



OPEN

Investigating the relationship between the *SNCA* gene and cognitive abilities in idiopathic Parkinson's disease using machine learning

Mehrafarin Ramezani^{1,2}, Pauline Mouches^{2,3,4}, Eunjin Yoon^{1,2}, Deepthi Rajashekar^{2,3,4}, Jennifer A. Ruskey^{5,6}, Etienne Leveille⁵, Kristina Martens², Mekale Kibreab^{1,2}, Tracy Hammer^{1,2}, Iris Kathol^{1,2}, Nadia Maarouf^{1,2}, Justyna Sarna^{1,2}, Davide Martino^{1,2}, Gerald Pfeffer^{1,2,7}, Ziv Gan-Or^{5,6,8}, Nils D. Forkert^{1,2,4} & Oury Monchi^{1,2,4,6}✉

Cognitive impairments are prevalent in Parkinson's disease (PD), but the underlying mechanisms of their development are unknown. In this study, we aimed to predict global cognition (GC) in PD with machine learning (ML) using structural neuroimaging, genetics and clinical and demographic characteristics. As a post-hoc analysis, we aimed to explore the connection between novel selected features and GC more precisely and to investigate whether this relationship is specific to GC or is driven by specific cognitive domains. 101 idiopathic PD patients had a cognitive assessment, structural MRI and blood draw. ML was performed on 102 input features including demographics, cortical thickness and subcortical measures, and several genetic variants (*APOE*, *MAPT*, *SNCA*, etc.). Using the combination of RRELIEFF and Support Vector Regression, 11 features were found to be predictive of GC including sex, rs894280, Edinburgh Handedness Inventory, UPDRS-III, education, five cortical thickness measures (R-parahippocampal, L-entorhinal, R-rostral anterior cingulate, L-middle temporal, and R-transverse temporal), and R-caudate volume. The rs894280 of *SNCA* gene was selected as the most novel finding of ML. Post-hoc analysis revealed a robust association between rs894280 and GC, attention, and visuospatial abilities. This variant indicates a potential role for the *SNCA* gene in cognitive impairments of idiopathic PD.

Parkinson's disease (PD) is traditionally associated with motor symptoms. Non-motor deficits have received less attention but are also quite common and can even precede motor symptoms. PD patients can display a wide range of cognitive deficits of various intensity from Mild Cognitive Impairment (MCI: neuropsychological deficits that do not interfere with daily life activities) to dementia¹. PD patients with cognitive deficits are more likely to develop Parkinson's disease dementia (PDD) as the disease progresses, compared to patients without such deficits². To address the high prevalence and the inevitable negative impact of cognitive deficits, the Movement Disorders Society (MDS) prepared a task force to assist identification of MCI in PD patients^{3,4}. Nevertheless, the underlying mechanisms of PD-MCI development are still obscure and further investigation is required to unravel its nature. At the cellular level, PD is characterized by presence of Lewy Body (LB) deposits in the substantia nigra⁵. LB mainly consists of alpha-synuclein⁶, a protein that is encoded by the *SNCA* gene⁷. *SNCA* is a prominent potential genetic marker in PD due to its involvement in familial PD through point mutations or

¹Department of Clinical Neurosciences, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada. ²Hotchkiss Brain Institute (HBI), Cummings School of Medicine, University of Calgary, Calgary, AB, Canada. ³Biomedical Engineering Graduate Program, University of Calgary, Calgary, Canada. ⁴Department of Radiology, University of Calgary, Calgary, AB, Canada. ⁵Montreal Neurological Institute, McGill University, Montreal, QC, Canada. ⁶Department of Neurology and Neurosurgery, McGill University, Montreal, QC, Canada. ⁷Department of Medical Genetics, Alberta Children's Hospital Research Institute, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada. ⁸Department of Human Genetics, McGill University, Montreal, QC, Canada. ✉email: oury.monchi@ucalgary.ca

