Low-frequency and rare coding variants of *NUS1* contribute to pathogenesis and phenotype of Parkinson's disease: a case-control study

Li Jiang
Xiangya Hospital Central South University
Jun-pu Mei
Xiangya Hospital Central South University
Yu-wen Zhao
Xiangya Hospital Central South University
Rui Zhang
Xiangya Hospital Central South University
Hong-xu Pan
Xiangya Hospital Central South University
Yang Yang
Xiangya Hospital Central South University
Qi-ying Sun
Xiangya Hospital Central South University
Qian Xu
Xiangya Hospital Central South University
Xin-xiang Yan
Xiangya Hospital Central South University
Jie-qiong Tan
Xiangya Hospital Central South University
Jin-chen Li
Xiangya Hospital Central South University
Bei-sha Tang
Xiangya Hospital Central South University
Jifeng Guo (guojifeng2003@163.com)
Xiangya Hospital Central South University https://orcid.org/0000-0002-3658-3928

Research

Keywords: Parkinson's disease, NUS1, rare variants, low frequency variants, association study

Posted Date: August 4th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-51745/v1

Abstract

Background NUS1 has recently been identified as a candidate risk gene for Parkinson's disease (PD), but the contribution of NUS1 rare and low-frequency variants to PD susceptibility and phenotypes is largely unknown. Methods In our case-control study, whole-exome or Sanger sequencing was performed on the subjects (4,779 cases vs. 4,442 controls) to analyze the coding sequence of NUS1 . The associations between variants and phenotypic data were analyzed using sequence kernel association test and regression models. Results A total of 13 variants were identified. Ten of them in 12 patients and one control were rare variants and three were low-frequency variants. Three rare variants (R86L, N144K, D163H) might be pathogenic. We identified a significant burden of rare NUS1 variants in PD (adjusted P=0.016). Two low-frequency variants, rs550854234 and rs539668656, were associated with PD (odds ratio = 0.76, adjusted P = 0.041; odds ratio = 2.80, adjusted P = 0.016; respectively). Analyses stratified by age at onset showed that the same two variants were associated with late-onset PD (odds ratio = 0.66, adjusted P = 0.025; odds ratio = 2.96, adjusted P = 0.025; respectively). The genotype-phenotype associations of these variants showed that patients with PD carrying rare variants, rs550854234 or rs539668656 were significantly associated with earlier onset age, emotional impairment and tremor severity. Conclusions Our study suggests that rare and low-frequency NUS1 variants play an important role in the pathogenesis and phenotype of PD. Moreover, our data will help understand the role of NUS1 plays in the pathogenesis of PD and further the development of personalized treatments for PD.

Introduction

Over the past 20 years, considerable efforts have been made to identify the genetic factors that cause the Parkinson's disease (PD) phenotype. Research has focused on the identification of PD-associated mutations, resulting in important discoveries in the genetics of PD[1]. This knowledge, in turn, may accelerate the development of mechanism-based therapies. To date, association studies have identified many genetic risk loci and many variants of the PD phenotype[2]. The discovery of PD-related genetic risk factors could help optimize the prevention and management of PD by enabling the identification of high-risk individuals. Recently, the targeted capture and sequencing of the protein-coding regions of the genome, known as exome sequencing, has been commonly used to the research of PD and other complex diseases[3, 4].Our recent whole-exome sequencing (WES) analysis[5] has identified *NUS1* as a candidate risk gene for PD. However, a recent association study with a small sample size[6] failed to observe any significant association between *NUS1* variants and PD. Considerable evidence suggests that a few low-frequency variants play a significantly greater role in establishing biomedical traits than do more common variants, and rare variants with large effect sizes are particularly relevant clinically[7]. However, the detection of associations for individual low-frequency and rare variants lacks statistical power with readily attainable sample sizes. Hence, exceptional sample sizes are often required for detecting these variants[8].In this study, we explored the contribution of low-frequency and rare *NUS1* variants to PD involved three steps. First, we sequenced the *NUS1* coding regions in 1,542 PD patients and 1,625 controls(Cohort A) by WES to explore which aggregate variants are associated with PD. Second, to further increase power, we combined our previous study cohort[5], which included 3,237 PD patients and 2,817 controls(Cohort B), with Cohort A to analyze the association of individual low-frequency variants with PD. Third, to explore the influence of the PD-associated genetic variants on clinical symptoms, linear and logistic regression analyses were utilized to reveal genotype-phenotype correlations. Our study systematically investigated the association of *NUS1* low-frequency and rare coding variations with PD susceptibility/phenotypes in the Han Chinese population.

