopathies. An independent iRBD-specific suggestive association (rs11732740) was detected at the 3' of *SNCA* (OR=1.32,  $p=4.7E-04$ , not statistically significant after Bonferroni correction). Homozygous carriers of both iRBD-specific SNPs were at highly increased risk for iRBD (OR=5.74,  $p=2E-06$ ). The known top PD-associated variant (3' variant rs356182) had an opposite direction of effect in iRBD compared to PD.

**Interpretation:** There is a distinct pattern of association at the *SNCA* locus in RBD as compared to PD, with an opposite direction of effect at the 3' of *SNCA*. Several 5' *SNCA* variants are associated with iRBD and with pRBD in overt synucleinopathies.

### INTRODUCTION

Rapid eye movement (REM) sleep behavior disorder (RBD) is characterized by lack of muscle atonia during REM sleep, causing patients to enact their dreams. Population-wide prevalence of RBD is estimated to be approximately 0.5–1%. {Haba-Rubio, 2017

---

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

Author contributions

*Conception and design of the study:* LK, ZG-O

*Acquisition and analysis of data:* LK, RYJW, KH, JAR, SBL, CB, AA, IA, MTMH, YD, BH, MT, KAB, AS, EH, CCM, AB, GP, EA, LF-S, PY, AH, VCC, BM, FS-D, CT, KS, DK, MF, MP, FD, VM, WO, MT, GLG, MV, J-FG, MAN, ABS, 23M, AD, JYM, PC, OAR, BFB, ND, EAF, RBP, LP, GAR, ZG-O

*Drafting a significant portion of the manuscript or figures:* LK, ZG-O

Potential conflict of Interests

The authors have no conflicts to report.

#105;Pujol, 2017 #106;Kang, 2013 #107} Isolated (also referred to as ‘idiopathic’) RBD (iRBD) is a prodromal condition for synucleinopathies, neurodegenerative disorders pathologically characterized by  $\alpha$ -synuclein deposition. Within an average of 10–15 years, more than 80% of iRBD cases will convert to Parkinson’s disease (PD, most of whom will eventually develop PD with dementia, PDD{Fereshtehnejad, 2015 #49}), dementia with Lewy bodies (DLB), or in fewer cases multiple system atrophy (MSA).{Iranzo, 2016 #75;Iranzo, 2013 #18;Schenck, 2013 #76;Postuma, 2012 #47;Postuma, 2009 #77} The time between iRBD onset or diagnosis to conversion is highly variable; some iRBD patients convert very rapidly, while others may convert decades after the diagnosis of iRBD, {Postuma, 2016 #108} and the mechanisms affecting the risk for RBD and rate of conversion are mostly unknown.{Högl, 2018 #74}

In recent years, preliminary studies suggested that the genetics of iRBD only partially overlap with those of PD or DLB.{Dauvilliers, 2018 #110} Similar to PD, iRBD has been associated with risk variants in *GBA*,{Gan-Or, 2015 #52} to the *LRRK2* protective haplotype{Bencheikh, 2018 #23} and to *TMEM175* variants.{Krohn, #128} However, no association has been found with other key PD or DLB risk variants in *LRRK2*,{Fernández-Santiago, 2016 #111} *MAPT* haplotypes,{Li, 2018 #95} and the *APOE*  $\epsilon$ 4 risk haplotype. {Gan-Or, 2017 #48} Thus far, the role of the *SNCA* gene in iRBD has not been thoroughly studied. *SNCA* encodes  $\alpha$ -synuclein, the main protein component of Lewy-bodies and neurites in synucleinopathies.{Kim, 2014 #112} Interestingly, in different forms of synucleinopathies, there are different, reportedly independent variants in the *SNCA* locus that have been associated with the risk for the disease. In PD, multiple independent *SNCA* association signals have been identified, but a downstream (3’) *SNCA* single nucleotide polymorphism (SNP) rs356182 has consistently been the top genome-wide association study (GWAS) signal in all large-scale meta-analyses, with a secondary independent association at the 5’ region of *SNCA*.{Chang, 2017 #4;Nalls, 2014 #5} However, the top 3’ PD-associated variant is not associated with DLB. Rather, in DLB a strong association has been demonstrated with a SNP at the 5’ region of *SNCA* (rs7681440).{Guerreiro, 2018 #99} The latter is in linkage disequilibrium (LD) with the secondary signal in PD, i.e. both variants appear on the same haplotype and are inherited together. Another SNP in LD with this 5’ region haplotype was associated with probable RBD (pRBD, determined by a validated questionnaire) in PD patients in a previous study.{Bjørnarå, 2018 #72} Other SNPs in the *SNCA* locus have been suggested to be involved in PDD{Guella, 2016 #64} and Alzheimer’s disease with Lewy body pathology (ADLBV, a variant of AD which also demonstrates diffuse Lewy body pathology).{Linnertz, 2014 #53} In MSA, contradicting results have been reported regarding the *SNCA* locus, and the largest analysis thus far refuted the previously reported association.{Sailer, 2016 #104}

In the current study, we aimed to thoroughly analyze the *SNCA* locus in RBD. We used a combination of full *SNCA* sequencing and comprehensive SNP genotyping in the largest genetic study of iRBD to date, as well as in PD and DLB patients with and without pRBD, to examine *SNCA* association with RBD risk and phenoconversion from RBD to defined neurodegenerative disease.

## METHODS

### Population

**Discovery cohort**—The discovery cohort included unrelated, consecutively recruited iRBD (n=1,076) and PD patients (n=733), and controls (n=6,155) of European ancestry (determined for all cases and controls using HapMap v.3 in hg19/GRCh37 and principal component [PC] analysis). iRBD refers to cases who were initially diagnosed with RBD only, without another defined neurodegenerative synucleinopathy at the time of diagnosis. Data on conversion from 432 of these patients who have been followed-up by their clinicians were collected, 237 of whom had converted. Patients were not pre-screened for any known pathogenic variants prior to enrollment. 81% of iRBD cases were men, with an average age at enrollment of  $68\pm 9.2$  years and average disease duration of  $8.4\pm 6.9$  years, collected from 17 centres; detailed cohort information by centre is detailed in Table 1. The control group was composed of elderly (n=225,  $63.5\pm 8$  years) and young controls (n=650,  $36\pm 7$  years) of European origin collected in Montreal (52% men), and elderly controls from CARTaGENE (n=5,245,  $55\pm 8$  years, 41% men), a registry that collects clinical data and DNA in Canada (<https://www.cartagene.qc.ca/en/home>).{Awadalla, 2012 #114} Since the variant frequencies were similar in the young and elderly control groups, and since the low prevalence of iRBD in the general population leaves little likelihood of undetected RBD patients in the control groups, they were combined for analysis. Because of the disparity in sex distribution between cases and controls, sex is included as a covariate in all statistical analyses. Homogeneity of the cohorts was confirmed using PC analysis, and PCs were included in the statistical analysis to account for unknown ethnical differences that could bias the genetic associations. Variability in sex and age were also taken into account and adjusted for in the statistical analysis.

RBD was diagnosed with video polysomnography (vPSG) according to the International Classification of Sleep Disorders, version 2 (ICSD-2) criteria.{Thorpy, 2012 #126} Additional data including age at onset and diagnosis of RBD, eventual phenoconversion to an overt synucleinopathy, and rate of phenoconversion were available for a subset of samples (n=432). In all cohorts, PD was diagnosed by movement disorder specialists according to the UK Brain Bank Criteria{Hughes, 1992 #25} without excluding patients who had relatives with PD (to 2015), or International Parkinson Disease and Movement Disorders Society criteria (after 2015).{Postuma, 2015 #115} DLB was diagnosed based on the 3rd (prior to 2017) and 4th (after 2017) DLB consensus criteria.{McKeith, 2017 #50} All study participants signed informed consent forms, and the study protocol was approved by the institutional review boards.

**Replication cohorts**—For analysis of probable RBD (pRBD) in PD and DLB, data was available for a total of 2,450 synucleinopathy patients with pRBD, 774 patients without pRBD, and 131,250 controls. These include a European cohort from 23andMe (PD+pRBD n=1,782, controls n=131,250), Montreal cohort (n=183 with pRBD [PD+pRBD], n=243 without pRBD [PD-pRBD]), a PD cohort from Oslo, Norway (PD+pRBD n=123, PD-pRBD n=147), a PD cohort from the Parkinson's Progression Marker Initiative{Marek, 2011 #125} (PPMI, PD+pRBD n=106, PD-pRBD n=276) and a DLB cohort from the Mayo Clinic (DLB

+pRBD n=256, DLB-pRBD n=108). Association results including the samples from the Oslo cohort were reported for three *SNCA* risk SNPs in a previous publication. {Bjørnarå, 2018 #72} RBD was assessed by the RBD screening questionnaire (RBDSQ) in the Oslo and PPMI cohorts, and by the RBD1Q question in the Montreal PD, 23andMe, and DLB cohorts. Both questionnaires have high sensitivity and specificity in PD. {Postuma, 2012 #80; Nomura, 2011 #86} The diagnoses of PD and DLB in the replication cohorts (except for the 23andMe cohort) was done using the same criteria as the discovery cohort. PD in the 23andMe cohort is self-reported.