Demographic/clinical characteristics	Mean, SD (Min, Max)
Age (years)	70, \pm 7.3 (53, 84)
Education (years)	14.4, \pm 2.9 (8, 21)
Disease duration (years)	5.7, \pm 4.0 (1, 18)
LEDD (mg/d)	869.0, \pm 457.6 (200, 2650)
UPDRS-III	19.46, \pm 9.41 (4, 53)
EHI	68.9, \pm 41.6 (-100, 100)
Sex (female %)	32%
Ethnicity	
European descent	84%
Others	10%
NA	6%
Global cognition Z-score	-0.27, \pm 0.6 (-2, 1)
Attention Z-score	-0.24, \pm 0.6 (-1, 1)
Executive function	-0.43, \pm 0.8 (-3, 1)
Language	-0.13, \pm 0.8 (-2, 1)
Memory	-0.15, \pm 0.8 (-2, 1)
Visuo-spatial	-0.43, \pm 0.9 (-3, 1)

Table 1. The demographic and clinical characteristics. Abbreviations; LEDD = Levodopa Equivalency Daily dosage, UPDRS-III = Unified Parkinson's disease Rating Scale part III, EHI = Edinburgh Handedness Inventory, NA = Not available.

gene dosage effect^{8–10}. *SNCA* is also implicated in a class of disorders called synucleinopathies, which have LB pathology in common, e.g. PD (with or without dementia), dementia with LB (DLB), Multiple System Atrophy (MSA), idiopathic REM Behavioral Disorder (RBD). PDD and DLB share numerous similarities contributing to the challenge of distinguishing them from each other^{11,12}. The substantial involvement of *SNCA* in PDD and DLB pathology through LB and the similarity of the symptomatology, such as the presence of dementia in these two diseases, indicate a potential role for the *SNCA* gene in cognitive decline of idiopathic PD patients and necessitates further investigation. In recent years, several variants of *SNCA* have been discovered in connection with cognitive impairments or dementia in PD^{12,13}. Exploring possible genetic variants associated with cognitive impairments in PD could improve the understanding of the primary biological mechanisms of PD-MCI and identification of patients at risk of cognitive decline. However, given the complexity of PD, further investigation is required to unravel the true involvement of the known and novel genetic variants linked to cognitive deficits in PD using advanced techniques¹⁴.

In addition to genetic risk factors, structural neuroimaging has been used in PD participants to identify image-based biomarkers of cognitive decline^{15–19}. A growing body of evidence supports the use of structural neuroimaging as a biomarker for PD-MCI identification. One example is thinning of frontal and temporal cortices, which has been associated with cognitive decline in PD^{15,19,20}. Also, corticometric and volumetric analyses have shown a reduced volume in frontal and limbic regions in PD-MCI compared to PD-nonMCI patients^{16,21,22}.

In this study, we aimed to predict global cognition in PD participants using a machine learning approach including genetic, structural neuroimaging, clinical, and demographic data as input. In a second step, using post-hoc analyses, we aimed to identify the relationship between the most novel genetic features and the cognitive profile of PD patients in more depth. Furthermore, we aimed to investigate whether this relationship is specific to global cognition or driven by specific cognitive domains including executive function, attention, visuospatial abilities, memory, and language.

Results

Participants. The demographic and clinical characteristics of 101 PD participants included in this study are summarized in Table 1. The mean age of participants was 70 years with a mean disease duration of 6 years. Thirty-two percents of the participants were female (n = 32). Ninety participants were right-handed (89%), six participants were left-handed (6%), and the remaining five participants were identified as ambidextrous (5%). The majority of the participants were of European descent (84%).

Machine learning analysis. The global cognition scores were predicted using the 102 features employing the machine learning framework including feature ranking and a support vector regression model. The best model performance predicting global cognition was achieved when including the 11 top-ranked features (Table 2). This resulted in a correlation coefficient of 0.54 and mean absolute error of 0.39. The selected features were (in order of descending importance): sex, rs894280, EHI, UPDRS-III, years of education, five measures of cortical thickness (right parahippocampal cortex, left entorhinal cortex, right rostral anterior cingulate cortex, left middle temporal cortex, and right transverse temporal cortex), and right caudate volume.

Cognitive domain	Correlation coefficient, R ²	Features
Global cognition	0.54, 0.29	Sex rs894280 EHI score UPDRS-III years of Education R parahippocampal-thickness L entorhinal-thickness R rostralanteriorcingulate-thickness L middletemporal-thickness R transeversetemporal-thickness R caudate-volume

Table 2. Machine learning analysis results. Features are presented in the right column based on their importance in the machine learning model. The features are presented in a descending order (the most important to the least important). Abbreviations: UPDRS-III = Unified Parkinson's disease Rating Scale part III, EHI = Edinburg Handedness Inventory.