Materials And Methods

Patients and controls

We recruited consecutive patients with PD admitted to the Department of Neurology of Xiangya Hospital from February 2006 to December 2019 and at other cooperating centers of the Parkinson's Disease & Movement Disorders Multicenter Database and Collaborative Network in China (PD-MDCNC, http://pdmdcnc.com:3111/). Two independent cohorts were included: (A) Cohort A, a part of samples(816 cases vs. 0 controls) overlapping with our previous cohort genotyped by molecular inversion probe technique[5], including 1,542 Han Chinese PD patients from mainland China (mean onset age, 46.03 ± 8.25 y; males, 54.28%), and 1,625 age- (mean age, 44.39 ± 8.14 y), gender- (males, 53.29%) and race-matched healthy controls selected from the Health Examination Center of Xiangya Hospital, the recruited patients' spouses, and the communities of Changsha; and (B) Cohort B, our previous cohort screened with Sanger sequencing, including 3,237 patients with PD and 2,817 healthy controls[5]. Cohort A included patients with a family history of PD (at least 1 relative with PD or parkinsonism) or born to consanguineous parents or with age at onset (AAO) no more than 50 y, and the controls without a history of neurological disease. Cohort B mainly included late-onset sporadic PD patients(90.64%) and healthy controls. All patients were diagnosed with PD by a movement disorder neurologist according to the clinical diagnostic criteria of either the UK PD Society Brain Bank [9] or the Movement Disorders Society[10]. Pathogenic variants of high-confidence PD disease-causing genes were ruled out in the patients of Cohort A[11, 12]. Patients with PD were defined as having early onset PD (EOPD) if AAO was ≤50 y[13] and late onset PD (LOPD) if AAO was >50 y. Clinical information collected on PD patients included disease duration, motor, and non-motor manifestations[14]**.** Blood samples were obtained from all participants. Genomic DNA was prepared from peripheral blood leukocytes according to standard procedures. Our protocol was approved by the Ethics Committee of Xiangya Hospital of Central South University, and written informed consent was collected from all participants according to the Declaration of Helsinki.

Genotyping method, quality control, and analysis of population structure

Cohort A was sequenced using WES technology. The average sequencing depth was 123× and a minimum of 10× coverage was achieved for 99.32% of the targeted regions. The data processing and analysis of sequencing data were carried out as described previously[15, 16] to obtain high-quality variants. ANNOVAR[17] was used to annotate the variants based on the human genome hg19 RefSeq, including gene regions, amino acid alterations, functional effects, and allele frequencies for East Asian and all populations taken from the gnomAD database.

Prior to the association study, we used PLINK v1.90[18] to screen the WES data for individual and variant quality control. In this step, genotypes were filtered to exclude those with a missing rate >5% and deviations from Hardy-Weinberg equilibrium in cases and controls (P < 1.0E-4)**.** In addition, samples with discordant gender, an unusual heterozygosity of >3 standard deviations, and unusual relatedness (identity by descent > 0.15) were also excluded. To assess potential population-structure factors, a principal-components analysis (PCA) was then performed with PLINK v1.90 after linkage disequilibrium (LD) pruning (Plink options: indep-pairwise 50,5,0.2). The top two principal components calculated from independent single-nucleotide polymorphisms were included as covariates in subsequent analyses as fixed effects to control for population structure.

Selection of *NUS1* variants

The *NUS1* variants, which comprised exon-coding sequences and exon-intron boundaries (±2bp) within the transcript region (NM_138459 at chr6:117996834-118028178)[19], were then extracted. The minor allele frequency (MAF) of each variant was then calculated according to Cohort A to define common variants (MAF ≥ 5%), low-frequency variants (0.1% < MAF < 5%), and rare variants (MAF ≤ 0.1%). The variants were then further classified into synonymous, nonsynonymous, and deleterious (i.e., predicted damaging by at least five of 11 prediction algorithms by the procedure of Quadri et al.[20]).

Statistical analysis

For variant burden analysis, sequence kernel association test (SKAT)[21] was performed using the SKAT R package. Optimized SKAT (SKAT-O) was applied to Cohort A to analyze the joint effect of rare variants, low-frequency variants, and variant sets stratified by functional level after adjusting for AAO in cases (age at entry in controls), sex and the first two principal components. To further increase the detection power of single low-frequency variant association analyses, we pooled Cohort A and Cohort B. Power to identify associations between low-frequency variation and PD was estimated using QUANTO 1.2 (http://biostats.usc.edu/software) under a log-additive genetic model. Power to detect variants contributing 1% to the phenotypic variation was >90%, depending on the MAF of the variants (range, 0.1%–5%). Further stratification by AAO was performed to assess the association of low-frequency variants with EOPD and LOPD. Logistic regression analyses were performed using PLINK v1.90, with adjustment for sex and AAO in cases (age at entry in controls). For the identified variants and clinical data association analyses, linear and logistic regression analyses were performed using PLINK v1.90, with adjustment for sex, AAO, and disease duration.