## Genetic Analysis

**Single nucleotide polymorphism analysis**—DNA was extracted using a standard salting out protocol. iRBD and PD cohorts from Montreal, and DLB cohort from the Mayo Clinic were genotyped using the OmniExpress-24 v1.2 chip (Illumina Inc., approximately 700,000 SNPs) with added NeuroX GWAS custom SNPs, comprised of over 24,000 SNPs associated with neurological diseases. CARTaGENE controls were genotyped using the Infinium Global Screening Array (GSA, Illumina). Data was converted to PLINK {Purcell, 2007 #14} format and merged using only SNPs genotyped on both platforms (n~160k). Oslo and PPMI PD samples were genotyped as previously described. {Pihlstrøm, 2018 #71; Bjørnarå, 2018 #72}

To exclude low quality SNPs and samples, quality control (QC) was performed on the variant level (SNPs excluded if genotype quality < 95%, missingness > 5%, divergent call rates between cases and controls  $p < 1E-04$ , or departure from the Hardy-Weinberg equilibrium  $p < 1E-04$ ) and on the sample level (samples excluded if genotype data showed missingness in > 5%, abnormal heterozygosity, conflicting sex assignment, cryptic relatedness  $\text{pihat} > 0.125$ , or non-European ancestry). This QC was separately performed on each cohort and then performed again when data was merged for analysis. To avoid potential bias due to different ethnicities, sample QC including ancestry and relatedness confirmed European ancestry and homogeneity across the cohorts. PCs were calculated on pre-imputed genome-wide merged data of pruned SNPs (pairwise  $R^2 > 0.5$ ) and minor allele frequency (MAF) > 0.05 for use in statistical analyses. Imputation was performed on the merged samples with the Michigan Imputation Server using the HaploType Reference Consortium {McCarthy, 2016 #89} r1.1 2016 reference panel and filtered for imputation quality > 0.8. Post imputation, a total of 1,862 SNPs in the *SNCA* locus (defined as +/- 500kb around the PD top GWAS hit, rs356182) were obtained.

In the 23andMe cohort, DNA extraction and genotyping were performed on saliva samples by the 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 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 is an Illumina Infinium Global Screening Array (~640,000 SNPs) supplemented with ~50,000 SNPs of custom content. Samples that failed to reach 98.5% call rate were re-analyzed. Those who did not reach a sufficient call rate were excluded. All individuals included in the analyses provided informed consent and answered surveys online according to the 23andMe human subject protocol, which was reviewed and approved by Ethical & Independent Review Services, a private institutional review board (<http://www.eandireview.com>). QC was performed similar to the procedure outlined above.

**Full sequencing of SNCA**—The coding and 3' and 5' untranslated regions (UTR) of *SNCA* were sequenced in iRBD patients (n=1,076) and controls (n=910) using Molecular Inversion Probes (MIPs) designed, targeted, and amplified as previously described. {Ross, 2016 #10} These probes are designed to specifically target genomic regions of interest and allow sequencing of selected genes. Targeting probes are detailed in Table S1 and the full protocol is available upon request. The MIPs library was sequenced using Illumina HiSeq 4000 platform at the McGill University and Genome Québec Innovation Centre. Sequencing data processing was done by Burrows-Wheeler Alignment (BWA), Genome Analysis Toolkit {McKenna, 2010 #8} (GATK v3.8) for post-alignment adjustments and variant calling, and ANNOVAR {Wang, 2010 #9} for annotation. Variant frequencies were extracted from two public databases: gnomAD {Lek, 2016 #12} and PDgene. {Lill, 2012 #11} To only include variants of high quality, variants were filtered for minimum depth of coverage at 30x, Hardy Weinberg equilibrium  $p > 0.001$ , genotype quality  $> 90\%$ , and missingness  $< 10\%$  in both variants and samples.

### Statistical Analysis

To analyze the association between *SNCA* SNPs and iRBD, case-control logistic regression models were used, with the top 3 PCs (number of PCs required for adjustment were determined by scree plot), sex, and age as covariates. The number of independent tests for multiple testing correction was determined according to pruned number of *SNCA* SNPs with  $MAF > 0.05$  by  $R^2 > 0.5$  (n=191) to avoid an overly stringent significance threshold by correcting separately for SNPs that are in high LD and represent the same haplotype. This method set the Bonferroni-corrected significance threshold at  $p < 2.6E-04$ , which is in line with previously established thresholds in the same *SNCA* region. {Pihlstrøm, 2018 #71} Logistic regression was performed in a stepwise forward method {Cordell, 2002 #87} as previously described. {Pihlstrøm, 2018 #71} This approach allows for the identification of additional disease-related SNPs that are otherwise masked by the top association. This is achieved by including the top associated SNP as an additional covariate in a following logistic regression. Linkage disequilibrium (LD), the co-appearance of SNPs that are inherited together in the same genomic region, may lead to identification of multiple associations that are all driven by a single variant. This method ensures independence of the significantly associated SNPs. LD was calculated using the European reference cohort on LDlink (<https://ldlink.nci.nih.gov/>), and visualized for top hits and known synucleinopathy risk variants in Haploview. {Barrett, 2004 #91} Variants which passed the aforementioned significance threshold were replicated using 23andMe PD+pRBD versus controls including the covariates sex, age, and PCs 1–5. Next, top risk SNPs for iRBD and overt

synucleinopathies were analyzed in the PD and DLB cohorts, comparing patients with and without pRBD (PD+/-pRBD, DLB+/-pRBD). Each cohort was analyzed separately, followed by a meta-analysis with the R package metafor, { Viechtbauer, 2010 #13 } using logistic regression adjusted for sex, age, and PCs 1–3. The association of rare variants with RBD risk was evaluated in the MIPs sequencing data using optimized sequence Kernel association test (SKAT-O). { Lee, 2012 #92 }

To examine whether *SNCA* variants affect the rate of conversion of RBD, we estimated both duration from RBD onset to conversion onset and RBD diagnosis to conversion diagnosis. Top synucleinopathy-associated SNPs and 3' and 5' UTR variants with MAF >0.01 from the sequencing data were included in the analysis (n=8, Detailed in the results section). We examine the potential association between variants and rate of conversion with the intention to genetically stratify patients for clinical trials; those who convert quickly would be ideal for measuring protective interventions. Rare variants are not considered since large sample sizes are needed for these trials. The association between SNPs and rate of conversion was examined using Kaplan-Meier survival analysis. The Bonferroni-corrected p value threshold was  $p < 0.006$ , accounting for the 8 variants. Analyses were performed using R version 3.5.1.

## RESULTS

### *SNCA* variants included in the analysis

To perform the analysis on high quality GWAS data, a total of 1,862 SNPs in the *SNCA* locus were included in the final analysis (Table S2). From the targeted sequencing of *SNCA*, 53 rare variants were found and included in the SKAT-O analyses, as well as 3 common variants included in conversion analyses (Table S4, including coding, intronic and untranslated region variants). Of the 53 rare variants, 4 were nonsynonymous; p.K97R, p.P117S (rs145138372), and p.A124T (rs1358566725) were each found in a single iRBD patient, and p.N122S (rs749476922) was found in two controls. None were previously associated with PD or any other condition. Each variant is quite rare (MAF < 1e-04 on gnomAD).

### Risk for RBD is primarily associated with 5' *SNCA* SNPs

Two independent signals, one at the 5' (rs10005233, OR=1.43, 95%CI=1.27–1.62,  $p=1.1e-08$ ) and one at the 3' (rs11732740, OR=1.32, 95%CI=1.13–1.53,  $p=4.7e-04$ ) of *SNCA*, were associated with risk for iRBD (Table 2, Figure 2), yet only the 5' SNP remained statistically significant after Bonferroni correction. Analysis without the CARTaGENE control samples that were done on a different platform, yielded almost identical results (data not shown). Homozygous carriers of both SNPs were associated with highly increased risk for iRBD (OR=5.74, 95%CI=2.81–11.72,  $p=2E-06$ ), with a gradual increase in risk dependent on the number of alleles with risk variants in these two SNPs (Table 3). The association of the *SNCA* 5' variant with iRBD was then examined in the 23andMe cohort of PD+pRBD patients (n=1,782) and controls (n=131,250), and was replicated (OR=1.15, 95%CI=1.08–1.23,  $p=4.6E-05$ ). This 5' variant (rs10005233) is in LD with the previously published top signal for DLB, rs7681440 ( $R^2=0.94$ ,  $D'=0.99$ ), { Guerreiro, 2018 #99 } as well as the secondary PD 5' GWAS signal rs763443 ( $R^2=0.78$ ,

$D'=0.89$ ) {Chang, 2017 #4; Nalls, 2014 #5} and a variant previously associated with ADLBV rs2583988 ( $R^2=0.40$ ,  $D'=0.99$ ), {Linnertz, 2014 #53} suggesting that one haplotype may drive this association across all synucleinopathies (Figure 1A). The second independent iRBD-associated signal at the 3' of *SNCA* (RBD 3', rs11732740) did not reach Bonferroni-corrected statistical significance. Notably, this SNP is not in LD with the top GWAS 3' SNP associated with PD, rs356182 {Chang, 2017 #4; Nalls, 2014 #5} ( $R^2=0.003$ ,  $D'=0.14$ ), i.e., it is an independent risk factor. However, this SNP is in partial LD with another, independent 3' PD risk SNP rs2870004, which was recently reported {Pihlstrøm, 2018 #71} ( $R^2=0.04$ ,  $D'=0.81$ ), suggesting that these two associations (with iRBD and PD) might be driven by the same genetic variant.

To determine whether iRBD was associated with rare *SNCA* variants ( $MAF < 0.01$ ), all rare variants (detailed in Table S4) were analyzed together using SKAT-O (burden), followed by separate analyses of nonsynonymous, synonymous, and UTR variants with the same method. None of the analyses suggested association between any of the types of rare *SNCA* variants and iRBD.