Demographic/clinical characteristics Mean, SD, (Min, Max)	CC n = 33	CT/TT n = 68	Significance
Age (years)	71.5, ± 7.3	69.2, ± 7.3	0.14 ^a
Education (years)	14.6, ± 3.2	14.3, ± 2.3	0.66 ^a
Disease duration (years)	5.6, ± 4.2	5.7, ± 4.0	0.56 ^b
LEDD (mg/d)	767.7, ± 300.9	918.2, ± 511.7	0.31 ^b
UPDRS-III	18.14, ± 7.82	20.10, ± 10.09	0.44 ^b
EHI	69, ± 40	69, ± 40	0.83 ^b
Sex (female%)	33%	31%	0.82 ^c
Ethnicity			
European descent	88%	82%	0.67 ^d
Others	9%	10%	
NA	3%	7%	

Table 3. Demographic and clinical characteristics of the two groups of rs894280 variant. Abbreviations; SD = Standard Deviation, LEDD = Levodopa Equivalency Daily dosage, UPDRS-III = Unified Parkinson's disease Rating Scale part III, EHI = Edinburg Handedness Inventory, NA = Not available. ^aStudent-t test. ^bMann-Whitney U test. ^cFisher exact test. ^dChi-square test.

The high-ranked demographic features were further analyzed individually using Pearson correlation or chi-square test. Correlation analysis of sex and EHI score with global cognition were found to be non-significant ($\chi^2 = (1, n = 101) 75.59$ $p = 0.46$ and $R^2 = 0.005$, $p = 0.50$, respectively). Also, UPDRS-III and years of education exhibited significant correlation with global cognition. The UPDRS-III score had a negative correlation while years of education had a positive correlation with global cognition ($R^2 = -0.26$, $p < 0.001$ and $R^2 = 0.07$, $p = 0.007$, respectively). The structural measures were tested for correlation with global cognition and two measures showed significant correlation with global cognition: R rostral-anterior cingulate-thickness, and L middle temporal-thickness ($R^2 = 0.08$, $p = 0.004$, $R^2 = 0.04$, $p = 0.03$, respectively).

Association of the SNCA variant rs894280 and global cognition. Based on the machine learning results, post-hoc analyses were performed to study the association of the alleles of the novel variant rs894280 with global cognition and specific cognitive domains.

Out of 101 participants, 33 had CC genotype, 48 had CT genotype, and 20 had TT genotype. Based on the preliminary analysis of this SNP, participants with the T allele were pooled in one group, resulting in dividing the participants in two allelic groups²³. The demographic and clinical characteristics were not significantly different between the two allelic groups (Table 3).

This variant showed a significant association with global cognition when controlling for UPDRS-III, education, R rostral-anterior cingulate and L middle temporal thickness measures ($F(5,95) = 12.17$, $p < 0.001$, $R^2 = 0.35$). Based on the ANCOVA results, global cognition was significantly different between the CC and CT/TT groups, $F(5,95) = 4.20$, $p = 0.04$.

According to ANCOVA analysis, each participant's Z-score of global cognition increased by 0.25 (95% CI = 0.01–0.42) when the participant had CC genotype (i.e. reference sequence) for rs894280. The calculated Hedges' effect size for rs894280 is 0.4, which represents a medium effect size based on 95% CI.

Association analysis of rs894280 and each cognitive domain was performed using ANCOVA including any demographic/clinical factor with a significant correlation with the domain of interest as covariates. UPDRS-III, education, R rostral-anterior cingulate and L middle temporal thickness measures were included for all domains, except for the visuo-spatial domain for which the EHI score was also added to the model ($R^2 = 0.05$, $p = 0.02$).