Continuous variables are presented as mean ± standard deviation and categorical variables are presented as frequencies (percentages). The corresponding odds ratios (OR) or beta coefficients (β), 95% confidence intervals (CI), and P-values in association analyses are also provided. For multiple comparisons, traditional Bonferroni correction is considered overly conservative, which may result in the increase of false negative errors and rejection of significant associations[22]. We used the method of Benjamini-Hochberg to calculate adjusted P values for false discovery rate, which reduces the probability of false negative errors while still controlling for false positive errors. Adjusted P values <0.05 were considered statistically significant.

Results

NUS1 variants identified in Cohort A

From the 1,542 PD patients and 1,625 healthy controls in our study, we identified 13 variants including 10 rare and 3 low-frequency variants in 98 patients and 78 controls (No common variants were observed). Of these, eight polymorphisms (rs974335534, rs550854234, rs761121795, rs539668656, rs969919569, rs573584810, rs369403261, and rs28362519) have been reported previously and five single-nucleotide polymorphisms (S24S, F30F, A80A, N144K, and L159I) were novel. Among the 13 variants, 10 rare variants (H19H, S24S, F30F, K58N, A80A, R86L, T106T, N144K, L159I, and D163H) were identified in 12 patients and one healthy control. Considering the functional effect of the variants, seven were synonymous and seven were nonsynonymous as shown in Table 1. Pathogenicity predictions indicated that three (R86L, N144K and D163H) of the nonsynonymous variants may be deleterious (Table 1, Table S1).

Burden analyses

In Cohort A, significant associations comparing patients and controls were detected for the *NUS1* gene (P = 0.023, adjusted P = 0.046). Further stratification by the variants' predicted functional properties showed that the association was mainly in the nonsynonymous variants (P = 0.022, adjusted P = 0.046). Stratification by allele frequency showed that the significant association mainly involved the rare variants (P = 0.0026, adjusted P = 0.016). However, when synonymous variants (P = 0.071), deleterious variants (P = 0.20), or low-frequency variants (P = 0.21) were considered for analysis, no significant associations were observed (Table 2). Our burden results further support the view that *NUS1* is a risk gene for PD.

Association between low-frequency variants and PD

Additionally, we investigated the association between 3 low-frequency variants (rs550854234, rs539668656, and rs28362519) and PD. To further increase the power to detect associations, we collected the results for rs550854234, rs539668656, and rs28362519 from the Cohort B and combined the two cohorts for joint analysis. Among the three low-frequency variants, we identified two independent variants (rs550854234 and rs539668656, R^2<0.2) that were associated with PD (adjusted P < 0.05) (Table 3).The frequency of rs539668656 in PD patients was higher than that of healthy controls

even after correction for multiple comparisons (OR = 2.80, 95% CI = 1.36-5.80, P = 0.0054, adjusted P = 0.016). The rs550854234 variant was associated with a decreased PD risk after correction for multiple comparisons (OR = 0.76, 95% CI = 0.60-0.97, P = 0.027, adjusted P = 0.041). However, the rs28362519 variant was not significantly associated with PD (P = 0.48).

The patients with PD from the combined cohort were divided into EOPD and LOPD cohorts (i.e., stratified by AAO; Table 3). In cases of LOPD, we found that rs539668656 was associated with an increased LOPD risk after correction for multiple comparisons (OR = 2.96, 95% CI = 1.22-7.19, P = 0.017, adjusted P = 0.025), and rs550854234 was associated with a decreased LOPD risk after correction for multiple comparisons (OR = 0.66, 95% CI = 0.47-0.92, P = 0.016, adjusted P = 0.025). However, neither of the three variants were significantly associated with EOPD (P > 0.05).

Variant-clinical data association

We also analyzed the clinical features of *NUS1*-variant carriers and non-carriers (Table 4). We found that in Cohort A, patients with rare variants had onset with PD with 6.16-year earlier than did non-carriers (OR=-6.16, P=0.0087), and had suffered a 6.88 times potential high risk of depression (HAMD[23], OR=6.88, P=0.019) than those without rare variants. In the combined cohort, patients with rs550854234 had decreased risk for depression (HAMD[23], OR=0.51, P=0.035) than the non-carriers, and rs550854234 in patients were positively associated with tremor score (UPDRS Item 20 and 21[24], β=0.86, P=0.027). Patients with rs539668656 had more severe mental/behavioral symptoms and greater emotional impairment (UPDRS-Part I[24], β=1.38, P=0.012) than did non-carriers.