### SNPs previously linked to synucleinopathies and their association with RBD

To examine whether *SNCA* variants previously associated with synucleinopathies are also associated with RBD, we extracted the top hits from studies performed on PD, {Pihlstrøm, 2018 #71; Nalls, 2018 #116; Chang, 2017 #4} PDD, {Guella, 2016 #64} DLB, {Guerreiro, 2018 #99} and ADLBV. {Linnertz, 2014 #53} Table 4 details the association of these variants with risk for iRBD in our cohort, and the LD with the iRBD 5' risk variant. Additionally, Table 4 indicates whether the variant in question shows the same direction of effect in iRBD that it does in the cited synucleinopathy. Only PD-associated variants had opposite effects in PD and iRBD, as two variants are associated with increased risk for PD and reduced risk for iRBD. The top PD signal rs356182-G (3') is associated with increased risk for PD ( $OR=1.32$ , 95%  $CI=1.30-1.35$ ,  $p=3.9E-194$ ), yet decreased risk for RBD ( $OR=0.79$ , 95%  $CI=0.70-0.91$ ,  $p=6E-04$  in iRBD, Table 4). Since we found that some *SNCA* variants have opposite effects in PD and iRBD, we further examined whether the variants that are associated with the different synucleinopathies detailed in Table 4 also have differential effects in PD vs. iRBD. Figure 1B compares the effect sizes (betas) of these top hits between iRBD and PD. Interestingly, several variants had opposite effects in PD vs. iRBD (Figure 1B). For example, the ADLBV variant associated with increased risk for ADLBV (rs2583988-C) is associated with decreased risk for PD ( $OR=0.88$ , 95%  $CI=0.86-0.90$ ,  $p=3.2E-36$ ) and increased risk for iRBD ( $OR=1.27$ , 95%  $CI=1.10-1.48$ ,  $p=0.001$ ), although without statistical significance after Bonferroni correction in iRBD.

To further compare the frequencies of *SNCA* variants in RBD and PD, we analyzed the Montreal PD cohort ( $n=733$ , controls  $n=6,155$ ) in the same step-wise, conditional manner in which we analyzed iRBD. Contrary to iRBD and the PD+pRBD replication cohort, the strongest associated SNP with PD in the Montreal PD cohort is downstream (3'). The top signal is rs356181 (Figure 2,  $OR=1.62$ , 95%  $CI=1.13-1.41$ ,  $p=9.9E-05$ ), which is in LD with the top PD GWAS hit, rs356182 ( $R^2=0.63$ ,  $D'=0.99$ , Figure 1A), suggesting that these associations may be driven by the same variant. After conditioning on the top signal, a

second 3' signal was identified at rs2870006 (OR=0.84, 95%CI=0.75–0.95,  $p=0.0034$ ), also in LD with previously found 3' signal at rs2870004 {Pihlstrøm, 2018 #71} ( $R^2=0.21$ ,  $D'=0.99$ ), representing the first independent replication of this association. These SNPs were not in LD with the iRBD two top signals (Table 4), and at separate location than the top iRBD signals (Figure 2), suggesting that they are independent.

### Role of *SNCA* in synucleinopathies with probable RBD.

We further examined the associations of the 5' and 3' *SNCA* SNPs that are associated with iRBD in the current study or with synucleinopathies in previous studies (Table 4) with pRBD in PD and DLB. Bonferroni corrected significance threshold was set at  $p<0.01$ . Results are detailed in Figure 3.

**5' variants**—The 5' variants associated with iRBD, DLB, and PD (rs10005233, rs7681440, and rs763443, respectively) are in high LD with each other ( $R^2>0.78$ ,  $D'>0.9$ ) and are thus represented by the RBD variant rs10005233 in this analysis. We found that 5' variants are significantly associated with synucleinopathy+pRBD (Fixed Effect [FE] meta-analysis OR=1.22, 95%CI=1.05–1.41,  $p=0.008$ , Figure 3).

**3' variants**—Two independent 3' risk variants for PD and the 3' variant associated with pRBD were analyzed. Unlike the 5' variants, these are independent signals (PD top 3', PD novel 3' and RBD 3', Figure 1A). The recently reported novel 3' risk locus for PD (PD novel 3', rs2870004) {Pihlstrøm, 2018 #71} was significantly associated with pRBD (meta-analysis OR=0.76, CI=0.62–0.94,  $p=0.009$ ), however showed no differences in allele frequencies of DLB+pRBD versus DLB-pRBD, suggesting that this may be a PD-specific signal. Results are inconclusive for the PD top 3' GWAS variant rs356182, showing discrepant distributions of allele frequencies across these cohorts (Figure 3C). Similar to the iRBD 5' risk variant, the secondary iRBD signal (RBD 3', rs11732740, Figure 1A) shows consistently increased allele frequencies in synucleinopathy+pRBD, however without reaching statistical significance ( $p=0.052$ , Figure 3B).

### *SNCA* variants and conversion to overt synucleinopathies

Next, we examined whether the top 5' and 3' *SNCA* variants for synucleinopathy risk, as well as 3' and 5' UTR variants from the sequencing data, were associated with rate of conversion, defined as time from onset or diagnosis from iRBD to diagnosis of an overt neurodegenerative disease in 237 patients who have converted. It must be noted that phenoconversion data was gathered from a cohort where 53% converted to PD, 40% to DLB, and 5% to MSA (2% to “other”), however at random a smaller number of DLB converters had genetic data available, creating a PD-skewed cohort for these analyses. For this reason, we could not perform analysis of variant association with type of conversion.

Top synucleinopathy risk variants from the GWAS data and three common 5' and 3' UTR SNPs from the sequencing data were analyzed for association with rate of conversion. These SNPs include rs10005233 (allele frequency in the converted iRBD patients = 0.57), rs11732740 (0.21), rs356182 (0.31), rs2870004 (0.23), rs2737024 (0.23), rs1045722 (0.09), rs17016074 (0.81) and rs2583986 (0.22). One variant, 5' UTR rs2583986, was associated



with faster conversion to overt synucleinopathies (Figure 4, Kaplan-Meier survival analysis  $p=0.0029$ ), when analyzing the rate of conversion from reported onset of RBD to diagnosed onset of overt synucleinopathy. Homozygous carriers of this variant ( $n=4$ ) had converted within  $3.5 \pm 1.97$  years, Heterozygous carriers ( $n=30$ ) within  $6.2 \pm 2.71$  years, and non-carriers ( $n=76$ ) had converted within  $9.4 \pm 2.05$  years. We then performed the same calculation using age at diagnosis to conversion rather than age at onset, and found that this variant lost its significance ( $p>0.05$ ). Interestingly, this variant is in LD with the top risk SNP for iRBD ( $R^2=0.38$ ,  $D'=0.99$ ), however the wild type allele (slower conversion) is correlated with the iRBD risk allele (rs10005233-T). No other *SNCA* variants were associated with rate of conversion from either onset or diagnosis of iRBD.

## DISCUSSION

In the current study we utilized the largest genetic cohort of patients with iRBD published to date, and large replication cohorts with data on pRBD, to perform a fine-mapping study of the *SNCA* locus in RBD. Our results demonstrate that there is a distinct pattern in RBD compared to the most common synucleinopathy, PD. The top association with *SNCA* in RBD is a SNP in the 5' region of the gene (RBD 5', rs10005233, Figure 1A), contrary to the 3' SNP top association with PD (PD top 3', rs356182). Additionally, this RBD 5' variant is associated with susceptibility to pRBD in PD and DLB, suggesting that the 5' associations previously reported in PD and DLB may be driven by the subgroup of patients with RBD. A secondary, independent SNP association with risk for iRBD was found in the 3' region (RBD 3' rs11732740, Figure 1A), although without statistical significance after correction for multiple comparisons. Interestingly, the PD top 3' SNP (rs356182) shows an opposite direction of effect in iRBD, further highlighting the genetic differences. A very high risk for iRBD (OR=5.7) was noted for homozygous carriers of both the RBD 5' and RBD 3' risk variants (rs10005233 and rs11732740, respectively). We also report a variant (rs2583986, which is in LD with 5' risk SNPs) that is potentially associated with rate of conversion to overt synucleinopathy, yet this association has several limitations discussed below, and requires replication.

We have identified two major associations in RBD, a primary association at the 5' side of *SNCA* (RBD 5') and a suggestive secondary association at the 3' of the gene (RBD 3'), both distinct from those identified in our cohort of PD patients (Figure 2). Interestingly, the RBD 5' SNP (rs10005233) is in nearly complete LD with the top risk locus for DLB (rs7681440, Figure 1A)<sup>17</sup> and is in LD with all 5' other risk SNPs for synucleinopathy (Figure 1A, Table 4), including the 5' risk variant for PD (PD 5', rs763443). This may suggest that all of these associations could be driven by the same variant. This hypothesis requires functional studies to identify the specific variant or combination of variants that drive these associations. Of note, it is possible that these *SNCA* 5' risk variants in PD and DLB may be driven by the sub-population of patients susceptible to RBD. This is supported by our analyses on PD and DLB cohorts for which we have data on pRBD, where 5' synucleinopathy risk alleles (rs10005233, rs763443 and rs7681440) are significantly associated with synucleinopathy +pRBD. It is important to note that the RBD 3' SNP, rs11732740 (which did not pass multiple testing correction), is not in LD with neither the PD nor DLB associated SNPs in previous GWASs, and therefore it may be specific to the RBD subtype. This RBD 3' risk

allele frequency is consistently elevated in synucleinopathy with RBD (Figure 3B), however without statistical significance ( $p=0.052$ ), therefore this association remains speculative.