Cognitive domain Mean, SD, (Min–Max)	CC N = 33	CT/TT N = 68	CI95% _{rs894280}	Significance
Global cognition	0, ± 0.5, (–1, 1)	–0.4, ± 0.6, (–2, 1)	0.006–0.42 ^a	0.04*
Attention	0, ± 0.5, (–1, 1)	–0.4, ± 0.6, (–1, 1)	0.12–0.56 ^a	0.003**
Executive function	–0.3, ± 0.7, (–2, 1)	–0.5 ± 0.9, (–3, 1)	–0.10–0.55 ^a	0.19
Language	0, ± 0.7, (–1, 1)	–0.2, ± 0.8, (–2, 1)	–0.16 – 0.44 ^a	0.36
Memory	0.2, ± 0.7, (–1, 1)	–0.3, ± 0.8, (–2, 1)	0.07–0.70 ^a	0.02
Visuo-spatial	–0.1, ± 0.6, (–2, 1)	–0.6, ± 1, (–3, 1)	0.16–0.87 ^b	0.005**

Table 4. Results of ANCOVA analysis of rs894280, and global cognition and the five cognitive domains. SD = Standard Deviation, CI 95% = 95% confidence intervals. ^aUPDRS-III, Education, R rostral-anterior cingulate and L middle temporal thickness measures were entered in the model. ^bUPDRS-III, Education, R rostral-anterior cingulate and L middle temporal thickness measures and EHI score were entered in the model. *Significance level was set to 0.05. **Significance level was set to 0.01, Bonferroni correction.

Significance level was set to 0.01 using Bonferroni correction to correct for multiple tests. A significant association was found between rs894280 and the attention domain score ($B = -0.34$, $p = 0.003$) with the CC group displaying better attention abilities. Moreover, this variant showed an association with the visuo-spatial domain score ($B = -0.51$, $p = 0.005$). Similar to results from global cognition and attention, PD patients with CC genotype had better visuo-spatial abilities compared to the other group (CT/TT). The results of association analysis are summarized in Table 4. A trend was observed for the memory domain ($p = 0.02$) while executive function and language did not show any association with this variant. PD participants homozygous for C allele of rs894280 displayed superior attention and visuo-spatial abilities compared to participants who had one or more T alleles.

Discussion

In this study, we used machine learning to predict global cognition in PD patients and post-hoc analysis to investigate the *SNCA* rs894280 variant as a feature associated with cognitive deficits in PD. Using the RRELIEFF feature selection algorithm and SVR, eleven features were selected as the best predictor of global cognition Z-score in this cohort; sex, rs894280, EHI, UPDRS-III, education, five measures of cortical thickness (right parahippocampal cortex, left entorhinal cortex, right rostral anterior cingulate cortex, left middle temporal cortex, and right transverse temporal cortex), and right caudate volume. The selection of features indicate that these variables are informative for prediction of the global cognition score but the direction for each single feature cannot be easily determined based on the machine learning model. Consistent with the machine learning results, which revealed rs894280 as the only genetic factor informative for PD cognition prediction, further analysis was performed on the association of this variant and global cognition. The results suggest that this variant is associated with differences in global cognition, as well as attention and visuo-spatial domains in our cohort, with a medium effect size (Hedges' $g = 0.4$).

The RRELIEFF approach was used to remove redundant and non-informative features and to select the optimal subset of features. This feature selection method has been used for optimal selection of genetic features in previous studies^{24–27}. The SVR was implemented to build the model based on the selected features in order to evade the collinearity issue of the features. The SVR model has been used previously to model PD diagnosis and progression^{28,29} but the current combination used in this study has not been applied before specifically in investigation of cognitive deficits in PD with similar set of inputs.

Except for rs894280, all other features used in the optimal regression model have been reported in different studies to be associated with cognitive decline in PD patients. There is a substantial body of evidence on the role of sex in cognitive decline in PD, with male patients showing greater risk of cognitive impairments^{30–32}. However, it also needs to be mentioned that several other studies reported no evidence of the impact of sex in cognitive decline in PD^{4,33}. The reason for these conflicting results remains speculative but could be related to the sample size. EHI is a well-known screening tool to determine handedness³⁴. The correlation between the dominant hand and the side of motor symptoms onset has been suggested by several studies^{35–37}. This relationship might extend to the cognitive impairments in PD as some studies suggest^{38,39}.

The severity of motor symptoms has been suggested to be one of the strongest risk factors for cognitive deficits in idiopathic PD^{4,40–42}. UPDRS-III is one of the most widely used screening tools for the severity of motor symptoms in PD and can accurately and efficiently note the presence and progression of those symptoms⁴³. In line with the findings of this study, other studies have also found evidence for a connection between the severity of motor symptoms and emergence of cognitive decline^{44,45}. The number of years of education was also found to be predictive of global cognition in our cohort. Similarly, a large body of evidence reported a negative correlation between higher education level and the likelihood of cognitive impairments in PD patients^{4,30,33}. It has been suggested that education has a role in preserving the cognitive reserve in PD patients at risk of cognitive decline⁴⁶.