Additionally, we further validated the effects of the *NUS1* variants on PD-related phenotype using a more conservative strategy: Permutation test by 10,000 times of random permutations, which does not require assumptions of distribution, also showed significant effects for the *NUS1* rare variants on the AAO (EMP2=0.019), the *NUS1* rare variants and rs550854234 on the incidence of depression (EMP2=0.0041; EMP2=0.041;respectively), rs550854234 on the tremor level (EMP2=0.022), and rs539668656 on mental/behavioral symptoms and emotional impairment (EMP2= 0.015).

Discussion

NUS1, encoding Nogo-B receptor and localized to the membrane of the endoplasmic reticulum, belongs to the undecaprenyl diphosphate synthase family. The neurodevelopmental phenotype, which includes epilepsy and tremors, is reportedly associated with a 6q22 deletion, which includes *NUS1*[25]. To date, six variants have been identified to be associated with developmental delays and epileptic encephalopathy, a congenital disorder of glycosylation, and early-onset PD[26].

In our previous research, we identified *NUS1* as a candidate risk gene for PD through genetic and functional studies. Further studies are needed to identify any variants within the protein-coding regions of *NUS1* and to explore whether these variations modify PD's phenotype. However, few studies have focused on the association between *NUS1* variants and PD. In the present research, we used WES to sequence the exons and exon-intron junctions of *NUS1* in 1,542 PD patients and 1,625 healthy controls from the Chinese Han population and totally identified 13 variants. Among the 13 variants, ten rare variants (H19H, S24S, F30F, K58N, A80A, R86L, T106T, N144K, L159I, and D163H) were identified in PD. Three of these rare variants (R86L, N144K, and D163H) could be deleterious according to *in silico* predictions. The three variants were extremely rare or absent in the 1,625 neurologically normal Han Chinese controls and in East Asian and all populations described by the gnomAD database. PhastCons analysis showed that these variants were located in the highly conserved region of the *NUS1* gene, suggesting that the three variations were pathogenic according to the criteria of the ACMG[24] (Table 1). Our results showed that rare *NUS1* mutations made a significant contribution to PD risk, especially in a large sample of familial PD(FPD) and EOPD population more likely to be driven by genetic factors. Enrichment of nonsynonymous variants in PD was also observed, which suggest that genetic variations at the *NUS1* locus is a significant risk factor for PD in our large sample of Han Chinese FPD and EOPD patients.

Three low-frequency *NUS1* variants identified here (rs550854234, rs539668656, and rs28362519) have been reported previously in a genetic analysis of *NUS1* in a Chinese PD population[6], and none were found to have a significant association with sporadic PD. In the present study, an association was found between *NUS1* rs550854234 and rs539668656 and PD; rs539668656 was associated with an increased risk of PD, while rs550854234 was associated with a decreased risk of PD, even after correction for multiple comparisons. The main reason for the discrepancy between our results and those of Xiang et al. may lie in the differences between the samples sizes and the genetic heterogeneity of the populations studied. Furthermore, we found that rs550854234 and rs539668656 were also associated with LOPD risk, suggesting that the role of rs550854234 and rs539668656 in the pathogenesis of LOPD is of importance. Notably, we found that rs550854234 and rs539668656 had entirely opposite effects on PD risk. Interestingly, *NUS1* is considered intolerant of loss-of-function variants, based on a pLI score of 0.87 (pLI, probability of loss-of-function-intolerant) found in the Exome Aggregation Consortium Browser database[27] (http://exac.broadinstitute.org), suggesting that *NUS1* variants may lead to disease through haploinsufficiency. Sequence variants within classical splice sites or splicing enhancer sequences leading to splicing defects can be reportedly an important mechanism of pathogenicity, especially for those genes in which loss of function is the common pathogenic mechanism[28]. A potential effect of rs550854234 and rs539668656 on splicing was found by searching the Human Splicing Finder (HSF)[29] (http://www.umd.be/HSF/), showing that rs539668656 can affect splicing through the alteration of an exonic splicing enhancer site (Table S2), which may have an effect on regulation of *NUS1* expression. Although rs550854234 seems to have no effect on the amino acid sequences of proteins or on splicing signals, it is commonly accepted that synonymous mutations can cause either enhancement or suppression of local translation rates, conformation, or substrate specificity, depending on the location of the mutation, thereby affecting the function of the protein[30]. These data suggest that rs550854234 and rs539668656 may contribute to PD pathogenesis through different mechanisms.