The mechanism underlying the association of *SNCA* variants remains unknown. As shown in Figure 1A, the top variants associated with synucleinopathies are concentrated in promoter and regulatory regions of *SNCA*, with a strong LD block towards the 5' region. The 5' section of intron four (chr4:90737000–90743400, including RBD 5' variant rs10005233) has been identified as a conserved regulatory region containing a ~160bp CT-rich region. {Lutz, 2015 #123} This region is marked by four distinct haplotypes, one of which is associated with Lewy body pathology and increased  $\alpha$ -synuclein levels in Alzheimer's disease. {Lutz, 2015 #123} This haplotype is tagged by rs2298728-A, which is in strong LD with the RBD risk allele rs10005233-T ( $R^2=0.06$ ,  $D'=1$ ).  $R^2$  is low due to differences in allele frequencies between the two SNPs, as one SNP is more common than the other, yet the SNP with low allele frequency always appear on the same haplotype with the SNP with the higher allele frequency. Therefore, there are only three haplotypes in the European population composed of these two SNPs (rs2298728G>A and rs10005233T>C), G-C (haplotype frequency = 0.505), G-T (0.438) and A-T (0.057).. In contrast, a variant in the 3' end of the intron-four enhancer region rs356168-A has been associated with increased levels of  $\alpha$ -synuclein in human iPSCs, {Soldner, 2016 #122} and is in LD with the top signal for PD rs356182-A ( $R^2=0.50$ ,  $D'=0.91$ ). Finally, the RBD 5' variant rs10005233 has been linked to novel alternative 3'-end *SNCA* isoforms (PB.1016.383, PB.1016.384), associated with a truncated open reading frame prediction. {Tseng, 2019 #124} These findings suggest potential functional effects of some of these variants, but they need to be further replicated and studied.

The independent associations of the 5' and 3' variants across the different synucleinopathies may suggest phenotype-specific effects in the regulatory regions on opposing ends of the gene. Although the pathogenic mechanisms of the *SNCA* gene and its encoded protein  $\alpha$ -synuclein in synucleinopathies are still not fully understood, our findings suggest that the 5' region of *SNCA* might affect cognitive components of synucleinopathy. This region is associated with PDD, DLB and ADLBV, which are characterized by cognitive impairment, and now also with RBD, which is known to be the strongest risk factor for rapid and severe cognitive decline in PD. {Vendette, 2007 #117} The differential effects of the 5' and 3' SNPs may be due to differential effects in different parts of the brain, and future functional and clinical studies of these regulatory regions will be essential for understanding the pathological mechanisms underlying RBD and these synucleinopathies. Of note, it is possible that the effects of the different 5' and 3' alleles on the different phenotypes will allow for future stratification for basic and clinical studies based on genetic data. The frequencies of the PD 3' SNP (rs356182) and the RBD 5' SNP (rs10005233) are >0.3, common enough for stratification of patient cohorts, which may uncover additional correlations with pathological, clinical and perhaps therapeutic effects.

Previous genetic findings also supported an only partially overlapping genetic background between iRBD, PD and DLB. {Dauvilliers, 2018 #110} While *GBA* variants, implicated in both PD {Gan-Or, 2008 #118} and DLB, {Guerreiro, 2018 #99; Clark, 2009 #119} were also strongly associated with RBD, {Gan-Or, 2015 #52} the PD-causing *LRKK2* mutations and

the DLB-associated *APOE*  $\epsilon$ 4 allele were not associated with RBD. {Gan-Or, 2017 #48; Fernández-Santiago, 2016 #111} The H1/H2 *MAPT* haplotypes, associated with PD and other neurodegenerative disorders, were not associated with RBD either. {Li, 2018 #95} More recently, it was demonstrated that the PD-associated *TMEM175* coding variant p.M393T was associated with RBD and potentially affects the activity of the lysosomal enzyme glucocerebrosidase (encoded by *GBA*). {Jinn, 2019 #120; Krohn, #128} Therefore, it is possible that some genetic variants are relevant for all types of synucleinopathy, with and without RBD, while other variants are specifically relevant for RBD-associated synucleinopathy. Currently, the division between PD and DLB is somewhat arbitrary, determined by a cut-off of one year between the onset of parkinsonism and the onset of dementia. {Lippa, 2007 #121} In late stages of their disease, patients with RBD who converted to PD first and later on developed dementia, and RBD patients who first presented with dementia and later with parkinsonism, can eventually have an undistinguishable clinical presentation. It is possible that in the future this division will be based on genetic background and/or the molecular mechanism involved, such as “*GBA*-associated synucleinopathy”. This would be especially true if treatments that target the underlying genetic cause would be available.

There are several limitations to the current study. One limitation is sample size, as the PD meta-analyses are much larger than the iRBD cohort studied here, which makes it possible that smaller effect size in the *SNCA* locus were not detected in the iRBD cohort. However, with the world’s largest iRBD cohort including more than 1000 cases with available genetic data, this study is comparable to previous DLB studies {Guerreiro, 2018 #99; Guerreiro, 2016 #97} and is sufficiently powered to detect strong associations between *SNCA* and RBD risk. A second limitation lies in the analysis of RBD conversion, which is based on a small number of iRBD patients who had converted to overt synucleinopathies, therefore underpowered to reach corrected statistical significance. Furthermore, the time from onset or diagnosis to conversion is likely an inaccurate estimate of disease duration. Onset of RBD is based on patients’ report, who can be unaware of the actual time when RBD initially presented. Diagnosis time is also not necessarily a good indicator for disease duration, as patients can contact physicians many years after disease onset, depending on whether or not the RBD symptoms disturb the patients or their spouses. Therefore, the results presented here on rate of conversion should be taken as preliminary and with caution. Future studies on larger datasets with adjustments for other clinical variables will enable a more accurate study of the effect of genetics on rate of RBD conversion.

Further studies of RBD genetics are of great importance. As clinical trials in PD have repeatedly failed, it is possible that performing studies on prodromal patients such as iRBD would increase the chances of success in clinical trials since the neurons carry less damage and may therefore be more responsive to treatment. Understanding the underlying genetics of iRBD will enable genetic stratification of patients and may potentially help identify individuals at risk for iRBD and overt synucleinopathies at an earlier stage. Furthermore, genetics can provide targets for drug development (e.g. *GBA* and *LRRK2* in PD) and drive molecular and cellular studies to understand the underlying mechanisms of RBD and synucleinopathies.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Authors

Lynne Krohn, MSc<sup>1,2</sup>, Richard YJ Wu, BSc (Hons)<sup>1,3</sup>, Karl Heilbron, PhD<sup>4</sup>, Jennifer A. Ruskey, MSc<sup>2,5</sup>, Sandra B. Laurent, BTS<sup>2,5</sup>, Cornelis Blauwendraat, PhD<sup>6</sup>, Armaghan Alam, BSc<sup>1,2</sup>, Isabelle Arnulf, MD, PhD<sup>7</sup>, Michele T.M. Hu, MBBS, FRCP, PhD<sup>8,9</sup>, Yves Dauvilliers, MD, PhD<sup>10</sup>, Birgit Högl, MD<sup>11</sup>, Mathias Toft, MD, PhD<sup>12,13</sup>, Kari Anne Bjørnarå, MD, PhD<sup>12</sup>, Ambra Stefani, MD<sup>11</sup>, Evi Holzknecht, MD<sup>11</sup>, Christelle Charley Monaca, MD, PhD<sup>14</sup>, Abril Beatriz, MD<sup>15</sup>, Giuseppe Plazzi, MD<sup>16,17</sup>, Elena Antelmi, MD, PhD<sup>16,17</sup>, Luigi Ferini-Strambi, MD<sup>18</sup>, Peter Young, MD<sup>19</sup>, Anna Heidebreder, MD<sup>19</sup>, Valérie Cochen De Cock, MD, PhD<sup>20,21</sup>, Brit Mollenhauer, MD<sup>22,23</sup>, Friederike Sixel-Döring, MD<sup>22,23</sup>, Claudia Trenkwalder, MD<sup>22,23</sup>, Karel Sonka, MD<sup>24</sup>, David Kemlink, MD, PhD<sup>24</sup>, Michela Figorilli, MD, PhD<sup>25</sup>, Monica Puligheddu, MD, PhD<sup>25</sup>, Femke Dijkstra, MD<sup>26,27</sup>, Mineke Viaene, MD, PhD<sup>26,27</sup>, Wolfgang Oertel, MD<sup>28</sup>, Marco Toffoli, MD<sup>29,30</sup>, Gian Luigi Gigli, MD<sup>31,32</sup>, Mariarosaria Valente, MD<sup>29,31</sup>, Jean-François Gagnon, PhD<sup>33,34</sup>, Mike A. Nalls, PhD<sup>8,35</sup>, Andrew B. Singleton, PhD<sup>8</sup>, 23andMe Research Team<sup>4</sup>, Alex Desautels, MD, PhD<sup>33,34</sup>, Jacques Y. Montplaisir, MD, PhD<sup>33,36</sup>, Paul Cannon, PhD<sup>4</sup>, Owen A. Ross, PhD<sup>37</sup>, Bradley F. Boeve, MD<sup>38</sup>, Nicolas Dupré, MD, MSc<sup>39,40</sup>, Edward A. Fon, MD, FRCPC<sup>2</sup>, Ronald B. Postuma, MD, MSc<sup>2,5,33</sup>, Lasse Pihlstrøm, MD, PhD<sup>12</sup>, Guy A. Rouleau, MD, PhD, FRCPC<sup>1,2,5</sup>, Ziv Gan-Or, MD, PhD<sup>1,2,5</sup>

## Affiliations

- <sup>1</sup>Department of Human Genetics, McGill University, Montréal, QC, Canada.
- <sup>2</sup>Montreal Neurological Institute, McGill University, Montréal, QC, Canada.
- <sup>3</sup>Department of Medicine, Imperial College London, London, United Kingdom.
- <sup>4</sup>23andMe, Inc., 899 W Evelyn Avenue, Mountain View, California 94041 USA.
- <sup>5</sup>Department of Neurology and Neurosurgery, McGill University, Montréal, QC, Canada.
- <sup>6</sup>Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA.
- <sup>7</sup>Sleep Disorders Unit, Pitié Salpêtrière Hospital, Centre de Recherche de l'Institut du Cerveau et de la Moelle Epinière and Sorbonne Universities, Paris, France.
- <sup>8</sup>Oxford Parkinson's Disease Centre (OPDC), University of Oxford, Oxford, United Kingdom.
- <sup>9</sup>Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, United Kingdom.