The structural biomarkers of cognitive decline in PD have been investigated extensively using different techniques^{33,47,48}. The right parahippocampal gyrus, the top anatomical feature identified in this study, has been reported as one of the main brain regions showing significant Dopamine receptor (D2) binding reduction in PD patients⁴⁹. Another machine learning study reported the parahippocampal region as a top feature showing the highest correlation with the motor score in PD⁵⁰. Similarly, this region was also identified as one of the top features in Alzheimer's disease (AD). These findings suggest a more general function of the parahippocampal

region in neurodegenerative diseases given its prominent role in memory⁵¹. The entorhinal cortex was previously reported as one of the main brain regions allowing a fine distinction between PD-MCI and PD-nonMCI patients⁵². The right entorhinal volume was observed to be positively correlated with memory abilities in early drug-naïve PD-MCI patients⁵². Additionally, cortical thinning of the entorhinal region was found to correlate significantly with memory impairments in PD patients⁵³. The anterior cingulate cortex is another ROI associated with cognitive impairment in PD. A large body of evidence indicates a link between PD-MCI cognitive status and the anterior cingulate^{49,54–56}. These findings are in accordance with the results of the present study and indicate a potential relationship between this region and cognition in PD. The last anatomical feature predictive of global cognition was the right caudate volume. The caudate nucleus is one of the chief regions in PD pathology and extensive loss of neurons in this nucleus was reported in association with cognitive impairment and dementia in PD^{57–60}.

We used a combination of known genetic risk factors (*H1 MAPT*, $\epsilon 4$ *APOE*, *COMT p.Val158Met*, *DAT1 VNTR*, *BDNF p.Val66Met*) and novel genetic variants to predict the global cognition in this cohort^{12,30,60}. The rs894280 was selected as the novel finding for the post-hoc analysis because of its importance in the machine learning model. Ranked as the second top feature, this variant could present a meaningful role in prediction of global cognition in this cohort. This finding was in contrast with the known genetic risk factors included in this study which were not selected by the machine learning model. This variant is an intronic polymorphism located on the 5' region of *SNCA* gene and was initially reported in association with dementia with Lewy bodies (DLB)⁶¹. The role of *SNCA* gene mutations in familial PD has been known for decades. However, new data suggest a role for this gene in cognitive deficits and dementia in idiopathic PD^{12,13,62,63}. A recent study indicated association between several *SNCA* variants and worse performance in Mini Mental state examination (MMSE) in PD patients⁶³. Similar association was reported on the association of several *SNCA* variants and PDD¹². A microsatellite (Rep1) is located on the 3' region of the *SNCA* gene and has two common alleles (short repeat and long repeat). The long repeat allele of Rep1 seems to increase *SNCA* transcription and was reported to be linked to lower MMSE scores in PD patients⁶². On the other hand, the 5' region of the *SNCA* gene was considered as a haplotype specific for DLB and not PD. This evidence was further supported by another study investigating the *SNCA* role in both DLB and PD¹². Both PDD and DLB are classified as synucleinopathies and share substantial similarities in symptoms and pathology, to the point that the exact differentiation of these two disorders clinically and pathologically are still a matter of debate⁶⁴. The rs894280 has been reported in association with both DLB and PDD and this might suggest a more general role for this variant in LB pathology.

Furthermore, rs894280 is in linkage disequilibrium (LD) with rs1348224 with comparable odds ratio ($D' = 1.0$, $R^2 = 1.0$). The rs1348224 variant was previously reported in association with PDD, surviving multiple testing in a sample of 1492 PD patients¹². Moreover, a strong correlation was reported between rs894280 and the Hopkins Verbal Learning Test-Revised (HVLT-R) total recall in PD patients ($p = 6.1 \times 10^{-4}$), and it displayed the strongest relationship with cognitive abilities out of 39 *SNCA* variants included in that study. However, this association did not survive after correction for multiple comparisons. This could indicate a role for rs894280 in PD cognitive abilities, especially in the memory domain⁶⁵. A Brazilian study found cognitive impairments in PD patients carrying T allele of rs2583988 of *SNCA*. The rs2583988 is in a strong LD with rs894280 ($D' = 1.0$, $R^2 = 0.40$ in European descendent populations), which further indicates a possible role for rs894280 in cognitive decline of PD patients¹³.