In this study, genotype-phenotype analysis showed that patients with rare variants showed a significant association with earlier AAO in FPD and EOPD patients. Our results revealed that patients with rare *NUS1* variants were associated with a greater risk of depression, while rs550854234 carriers showed a lower risk of depression and rs539668656 was also associated with more severe mental/behavioral symptoms and emotional impairment. Previous studies have shown the association between depression[31], affective disorder and PD risk[32]. The study of Fang et al[33] suggested that depression might be an early symptom in PD and shared common etiological basis with PD. We speculate that the *NUS1* rare and low-frequency variants affecting the onset of depression or emotional impairment may contribute to PD risk through an unknown mechanism. Further researches to know the pathogenesis role of *NUS1* variants underlying PD susceptibility/phenotypes are needed.

This study has several limitations. First, although we have used a stringent filtering strategy to identify deleterious variants, those identified in our study could not be analyzed for co-segregation to prove their true pathogenesis owing to the relatives of these patients refusing genetic testing or out of contact. Moreover, N144K has been identified in PD patients and absent in controls in the study of Xu et al[6]. Second, we lack an independent replication cohort to verify the associations between the two low-frequency variants (rs550854234 and rs539668656) and PD. We attempted to analyze the largest available genome-wide association studies dataset in PD[2] to identify the relationship, but these two variants were not found in the patients and controls. Third, no functional analysis was performed to verify an effect of rs550854234 and rs539668656 on alternative splicing and expression of *NUS1* due to a lack of RNA samples. Therefore, more replications and functional studies are needed to provide more evidence of these rare variants' pathogenicity and the two low-frequency variants' association with PD, and further clarify the biological functions of rs550854234 and rs539668656 in PD.

Conclusions

In conclusion, our identification of rare and low-frequency *NUS1* variants associated with PD suggests that additional studies should focus on this important gene in the Chinese and other populations, and our findings may facilitate a better

understanding of pathophysiological mechanisms of *NUS1* variants in PD and the development of personalized treatments.

Additional File

Additional file 1:Table S1. Functional Prediction of 7 nonsynonymous variants in *NUS1*

Additional file 2:Table S2. Splicing alterations prediction of 13 variants identified in *NUS1*

Abbreviations

PD: Parkinson's disease; WES: whole-exome sequencing; PD-MDCNC: Parkinson's Disease & Movement Disorders Multicenter Database and Collaborative Network in China; EOPD: early onset PD; LOPD: late onset PD; PCA: principal-components analysis; LD: linkage disequilibrium; MAF: minor allele frequency; SKAT: sequence kernel association test; SKAT-O: Optimized SKAT; OR: odds ratios; CI: confidence intervals; FPD: familial PD; pLI: probability of loss-of-function-intolerant; HSF: Human Splicing Finder.

Declarations

Acknowledgements

We thank all patients and healthy-control participants for permitting us to publish their information; PD-MDCDC for providing clinical materials; and all members of our research group for assistance in whole-exome sequencing.

Funding

This study was funded by the National Key Research and Development Program of China (Grant No.2017YFC0909101 and 2016YFC1306000), the Central Public-Interest Scientific Institution Basal Research Fund of the Chinese Academy of Medical Sciences (Grant No.2018-12M-HL-025), the National Natural Science Foundation of China (Grant No.81430023, 81974202, 81873785), Innovation-Driven Team Project from Central South University(Grant No. 2020CX016) and innovation capacity building project of National Clinical Research Center for Geriatric Disorders(Grant No.2019SK2335).

Availability of data and materials

All data generated or analysed during this study are included in this article or uploaded as additional files.

Authors' contributions

Li Jiang and Jun-pu Mei contributed equally to the work as coauthors. Li Jiang, Jun-pu Mei, Yu-wen Zhao, Rui Zhang, Hong-xu Pan participated in patients and controls recruitment; Li Jiang and Ji-feng Guo designed the study; Yang Yang, Qi-ying Sun, Qian Xu, Xin-xiang Yan, Jie-qiong Tan, Bei-sha Tang and Ji-feng Guo participated in the sample management; Li Jiang and Jin-chen Li analyzed the bioinformatics data and interpreted the data; Li Jiang participated in manuscript writing; Bei-sha Tang and Ji-feng Guo had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis; All authors revised the manuscript for important intellectual content; Bei-sha Tang and Ji-feng Guo obtained funding or material support.

Ethics approval and consent to participate

Our protocol was approved by the Ethics Committee of Xiangya Hospital of Central South University, and written informed consent was collected from all participants according to the Declaration of Helsinki.

Consent for publication

Not applicable.

Competing interests

The authors declare no conflicts of interest.