10. National Reference Center for Narcolepsy, Sleep Unit, Department of Neurology, Gui-de-Chauliac Hospital, CHU Montpellier, University of Montpellier, Inserm U1061, Montpellier, France.
11. Sleep Disorders Clinic, Department of Neurology, Medical University of Innsbruck, Innsbruck, Austria.
12. Department of Neurology, Oslo University Hospital, Oslo, Norway.
13. Institute of Clinical Medicine, University of Oslo, Oslo, Norway.
14. University Lille North of France, Department of Clinical Neurophysiology and Sleep Center, CHU Lille, Lille, France.
15. Sleep disorder Unit, Carémeau Hospital, University Hospital of Nîmes, France.
16. Department of Biomedical and Neuromotor Sciences (DIBINEM), Alma Mater Studiorum, University of Bologna, Bologna, Italy.
17. IRCCS, Istituto delle Scienze Neurologiche, Bologna, Italy.
18. Department of Neurological Sciences, Università Vita-Salute San Raffaele, Milan, Italy.
19. Department of Sleep Medicine and Neuromuscular Disorders, University of Muenster, Germany.
20. Sleep and Neurology Unit, Beau Soleil Clinic, Montpellier, France.
21. EuroMov, University of Montpellier, Montpellier, France.
22. Paracelsus-Elena-Klinik, Kassel, Germany.
23. Department of Neurosurgery, University Medical Centre Goettingen, Goettingen, Germany.
24. Department of Neurology and Centre of Clinical Neuroscience, Charles University, First Faculty of Medicine and General University Hospital, Prague, Czech Republic.
25. Department of Medical Sciences and Public Health, Sleep Disorder Research Center, University of Cagliari, Cagliari, Italy.
26. Laboratory for Sleep Disorders, St. Dimpna Regional Hospital, Geel, Belgium.
27. Department of Neurology, St. Dimpna Regional Hospital, Geel, Belgium.
28. Department of Neurology, Philipps University, Marburg, Germany.
29. Department of Medicine (DAME), University of Udine, Udine, Italy.
30. Department of Clinical and Movement Neurosciences, UCL Queen Square Institute of Neurology, London, UK.
31. Clinical Neurology Unit, Department of Neurosciences, University Hospital of Udine, Udine, Italy.
32. DMIF, University of Udine, Udine, Italy.

33. Centre d'Études Avancées en Médecine du Sommeil, Hôpital du Sacré-Cœur de Montréal, Montréal, QC, Canada.
34. Department of Neurosciences, Université de Montréal, Montréal, QC, Canada.
35. Data Tecnica International, Glen Echo, MD, USA.
36. Department of Psychiatry, Université de Montréal, Montréal, QC, Canada.
37. Departments of Neuroscience and Clinical Genomics, Mayo Clinic, Jacksonville, FL, USA.
38. Department of Neurology, Mayo Clinic, Rochester, MN, USA.
39. Division of Neurosciences, CHU de Québec, Université Laval, Québec City, QC, Canada;
40. Department of Medicine, Faculty of Medicine, Université Laval, Québec City, QC, Canada.

## Acknowledgements

We thank the patients and control subjects for their participation in this study. This work was financially supported by Parkinson's Society Canada, the Michael J. Fox Foundation, the Canadian Consortium on Neurodegeneration in Aging (CCNA), the Canadian Glycomics Network (GlycoNet) and the Canada First Research Excellence Fund (CFREF), awarded to McGill University for the Healthy Brains for Healthy Lives (HBHL) program. JFG holds a Canada Research Chair in Cognitive Decline in Pathological Aging. GAR holds a Canada Research Chair in Genetics of the Nervous System and the Wilder Penfield Chair in Neurosciences. WO is Hertie Senior Research Professor, supported by the Charitable Hertie Foundation, Frankfurt/Main, Germany. EAF holds a Canada Research Chair (Tier 1) in Parkinson Disease. ZGO is supported by the Fonds de recherche du Québec - Santé (FRQS) Chercheurs-boursiers award, and is a Parkinson's Disease Canada New Investigator awardee. The access to part of the participants for this research has been made possible thanks to the Quebec Parkinson's Network (<http://rpq-qpn.ca/en/>). We thank Daniel Rochefort, Helene Catoire and Vessela Zaharieva for their assistance. Mayo Clinic is supported in part by the Mangurian Foundation Lewy Body Dementia Program, The Little Family Foundation, an American Parkinson Disease Association (APDA) Mayo Clinic Information and Referral Center, an APDA Center for Advanced Research and the Mayo Clinic Lewy Body Dementia Association (LBDA) Research Center of Excellence.

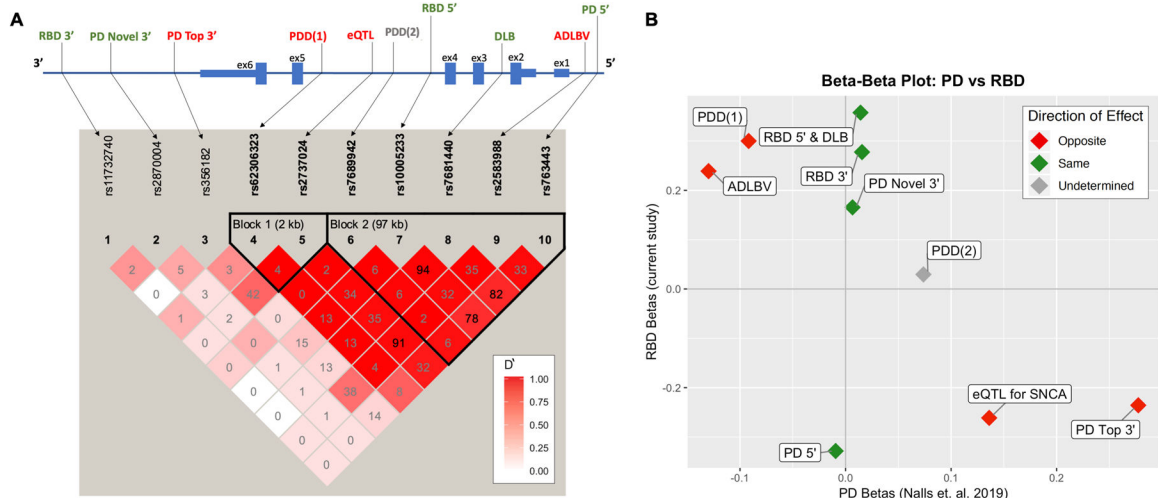
## REFERENCES

- Haba-Rubio J et al. Prevalence and determinants of rapid eye movement sleep behavior disorder in the general population. *Sleep* 41, zsx197 (2017).
- Pujol M et al. Idiopathic REM sleep behavior disorder in the elderly Spanish community: a primary care center study with a two-stage design using video-polysomnography. *Sleep medicine* 40, 116–121 (2017). [PubMed: 29042180]
- Kang S-H et al. REM sleep behavior disorder in the Korean elderly population: prevalence and clinical characteristics. *Sleep* 36, 1147–1152 (2013). [PubMed: 23904674]
- Fereshtehnejad S-M et al. New clinical subtypes of Parkinson disease and their longitudinal progression: a prospective cohort comparison with other phenotypes. *JAMA neurology* 72, 863–873 (2015). [PubMed: 26076039]
- Iranzo A, Santamaria J & Tolosa EJTLN Idiopathic rapid eye movement sleep behaviour disorder: diagnosis, management, and the need for neuroprotective interventions. *15*, 405–419 (2016).
- Iranzo A et al. Neurodegenerative disease status and post-mortem pathology in idiopathic rapid-eye-movement sleep behaviour disorder: an observational cohort study. *The Lancet Neurology* 12, 443–453 (2013). [PubMed: 23562390]
- Postuma R et al. Quantifying the risk of neurodegenerative disease in idiopathic REM sleep behavior disorder. *Neurology* 72, 1296–1300 (2009). [PubMed: 19109537]