Deficits in the attention, visuo-spatial, and memory domains are frequently reported in PD-MCI patients^{3,66,67}. Association of rs894280 with impairments in these domains in idiopathic PD patients may indicate a role for this variant in the development of such deficits. Specifically, this SNP might be connected to visuo-spatial abilities given that attention measures used in this study have a prominent visuo-spatial component. Studies have shown that attention measures with a visual component can tap on to visuo-spatial abilities⁶⁸. Out of three attention measures used in this study, two of them; Trail A and Symbol Span have the required component to engage both visuo-spatial and attention suggesting a potential role for this SNP in connection to visuo-spatial abilities.

We did not observe any association between executive function and language domains and rs894280 in our cohort. A possible explanation for this could be that executive function impairments involve the frontal-striatal areas while most of the cognitive deficits identified in this study are focused in more medial temporal lobe and posterior cortical regions³⁰. Although deficits in the language domain are reported in association with dementia in PD³⁰, this SNP did not show any link to language abilities in this cohort.

This study had some limitations that should be mentioned. We used a machine learning approach in this study in an effort to capture the underlying complex patterns of the various input features and focused on the most unique and relevant features for further investigation. These findings are preliminary and need replication in larger cohorts. The present cohort size was small for a genetic analysis, but our results displayed a fair level of robustness, in a cohort that is extensively phenotyped and well-characterized. These results need to be replicated in a larger cohort with higher number of genetic variants to avoid missing effect of other potential risk variants before a definite conclusion can be inferred on this specific variant and cognitive impairment in PD.

In conclusion, using machine learning, we found that rs894280 in *SNCA* was one of the top features predictive of cognition in PD patients. Further analysis in the same cohort revealed association of this variant (CC genotype) with attention and visuo-spatial abilities in PD patients with a trend in the same direction for memory abilities. These results indicate a potential involvement of *SNCA* variant rs894280 in the cognitive deficits and even dementia in idiopathic PD patients.

Gene	Variant	TaqMan genotyping assay
SNCA	rs7689942	C-28994912-10
	rs894280	C-8933128-10
APOE	rs7412	C-904973-10
	rs429358	C-3084793-20
COMT	rs4680	C-25746809-50
MAPT	rs393152	C-2265265-10
BDNF	rs6265	C-11592758_10
SLC6A3*	VNTR 9/10 R	F-TGTGGTGTAGGGAACGGCCTGAG R-CTTCTGGAGGTCACGGCTCAAGG

Table 5. The list of the TaqMan genotyping assays and the primers used for genotyping. **SLC6A3* VNTR was genotyped using custom made primers. For details please refer to methods.

Methods

Participants. 101 PD patients at Hoehn and Yahr stages II-III were recruited. All patients had a confirmed diagnosis of idiopathic PD by a movement disorder clinic neurologist, meeting the UK brain bank criteria for idiopathic PD. All participants were responsive to dopaminergic medications and took their usual dosage of medications during all study visits. None of the participants were asked to modify their medications for this study. Exclusion criteria were: 1) any neurological disorder other than PD, 2) alcohol dependency, 3) history or presence of a severe psychiatric disorder, and 4) cerebrovascular disorders. The severity of motor symptoms was assessed by a trained professional using the motor section of the Unified Parkinson's Disease Rating Scale (UPDRS-III). Levodopa Equivalency Daily Dosage (LEDD) and disease duration of all participants were calculated by a research nurse.

All participants provided written informed consent according to the declaration of Helsinki and the study was approved by the Conjoint Health Research Ethics Board (REB14-2463) at the University of Calgary, AB, Canada. All methods were carried out in accordance with the relevant guidelines and regulations.