References

1. Singleton A, Hardy J. The Evolution of Genetics: Alzheimer's and Parkinson's Diseases. Neuron. 2016;90(6):1154-63.
2. Nalls MA, Blauwendraat C, Vallerga CL, Heilbron K, Bandres-Ciga S, Chang D, et al. Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. Lancet Neurol. 2019;18(12):1091-102.
3. Xu Q, Li K, Sun Q, Ding D, Zhao Y, Yang N, et al. Rare GCH1 heterozygous variants contributing to Parkinson's disease. Brain. 2017;140(7):e41.
4. Groopman EE, Marasa M, Cameron-Christie S, Petrovski S, Aggarwal VS, Milo-Rasouly H, et al. Diagnostic Utility of Exome Sequencing for Kidney Disease. N Engl J Med. 2019;380(2):142-51.
5. Guo JF, Zhang L, Li K, Mei JP, Xue J, Chen J, et al. Coding mutations in NUS1 contribute to Parkinson's disease. Proc Natl Acad Sci U S A. 2018;115(45):11567-72.
6. Chen X, Xiao Y, Zhou M, Lin Y, Guo W, Huang S, et al. Genetic analysis of NUS1 in Chinese patients with Parkinson's disease. Neurobiol Aging. 2020;86(Z1):202.e5-202.e6.
7. Robinson MR, Wray NR, Visscher PM. Explaining additional genetic variation in complex traits. Trends Genet. 2014;30(4):124-32.
8. Rivas MA, Pirinen M, Neville MJ, Gaulton KJ, Moutsianas L, GoT2D Consortium, et al. Assessing association between protein truncating variants and quantitative traits. Bioinformatics. 2013;29(19):2419-26.
9. Samii A, Nutt JG, Ransom BR. Parkinson's disease. Lancet. 2004;363(9423):1783-93.
10. Postuma RB, Berg D, Stern M, Poewe W, Olanow CW, Oertel W, et al. MDS clinical diagnostic criteria for Parkinson's disease. Mov Disord. 2015;30(12):1591-601.
11. Zhao Y QL, Pan H LZ, Jiang L HY, Zeng Q ZX, Zhou X ZY, Fang Z WZ, et al. The role of genetics in Parkinson's Disease a large cohort study in Chinese mainland population. Brain. 2020.
12. Blauwendraat C, Nalls MA, Singleton AB. The genetic architecture of Parkinson's disease. Lancet Neurol. 2020;19(2):170-8.
13. Lv Z, Guo J, Sun Q, Li K, Yu R, Tian J, et al. Association between PLA2G6 gene polymorphisms and Parkinson's disease in the Chinese Han population. Parkinsonism Relat Disord. 2012;18(5):641-4.
14. Xiang YQ, Xu Q, Sun QY, Wang ZQ, Tian Y, Fang LJ, et al. Clinical Features and Correlates of Excessive Daytime Sleepiness in Parkinson's Disease. Front Neurol. 2019;10:121.
15. Tian Y, Wang JL, Huang W, Zeng S, Jiao B, Liu Z, et al. Expansion of Human-Specific GGC Repeat in Neuronal Intranuclear Inclusion Disease-Related Disorders. Am J Hum Genet. 2019;105(1):166-76.
16. Zeng S, Zhang MY, Wang XJ, Hu ZM, Li JC, Li N, et al. Long-read sequencing identified intronic repeat expansions in SAMD12 from Chinese pedigrees affected with familial cortical myoclonic tremor with epilepsy. J Med Genet. 2019;56(4):265-70.
17. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38(16):e164.
18. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81(3):559-75.
19. Zweig AS, Karolchik D, Kuhn RM, Haussler D, Kent WJ. UCSC genome browser tutorial. Genomics. 2008;92(2):75-84.