8. Postuma RB et al. Rapid eye movement sleep behavior disorder and risk of dementia in Parkinson's disease: a prospective study. *Movement disorders* 27, 720–726 (2012). [PubMed: 22322798]
9. Schenck CH, Boeve BF & Mahowald M.W.J.S.m. Delayed emergence of a parkinsonian disorder or dementia in 81% of older men initially diagnosed with idiopathic rapid eye movement sleep behavior disorder: a 16-year update on a previously reported series. *Sleep Medicine* 14, 744–748 (2013). [PubMed: 23347909]
10. Postuma RB & Berg D Advances in markers of prodromal Parkinson disease. *Nature Reviews Neurology* 12, 622 (2016). [PubMed: 27786242]
11. Högl B, Stefani A & Videnovic A Idiopathic REM sleep behaviour disorder and neurodegeneration —an update. *Nature Reviews Neurology* 14, 40 (2018). [PubMed: 29170501]
12. Dauvilliers Y et al. REM sleep behaviour disorder. *Nature Reviews Disease Primers* 4, 19 (2018).
13. Gan-Or Z et al. GBA mutations are associated with rapid eye movement sleep behavior disorder. *Annals of clinical and translational neurology* 2, 941–945 (2015). [PubMed: 26401515]
14. Bencheikh BOA et al. LRRK2 protective haplotype and full sequencing study in REM sleep behavior disorder. *Parkinsonism & related disorders* (2018).
15. Fernández-Santiago R et al. Absence of LRRK2 mutations in a cohort of patients with idiopathic REM sleep behavior disorder. *Neurology* 86, 1072–1073 (2016). [PubMed: 26747879]
16. Li J et al. Full sequencing and haplotype analysis of MAPT in Parkinson's disease and rapid eye movement sleep behavior disorder. *Movement Disorders* (2018).
17. Gan-Or Z et al. The dementia-associated APOE  $\epsilon$ 4 allele is not associated with rapid eye movement sleep behavior disorder. *Neurobiology of Aging* 49, 218. e13–218. e15 (2017).
18. Kim WS, Kågedal K & Halliday GM Alpha-synuclein biology in Lewy body diseases. *Alzheimer's research & therapy* 6, 73 (2014).
19. Chang D et al. A meta-analysis of genome-wide association studies identifies 17 new Parkinson's disease risk loci. *Nature genetics* 49, 1511 (2017). [PubMed: 28892059]
20. Nalls MA et al. Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. *Nature genetics* 46, 989 (2014). [PubMed: 25064009]
21. Guerreiro R et al. Investigating the genetic architecture of dementia with Lewy bodies: a two-stage genome-wide association study. *The Lancet Neurology* 17, 64–74 (2018). [PubMed: 29263008]
22. Bjørnarå KA, Pihlstrøm L, Dietrichs E & Toft M Risk variants of the  $\alpha$ -synuclein locus and REM sleep behavior disorder in Parkinson's disease: a genetic association study. *BMC neurology* 18, 20 (2018). [PubMed: 29466944]
23. Guella I et al.  $\alpha$ -synuclein genetic variability: A biomarker for dementia in Parkinson disease. *Annals of neurology* 79, 991–999 (2016). [PubMed: 27091628]
24. Linnertz C et al. The genetic contributions of SNCA and LRRK2 genes to Lewy Body pathology in Alzheimer's disease. *Human molecular genetics* 23, 4814–4821 (2014). [PubMed: 24777780]
25. Sailer A et al. A genome-wide association study in multiple system atrophy. *Neurology* 87, 1591–1598 (2016). [PubMed: 27629089]
26. Awadalla P et al. Cohort profile of the CARTaGENE study: Quebec's population-based biobank for public health and personalized genomics. *International journal of epidemiology* 42, 1285–1299 (2012). [PubMed: 23071140]
27. Thorpy MJ Classification of sleep disorders. *Neurotherapeutics* 9, 687–701 (2012). [PubMed: 22976557]
28. Hughes AJ, Daniel SE, Kilford L & Lees AJ Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. *Journal of Neurology, Neurosurgery & Psychiatry* 55, 181–184 (1992).
29. Postuma RB et al. MDS clinical diagnostic criteria for Parkinson's disease. *Movement Disorders* 30, 1591–1601 (2015). [PubMed: 26474316]
30. Marek K et al. The parkinson progression marker initiative (PPMI). *Progress in neurobiology* 95, 629–635 (2011). [PubMed: 21930184]
31. Nomura T, Inoue Y, Kagimura T, Uemura Y & Nakashima K.J.S.m. Utility of the REM sleep behavior disorder screening questionnaire (RBDSQ) in Parkinson's disease patients. *Sleep Medicine* 12, 711–713 (2011). [PubMed: 21700495]

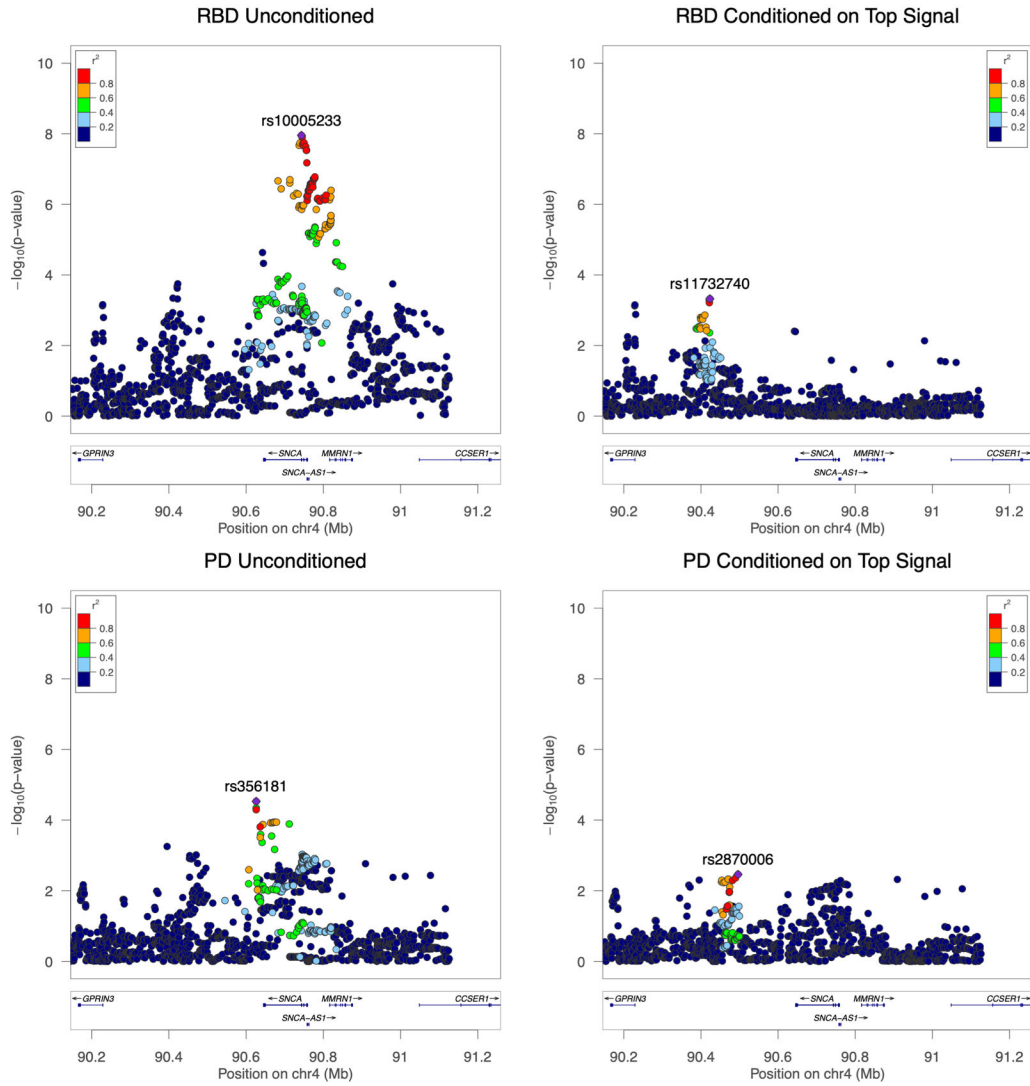
32. Postuma RB et al. A single-question screen for rapid eye movement sleep behavior disorder: a multicenter validation study. *Movement Disorders* 27, 913–916 (2012). [PubMed: 22729987]
33. Purcell S et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics* 81, 559–575 (2007). [PubMed: 17701901]
34. Pihlstrøm L et al. A comprehensive analysis of SNCA-related genetic risk in sporadic parkinson disease. *Annals of neurology* 84, 117–129 (2018). [PubMed: 30146727]
35. McCarthy S et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nature Genetics* 48, 1279 (2016). [PubMed: 27548312]
36. Ross JP et al. Analysis of DNAJC13 mutations in French-Canadian/French cohort of Parkinson's disease. *Neurobiology of aging* 45, 212. e13–212. e17 (2016).
37. McKenna A et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research* 20, 1297–1303 (2010). [PubMed: 20644199]
38. Wang K, Li M & Hakonarson H ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic acids research* 38, e164–e164 (2010). [PubMed: 20601685]
39. Lek M et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 536, 285 (2016). [PubMed: 27535533]
40. Lill CM et al. Comprehensive research synopsis and systematic meta-analyses in Parkinson's disease genetics: The PDGene database. *PLoS genetics* 8, e1002548 (2012). [PubMed: 22438815]
41. Cordell HJ & Clayton DG JTAJo.H.G. A unified stepwise regression procedure for evaluating the relative effects of polymorphisms within a gene using case/control or family data: application to HLA in type 1 diabetes. *American Journal of Human Genetics* 70, 124–141 (2002). [PubMed: 11719900]
42. Barrett JC, Fry B, Maller J & Daly M JJB Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21, 263–265 (2004). [PubMed: 15297300]
43. Viechtbauer W Conducting meta-analyses in R with the metafor package. *J Stat Softw* 36, 1–48 (2010).
44. Lee S, Wu MC & Lin XJB Optimal tests for rare variant effects in sequencing association studies. *Biostatistics* 13, 762–775 (2012). [PubMed: 22699862]
45. Nalls MA et al. Parkinson's disease genetics: Identifying novel risk loci, providing causal insights and improving estimates of heritable risk. *BioRxiv*, 388165 (2018).
46. Lutz MW et al. A cytosine-thymine (CT)-rich haplotype in intron 4 of SNCA confers risk for Lewy body pathology in Alzheimer's disease and affects SNCA expression. *Alzheimer's & Dementia* 11, 1133–1143 (2015).
47. Soldner F et al. Parkinson-associated risk variant in distal enhancer of  $\alpha$ -synuclein modulates target gene expression. *Nature* 533, 95 (2016). [PubMed: 27096366]
48. Tseng E et al. The landscape of SNCA transcripts across synucleinopathies: New insights from long reads sequencing analysis. *bioRxiv*, 524827 (2019).
49. Vendette M et al. REM sleep behavior disorder predicts cognitive impairment in Parkinson disease without dementia. *Neurology* 69, 1843–1849 (2007). [PubMed: 17984452]
50. Gan-Or Z et al. Genotype-phenotype correlations between GBA mutations and Parkinson disease risk and onset. *Neurology* 70, 2277–2283 (2008). [PubMed: 18434642]
51. Clark LN et al. Association of glucocerebrosidase mutations with dementia with lewy bodies. *Archives of neurology* 66, 578–583 (2009). [PubMed: 19433657]
52. Jinn S et al. Functionalization of the TMEM175 p. M393T Variant as a risk factor for Parkinson Disease. *Human molecular genetics* (2019).
53. Lippa C et al. DLB and PDD boundary issues: diagnosis, treatment, molecular pathology, and biomarkers. *Neurology* 68, 812–819 (2007). [PubMed: 17353469]
54. Guerreiro R et al. Genome-wide analysis of genetic correlation in dementia with Lewy bodies, Parkinson's and Alzheimer's diseases. *Neurobiology of Aging* 38, 214. e7–214. e10 (2016).