Genotyping. A blood sample was collected from each participant and DNA was extracted using an isopropanol-based protocol. DNA samples were screened for several Single Nucleotide Polymorphisms (SNP) using TaqMan genotyping assays on a C-1000 Touch Thermal Cycler. The list of all SNPs and TaqMan assays investigated in this study are shown in Table 5. TaqMan assay reading was done using Applied Biosystem Quantstudio Flex 7 Real-Time PCR system (Fisher Scientific) according to the manufacturer's instructions. TaqMan (assays) results were analyzed using Bio-Rad CFX Maestro software. The 40 bp Variable Tandem Repeats (VNTR) located on the 3' region of Solute Carrier 6 family 3 (*SLC6A3*) was amplified using PCR (30 s at 95 °C, 36 cycles of (95 °C for 15 s, 60 °C for 30 s, 70 °C for 60 s), 68 °C for 5 min and 4 °C for hold) on a C-1000 Touch Thermal Cycler (Biorad), using the primers and protocol described previously⁶⁹. PCR products were mixed with loading dye and loaded on 2% agarose gel containing gelstar and run at 120 V for 30 min followed by 60 min at 100 V. A 100 bp DNA ladder (Biohelix, DM 001-R500F, FroggaBio) was loaded on each gel to determine the molecular size of PCR products. The length of PCR products was captured using the Chemidoc Imaging System (Biorad).

Neuropsychological assessment. All participants completed a comprehensive cognitive assessment. The cognitive battery applied in this study consists of tests and measures covering five cognitive domains: executive function, memory, attention, visuo-spatial, and language. The full list of cognitive tests is shown in Table 6. All tests were scored by a trained psychometrist. The cognitive tests were first scored using the test makers manual, which details specific parameters to evaluate an examinee's performance. The total raw score is then converted to a standardized score by comparing the examinee's score to other healthy controls matched for age. All neuropsychological tests that were corrected for age, and most were corrected for years of education, and sex.

The measures corresponding to the same cognitive domain were averaged to obtain the average Z-score for each of the cognitive domains. The global cognition Z-score was calculated by averaging all cognitive domains' Z-scores. The Edinburgh Handedness Inventory (EHI) was administered to the participants and scored to identify each participant's dominant hand.

MRI acquisition. Each participant had an MRI scan within two weeks of the neuropsychological assessment using the GE Discovery 750 3 T MRI at the Seaman Family Imaging Centre at the University of Calgary, Calgary, Alberta, Canada. A high-resolution T1-weighted 3D inversion recovery prepared fast spoiled gradient recalled (IR-FSPGR) sequence was acquired for each participant (repetition time = 7.176 ms, echo time = 2.252 ms, flip angle = 10°, acquisition matrix = 256 × 256, voxel size = 1 × 1 × 1 mm³, 172 slices).

Cortical thickness and subcortical volume. Freesurfer (<http://surfer.nmr.mgh.harvard.edu/>; version 6.0.0) was used to perform cortical thickness and subcortical volumetric analyses. The analysis was performed following the procedure detailed in prior publications⁷⁰⁻⁷². Cortical segmentation was performed automatically and upon visual inspection, appropriate manual adjustments were made. The manual editing was carefully performed in accordance with the Freesurfer manual in several steps (<https://surfer.nmr.mgh.harvard.edu/fswiki/>

Cognitive domain	Cognitive test/measure
Attention	Trail A Wechsler Memory Scale (WMS) symbol span Digit-span Forward
Executive function	Hayling 2 Brixton Trail B Stroop Colour/Word Clock drawing command
Language	Animals Actions Boston Naming test (BNT)
Memory	Rey Complex Figure Test (RCFT) delay recall Hopkins Verbal Learning Test (HVLT) Retention HVLT Recognition Logical memory delayed recall
Visuo-spatial	RCFT copy Hooper Visual Organizational Test

Table 6. The neuropsychiatric battery.

Edits). The errors resulting from imperfect intensity normalization were corrected by inserting control points and where appropriate, extraneous tissue were removed from the brain volumes (wm.mgz and brainmask.mgz). A total of 29 segmentation results were manually edited. Cortical thickness was computed for 68 Regions-of-Interest (ROIs) using the Desikan-Killiany brain atlas in Freesurfer⁷³. Furthermore, the mean cortical thickness was calculated for each hemisphere.

Subcortical volumetric measures were computed from eight regions per hemisphere including caudate nucleus, putamen, pallidum, nucleus accumbens, hippocampus, amygdala, thalamus, and ventral diencephalon, plus the brain stem. All subcortical volumetric measures were corrected for the intracranial volume.