20. Quadri M, Mandemakers W, Grochowska MM, Masius R, Geut H, Fabrizio E, et al. LRP10 genetic variants in familial Parkinson's disease and dementia with Lewy bodies: a genome-wide linkage and sequencing study. Lancet Neurol. 2018;17(7):597-608.
21. Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X. Rare-variant association testing for sequencing data with the sequence kernel association test. Am J Hum Genet. 2011;89(1):82-93.
22. Nyholt DR. A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. Am J Hum Genet. 2004;74(4):765-9.
23. HAMILTON M. A rating scale for depression. J Neurol Neurosurg Psychiatry. 1960;23:56-62.
24. Starkstein SE, Merello M. The Unified Parkinson's Disease Rating Scale: validation study of the mentation, behavior, and mood section. Mov Disord. 2007;22(15):2156-61.
25. Szafranski P, Von Allmen GK, Graham BH, Wilfong AA, Kang SH, Ferreira JA, et al. 6q22.1 microdeletion and susceptibility to pediatric epilepsy. Eur J Hum Genet. 2015;23(2):173-9.
26. Den K, Kudo Y, Kato M, Watanabe K, Doi H, Tanaka F, et al. Recurrent NUS1 canonical splice donor site mutation in two unrelated individuals with epilepsy, myoclonus, ataxia and scoliosis - a case report. BMC Neurol. 2019;19(1):253.
27. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature. 2016;536(7616):285-91.
28. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 2015;17(5):405-24.
29. Desmet FO, Hamroun D, Lalande M, Collod-Béroud G, Claustres M, Béroud C. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. Nucleic Acids Res. 2009;37(9):e67.
30. Sauna ZE, Kimchi-Sarfaty C. Understanding the contribution of synonymous mutations to human disease. Nat Rev Genet. 2011;12(10):683-91.
31. Alonso A, Rodríguez LA, Logroscino G, Hernán MA. Use of antidepressants and the risk of Parkinson's disease: a prospective study. J Neurol Neurosurg Psychiatry. 2009;80(6):671-4.
32. Nilsson FM, Kessing LV, Bolwig TG. Increased risk of developing Parkinson's disease for patients with major affective disorder: a register study. Acta Psychiatr Scand. 2001;104(5):380-6.
33. Fang F, Xu Q, Park Y, Huang X, Hollenbeck A, Blair A, et al. Depression and the subsequent risk of Parkinson's disease in the NIH-AARP Diet and Health Study. Mov Disord. 2010;25(9):1157-62.

Tables

Table 1. Analysis of *NUS1* coding variants identified in Cohort A

rsID	Position (hg19)	Exon	Nucleotide change	AA change	ExonicFunc.refGene	PD (1,542)	HC (1,625)	MAF[a]	phastCons	ACMG
rs974335534	117996890[d]	1	c.57C>T	p.H19H	synonymous	1	0	-/0/-/0.000007871/0.0001604	-	US
rs550854234	117996896[c]	1	c.63G>T	p.T21T	synonymous	45	49	0.0125/0.0142/0.0007/0.0012/0.01509	-	B
Novel	117996905[d]	1	c.72C>T	p.S24S	synonymous	1	0	-/-/-/-/0.0001598	-	US
Novel	117996923[d]	1	c.90C>T	p.F30F	synonymous	1	0	-/0.0006/-/0.00004782/0.0001591	-	US
rs761121795	117997007[d]	1	c.174G>T	p.K58N	nonsynonymous	1	0	0/0/0.0002/0.0003/0.0001591	C(0.999)	US
rs539668656	117997030[c]	1	c.197G>C	p.R66P	nonsynonymous	5	2	0.0019/0.0025/0.0012/0.0076 /0.001116	C(1.000)	US
Novel	117997073[d]	1	c.240C>A	p.A80A	synonymous	1	0	-/0.0002/-/0.00001459/0.0001623	-	US
rs969919569	117997090[bd]	1	c.257G>T	p.R86L	nonsynonymous	2	0	0.0006/0/0.0000325/0/0.0003301	C(1.000)	P
rs573584810	117997151[d]	1	c.318C>T	p.T106T	synonymous	1	0	0/0/0.0002/0.0004/0.0001591	-	US
Novel	118014221[bd]	2	c.432T>G	p.N144K	nonsynonymous	1	0	-/-/-/-/0.0001579	C(1.000)	P
Novel	118014264[d]	2	c.475C>A	p.L159I	nonsynonymous	1	0	-/-/-/-/0.000158	C(1.000)	US
rs369403261	118014276[bd]	2	c.487G>C	p.D163H	nonsynonymous	2	1	-/0.0001/-/0.00000814/0.0004738	C(1.000)	P
rs28362519	118014326[c]	2	c.537T>A	p.D179E	nonsynonymous	36	26	0.008/0.0098/0.0005/0.0007/0.009792	C(1.000)	B

a. MAF from gnomAD_genome_EAS, gnomAD_exome_EAS, gnomAD_genome_ALL, gnomAD_exome_ALL and our Cohort A

b. Deleterious variants as defined by at least 5 of the 11 protein prediction algorithms

c. The low-frequency variants identified in our cohort

d. The rare variants identified in our cohort

HC, healthy control; AA, amino acid; C, Conserved; US, Uncertain significance; B, Benign; P, Pathogenic; ACMG, American College of Medical Genetics and Genomics

Table 2. Burden analysis of *NUS1* coding variants

Variants group	Markers included	Number of samples carrying variants		P value	Adjusted P
		Case(1,542)	Control(1,625)		
All	13	98	78	0.023	0.046
Synonymous	6	50	49	0.071	0.11
Nonsynonymous	7	48	29	0.022	0.046
Deleterious	3	5	1	0.20	0.21
Rare	10	12	1	0.0026	0.016
Low-frequency	3	86	77	0.21	0.21