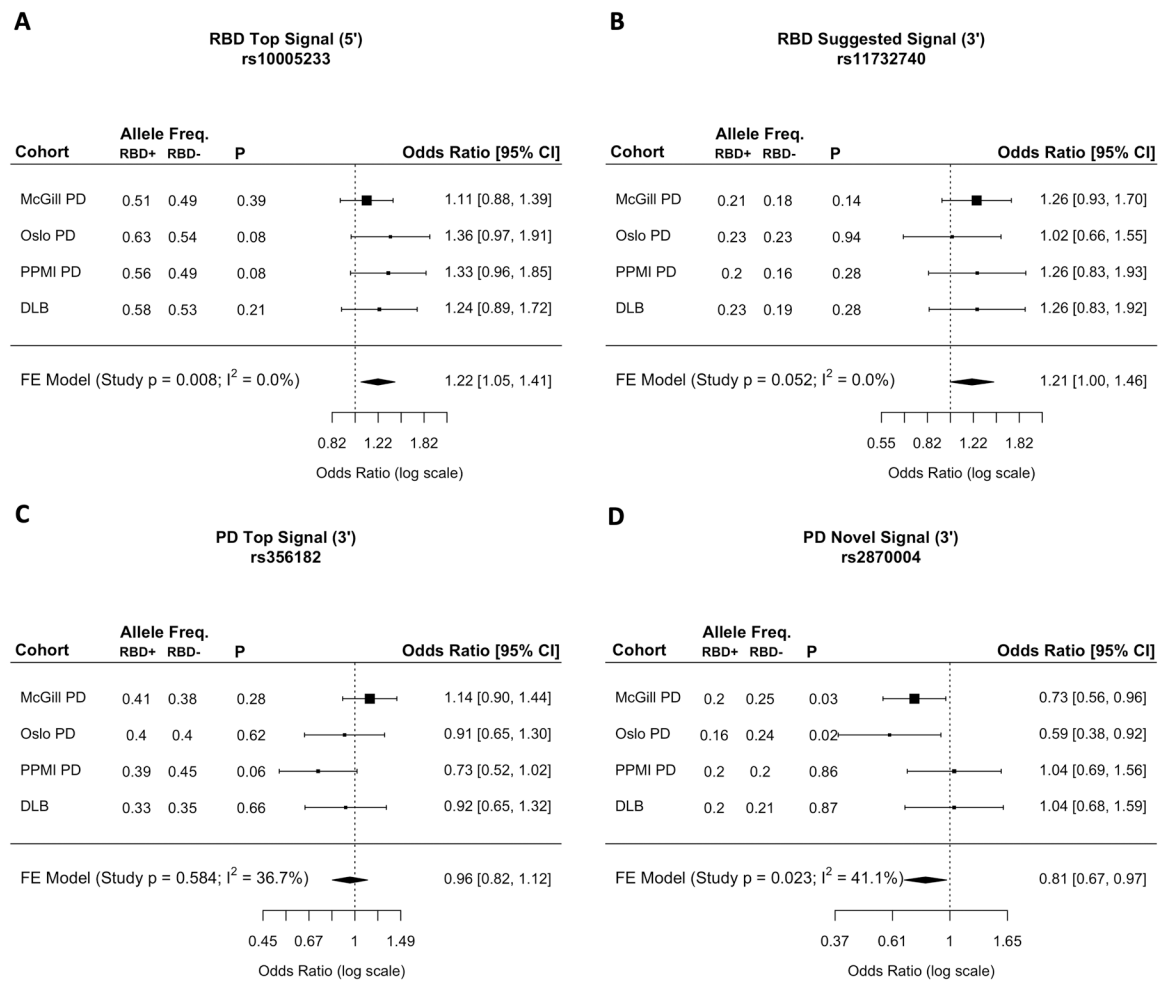
**Figure 1. Schematic of top *SNCA* genetic signals across synucleinopathies with a PD to RBD comparison.**

**A)** A schematic of the *SNCA* region representing the location of top signals for synucleinopathies and linkage disequilibrium (LD) plot by Haploview. In the LD plot, numbers represent  $R^2$  values and deeper color shade represents strength of  $D'$ . Of note, when  $R^2$  is low and  $D'$  is high, variants are still in strong LD, as the low  $R^2$  is a result of differences in allele frequencies, yet high  $D'$  means that the less common SNP is in most cases (or all cases if  $D'=1$ ) appears on the same allele as the more common SNP. Top RBD signal rs10005233 is in moderate to high LD with all 5' *SNCA* SNPs associated with synucleinopathy. The second RBD signal, 3' rs11732740, is independent and not in LD with previously reported 3' or 5' risk variants. Synucleinopathy risk variants are located in the promoter and regulatory regions, with a strong LD block in the 5' end. **B)** A beta-beta plot comparing *SNCA* betas found in the current study (y-axis) to betas from the latest PD GWAS (x-axis). Points in red represent different direction of effect between iRBD and PD, while green represents effect in the same direction. Interestingly, the strongest risk variants for PD at the 3' of the gene are less common in RBD (marked as “PD Top 3” and “eQTL for *SNCA*”).



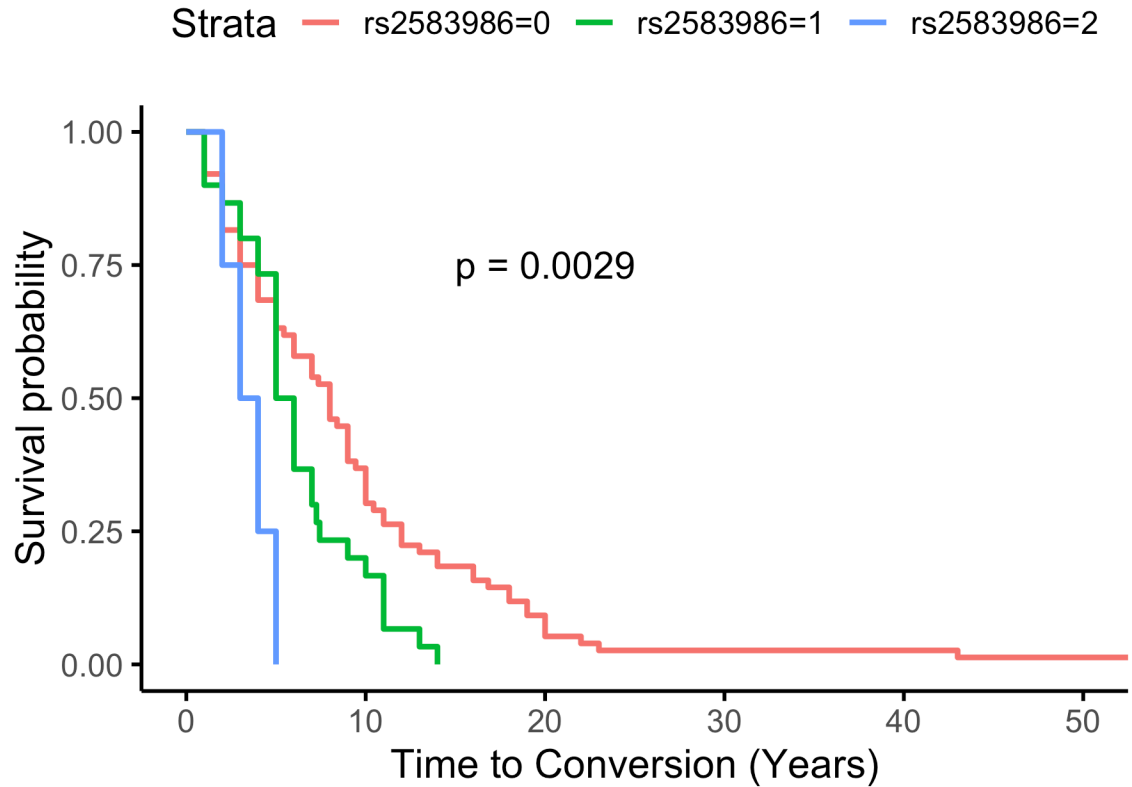
**Figure 2. The *SNCA* locus with association results for RBD and PD.**

Regional LocusZoom Manhattan plots show the top iRBD risk association signal (top left) and the second highest iRBD signal (top right), conditioned on the first. The bottom panels show the top PD hits in our PD cohort, demonstrating the different pattern of association between iRBD and PD in the *SNCA* locus. Conditional analysis assures independence of association signals, as shown by the disappearance of the LD block with the RBD top hit rs10005233 on the conditioned plot on the top right. In both iRBD plots, a less significant but notable signal is shown in the region of reported top hit for PD, rs356182, with an opposite direction of effect than the PD signal (OR=0.79, 95% CI=0.70–0.91,  $p=6e-04$ , in iRBD vs Controls).



**Figure 3. Association between synucleinopathy risk loci in SNCA with pRBD in PD and DLB.** The forest plots represent the odds ratios and 95% confidence intervals for association of top synucleinopathy risk loci in synucleinopathy with pRBD (PD+pRBD and DLB+pRBD) versus synucleinopathy without RBD (PD-pRBD and DLB-pRBD), as well as results from fixed effect meta-analyses. **A)** The RBD top hit rs10005233 represents all synucleinopathy 5' variants (rs7681440 and rs763443) because of high LD. 5' risk variants are associated with synucleinopathies+pRBD, possibly suggesting an RBD-driven signal. **B)** The RBD potential secondary signal shows a trend towards association to synucleinopathy with pRBD, without statistical significance, yet marginal ( $p=0.052$ ). **C)** The PD top GWAS hit rs356182 shows no apparent associations to synucleinopathy with or without pRBD. **D)** Secondary 3' PD variant rs2870004 is significantly associated with PD without pRBD, suggesting a PD specific signal (FE meta-analysis excluding DLB: OR = 0.76, CI=0.62–0.94,  $p=0.009$ ). This pattern is not present in DLB, thus is not significant in synucleinopathies overall. Freq.: frequency. OR: odds ratio. CI: confidence interval. FE: fixed effect. pRBD: probable RBD. Syn: Synucleinopathy.

## Time to Conversion to Synucleinopathy Kaplan-Meier Survival Plot



**Figure 4. The 5' UTR SNP rs2583986 and conversion to overt synucleinopathy.**

Kaplan-Meier survival plot demonstrating a possible effect of rs2583986 genotype of conversion from iRBD to overt synucleinopathy. Wild type carriers (n=76), had mean time to conversion (TTC) =  $9.4 \pm 2.05$  years, heterozygous carriers (n=30), had mean TTC =  $6.2 \pm 2.71$  years; and homozygous carriers (n=4), had mean TTC =  $3.5 \pm 1.97$  years. These results should be interpreted with much caution due to the limitations mentioned in the discussion, including the small numbers and the lack of confidence in patients' report on RBD onset. Of note, when we performed the same analysis from diagnosis of iRBD to conversion, there was no association between this SNP and conversion.

**Table 1.**

iRBD cohort information by centre.

Centre	n	Mean Age at Enrollment (n of available data)	Mean Age at RBD Onset (n of available data)	Mean Disease Duration (n of available data)	% Male (n available data)	n of Available Conversion data (n converted)
Nîmes, France	5	66.6 ± 5.4 (5)	65 ± 5.7 (4)	1.3 ± 0.5 (4)	80 (5)	5 (0)
Muenster, Germany	23	66.5 ± 8.2 (23)	62.2 ± 8.4 (18)	5.1 ± 3.6 (18)	82.6 (23)	23 (10)
Bologna, Italy	29	70.1 ± 12.2 (29)	58.6 ± 11.4 (28)	11.9 ± 8.3 (28)	79.3 (29)	19 (5)
Paris, France	220	67.7 ± 8.2 (217)	62.3 ± 8.5 (52)	6.8 ± 5.6 (52)	76.8 (220)	70 (50)
Kassel, Germany	28	67.8 ± 11.9 (28)	60.4 ± 8.6 (27)	9.1 ± 5.3 (27)	60.7 (28)	13 (3)
Montpellier, France (Beau Soleil Clinic)	26	67.9 ± 10.6 (26)	61.7 ± 10.1 (20)	7.8 ± 4.9 (20)	73.1 (26)	25 (10)
Montpellier, France (CHU Montpellier)	96	68.1 ± 8.5 (96)	64.6 ± 12.9 (4)	11.3 ± 8.3 (4)	79.2 (96)	23 (23)
Innsbruck, Austria	80	67 ± 10.3 (69)	61 ± 11.9 (35)	4.7 ± 6.4 (35)	87.5 (80)	28 (16)
Prague, Czech Republic	47	69.1 ± 7.5 (47)	61.5 ± 8.8 (46)	7.7 ± 4.6 (46)	91.5 (47)	26 (8)
Milan, Italy	19	69.6 ± 5.4 (19)	63.5 ± 5.8 (19)	6.1 ± 4.8 (19)	84.2 (19)	19 (10)
Udine, Italy	83	75.8 ± 7.2 (83)	61.9 ± 10.9 (75)	13.5 ± 8.4 (75)	78.3 (83)	0 (0)
Cagliari, Italy	28	71.2 ± 7.9 (28)	62.8 ± 9 (28)	8.4 ± 5.2 (28)	78.6 (28)	28 (10)
Lille, France	23	67.2 ± 6.2 (23)	63.8 ± 6.5 (20)	4.8 ± 3.7 (20)	87 (20)	15 (13)
Montreal, Canada	147	66.6 ± 9.1 (147)	57.5 ± 10.7 (107)	10.4 ± 8.6 (107)	74.8 (147)	84 (60)
Oxford, UK	184	65.7 ± 9.2 (184)	60 ± 9.2 (145)	7 ± 5.5 (145)	89.5 (184)	46 (14)
Geel, Belgium	9	59.3 ± 10.5 (9)	52.4 ± 11.1 (9)	6.9 ± 2.5 (9)	77.8 (9)	8 (5)
Marburg, Germany	29	NA (0)	NA (0)	NA (0)	93.1 (29)	0 (0)
<b>Total</b>	<b>1076</b>	<b>68 ± 9.2</b>	<b>60.6 ± 9.8</b>	<b>8.4 ± 6.9</b>	<b>80.9</b>	<b>432 (237)</b>

**Table 2.***SNCA* variants associated with REM Sleep Behavior Disorder.

SNP	rs10005233	rs11732740
Alleles, effect/reference	T/C	G/A
GRCh37/hg19 base position	4:90743331	4:90423029
<b>Analysis, n cases/controls = 1,031/5,864</b>		
Effect allele frequency cases/controls	0.58 / 0.49	0.21 / 0.17
Phred-Scaled CADD	4.71	4.67
<b>Stepwise Conditional Analysis</b>		
Conditioned on	NA	rs10005233
OR (95% CI)	1.43 (1.27 – 1.62)	1.32 (1.13 – 1.53)
<i>p</i>	1.11E-08	0.00047
<b>Combined Model</b>		
OR (95% CI)	1.42 (1.26–1.61)	1.32 (1.13–1.53)
<i>p</i>	2.73E-08	0.0005

SNP: single nucleotide polymorphism. CADD: combined annotation dependent depletion. OR: odds ratio. CI: confidence interval. NA: not applicable.

**Table 3.**Gradual increase in risk for iRBD in carriers of the top 5' and 3' *SNCA* variants.

Number of alleles with risk SNPs in rs10005233 and in rs11732740 <sup>a</sup>	% of carriers in RBD (n=1,076)	% of carriers in controls (n=5,897)	OR (95% CI) <sup>b</sup>	p-value <sup>b</sup>
0	12.6%	18.6%	reference	reference
1	33.6%	41.4%	1.40 (1.08–1.81)	0.012
2	39.0%	30.8%	2.08 (1.60–2.70)	3.6E-08
3	12.6%	8.5%	2.45 (1.75–3.42)	1.75E-07
4	2.0%	0.8%	5.74 (2.81–11.72)	2E-06

<sup>a</sup>0 – Non-carriers of both risk SNPs; 1 – Carriers of one risk variant, either in rs10005233 or rs11732740; 2 – Carriers of two risk variants, either one in each SNP, or homozygous carriers of one of the SNPs and non-carriers of the second SNP; 3 – Homozygous carriers of one of the risk SNPs and heterozygous carriers of the second risk SNP; 4 – Homozygous carriers of both risk SNPs

<sup>b</sup>Logistic regression adjusted for covariates as detailed in methods.

iRBD, isolated REM-sleep behavior disorder; OR, odds ratio; CI, confidence interval.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 4.**

*SNCA* SNPs previously associated with different forms of synucleinopathies and their association with RBD in the current study.

SNP	Reference	EA	EAF RBD/ controls	LD (R <sup>2</sup> /D')		Association with RBD in the current study	
				rs10005233 (5')	rs11732740 (3')	OR (95% CI)	<i>p</i>
<i>Parkinson's disease</i>							
<b>rs356182</b>	Top GWAS hit (3') Chang et. al. 2017	G	0.30 / 0.34	0.20 / 0.60	0.00 / 0.14	0.79 (0.70 – 0.91)	6E-04
rs763443	GWAS signal (5') Nalls, et. al. 2014	C	0.58 / 0.50	0.79 / 0.91	0.01 / 0.23	1.38 (1.22 – 1.57)	4E-07
rs2870004	Novel risk locus (3') Pihlström et. al. 2018	T	0.23 / 0.225	0.07 / 0.26	0.04 / 0.81	1.18 (1.022 – 1.37)	0.024
<b>rs2737024</b>	eQTL for <i>SNCA</i> <i>McClymont et. al. 2018</i>	G	0.22 / 0.27	0.42 / 1.0	0.01 / 0.32	0.77 (0.66 – 0.89)	4E-04
<i>Parkinson's disease with dementia</i>							
<b>rs7689942</b>	Guella et. al. 2016	T	0.06 / 0.06	0.07 / 1.0	0.00 / 0.27	1.03 (0.80 – 1.33)	0.80
<b>rs62306323</b>	Guella et. al. 2016	T	0.14 / 0.12	0.15 / 1.0	0.00 / 0.20	1.35 (1.13 – 1.63)	0.001
<i>Dementia with Lewy bodies</i>							
rs7681440	Top GWAS hit (5') Guerreiro et. al. 2018	G	0.58 / 0.50	0.94 / 0.99	0.00 / 0.00	1.42 (1.25 – 1.61)	3E-08
<i>Alzheimer's disease with Lewy body pathology</i>							
rs2583988	Linnertz et. al. 2014	C	0.78 / 0.74	0.40 / 0.99	0.01 / 0.29	1.27 (1.10 – 1.48)	0.001

In bold – variants for which the association with RBD is in the opposite direction of the reference association. For example, rs356182-G allele is associated with an increased risk for PD and decreased risk for RBD (see also Figure 1B).

SNP: single nucleotide polymorphism. EA: effect allele. EAF: effect allele frequency. LD: linkage disequilibrium. OR: odds ratio. GWAS: genome-wide association study. eQTL: expression quantitative trait loci.