Table 3. Association analysis of low-frequency coding variants with PD,EOPD and LOPD

rsID	The combined cohort					EOPD cohort					LOPD cohort				
	Case (Hom/Het/Wild)	Control (Hom/Het/Wild)	OR(95% CI)	P	Adjusted P	Case (Hom/Het/Wild)	Control (Hom/Het/Wild)	OR(95% CI)	P	Adjusted P	Case (Hom/Het/Wild)	Control (Hom/Het/Wild)	OR(95% CI)	P	Adjusted P
rs5508 54234	3/118/4614	0/154/4279	0.76(0.60,0.97)	0.027	0.041	0/60/1940	0/78/2294	0.89(0.63,1.26)	0.52	0.77	3/58/2674	0/76/1985	0.66(0.47,0.92)	0.016	0.025
rs5396 68656	1/29/4719	0/10/4432	2.80(1.36,5.80)	0.005	0.016	0/7/2007	0/4/2374	3.16(0.88,11.43)	0.079	0.24	1/22/2712	0/6/2058	2.96(1.22,7.19)	0.017	0.025
rs2836 2519	1/97/4680	0/83/4359	1.12(0.83,1.50)	0.48	0.48	0/40/1994	0/49/2329	0.97(0.63,1.48)	0.89	0.89	1/57/2686	0/34/2030	1.26(0.82,1.93)	0.30	0.30

Table 4. Comparison of demographic and clinical features in carriers and non-carriers of *NUS1* variants

Clinical features	Rare variants (Cohort A)			rs550854234 (The combined cohort)			rs539668656 (The combined cohort)		
	N	β/OR	P	N	β/OR	P	N	β/OR	P
Age at onset, years	1347	-6.16	0.0087	4735	-0.67	0.45 [a]	4749	-0.73	0.68 [a]
Disease duration, years	1347	0.53	0.70	4735	0.33	0.48 [b]	4749	-0.61	0.50 [b]
Motor manifestations									
UPDRS-Part I	1347	0.052	0.93	3138	-0.33	0.14	3152	1.38	0.012
UPDRS-Part II	1347	1.60	0.36	3137	0.36	0.60	3151	2.37	0.16
UPDRS-Part III	1347	5.74	0.18	3140	2.14	0.18	3154	4.67	0.23
Tremor score	1347	0.68	0.53	3140	0.86	0.027	3154	1.25	0.19
Stiffness score	1347	1.18	0.32	3140	0.65	0.14	3154	0.26	0.81
Bradykinesia score	1347	1.42	0.44	3140	0.55	0.42	3154	2.09	0.21
Postural instability score	1314	1.42	0.084	3140	-0.12	0.71	3154	0.26	0.74
H-Y stage	1347	0.40	0.16	3140	0.70	0.21	3154	0.46	0.16
Dyskinesia	1310	2.96	0.31	3067	0.87	0.65	3079	1.92	0.34
Freezing gait	1314	0.75	0.71	3058	1.23	0.40	3070	0.90	0.87
Non-motor manifestations									
SCOPA-AUT score	719	2.25	0.38	1634	-0.80	0.42	1650	2.28	0.45
MMSE score	1007	-0.43	0.69	2606	-0.027	0.95	2621	-0.26	0.82
HAMD	948	6.88	0.019	2519	0.51	0.035	2534	1.02	0.98
PDSS score	934	-12.60	0.22	2512	4.61	0.18	2526	-7.82	0.40
RBDQ-HK	928	1.92	0.39	2499	1.22	0.45	2514	2.74	0.14
ESS	923	1.07	0.93	2498	1.18	0.53	2512	3.15	0.09
HRS	922	0.57	0.50	2502	1.10	0.70	2517	1.79	0.39
PDQ-39 score	911	15.70	0.056	1553	-1.50	0.68	1568	3.74	0.71

a. Adjusting for sex and disease duration at entry.

b. Adjusting for sex and AAO.

UPDRS, Unified Parkinson's Disease Rating Scale; H-Y, Hoehn and Yahr; SCOPA-AUT, Scale for Outcomes in PD for Autonomic Symptoms; MMSE, Mini-Mental State Examination; HAMD, Hamilton Depression Scale; PDSS, Parkinson’s disease Sleep Scale; RBDQ-HK, REM sleep behavior disorder questionnaire–Hong Kong; ESS, Epworth Sleepiness Scale; HRS, Hyposmia rating scale; PDQ-39, Parkinson’s Disease Questionnaire-39

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryOnlineMaterials.docx