Machine learning. The input features used for machine learning were sex, age, EHI score, years of education, years of disease duration, LEDD, UPDRS-III, rs6265, rs7689942, rs894280, rs7412, rs393152, rs429358, rs4680, and *SLC6A3 VNTR*. Furthermore, the following imaging measures were included: cortical thickness measures (68 ROIs), subcortical volumetric measures adjusted for the intracranial volume (17 ROIs), and 2 measures of global thickness. In total, 102 measures were available per participant for the machine learning analysis to predict the outcome variable, the global cognition Z-score.

The machine learning analysis consisted of two main steps: (1) feature ranking and selection, and (2) regression analysis. The feature ranking was employed to rank the 102 features (8 genetic, 87 neuroimaging, 3 clinical, and 4 demographic) based on their contribution to the outcome variable and to select the most efficient combination of features that can predict the outcome variable of the regression problem. Reducing the number of features is essential for improving model performance by eliminating features that are redundant and non-informative. In this study, the RRELIEFF feature selection algorithm was used for this purpose⁷⁴. In the next step, the machine learning regression modelling was performed based on the ranked features using a Support Vector Regression (SVR) model with a polynomial kernel. The SVR is, in principle, very similar to the support vector machine classification models with slight differences for the adaption to a regression style problem⁷⁵. More precisely, an SVR model is built based on only a subset of training data within the predefined margins that minimize the generalization error. Therefore, the data is first transformed into a higher dimensional space employing the polynomial kernel, thereby allowing linear models to fit the training data. The SVR model was used in this study for regression modeling as it is less likely to overfit the data compared to other models, i.e. SVR is a model with adequate generalization capabilities and good prediction accuracy.

The least informative feature was iteratively removed from the set of ranked features until only two features were left for model training to identify the optimal subset of features. The model performance was evaluated for each iteration using the root mean squared error comparing the predicted with true observations. Finally, the model with the optimal feature subset was further evaluated using additional metrics including the coefficient of determination (R^2) and the correlation value. The coefficient of determination quantifies the amount of variance in the outcome variable that is explained by the selected features in the model. A nested leave-one-out method of the cross-validation was employed through the feature selection and regression in which the number of model validation was set to N where N is equal to the number of participants in the sample. At each validation test, one participant is used to test the model while N-1 participants were used to train the model. All metrics reported for machine learning were attained by averaging the metrics of these N models. This method was used to overcome the small sample size and to prevent double-dipping.

Statistical analysis. Statistical analyses of continuous variables were performed using either a student-t test or Mann-Whitney U test based on the data normality. The Fisher exact test was used for categorical variables.

The post-hoc statistical analysis was designed in compliance with the machine learning results. Pearson correlation test was used to select independent factors correlated with the target feature. ANCOVA was used to explore the allelic group differences in the rs894280 variant of the *SCNA* gene with regards to global cognition. Demographic and clinical factors that were significantly correlated with a cognitive domain score of (attention,

language, etc.) were included as covariates in the ANCOVA to control for them. A value of $p < 0.05$ was considered significant for the single tests, and Bonferroni correction was used to correct for multiple testing. The chi-square test was used to explore association of rs894280 with other genetic variants available in this cohort. All statistical tests were performed using IBM Corp. Released 2019. IBM SPSS Statistics for McIntosh, Version 26.0. Armonk, NY: IBM Corp.

Received: 31 October 2020; Accepted: 10 February 2021

Published online: 01 March 2021

References

- Aarsland, D. *et al.* Cognitive decline in Parkinson disease. *Nat. Rev. Neurol.* **13**, 217–231. <https://doi.org/10.1038/nrneuro.2017.27> (2017).
- Galtier, I., Nieto, A., Lorenzo, J. N. & Barroso, J. Mild cognitive impairment in Parkinson's disease: diagnosis and progression to dementia. *J. Clin. Exp. Neuropsychol.* **38**, 40–50 (2016).
- Litvan, I. *et al.* MDS Task Force on mild cognitive impairment in Parkinson's disease: critical review of PD-MCI. *Mov. Disord.* **26**, 1814–1824. <https://doi.org/10.1002/mds.23823> (2011).
- Litvan, I. *et al.* Diagnostic criteria for mild cognitive impairment in Parkinson's disease: Movement Disorder Society Task Force guidelines. *Mov. Disord.* **27**, 349–356. <https://doi.org/10.1002/mds.24893> (2012).
- Moore, D. J., West, A. B., Dawson, V. L. & Dawson, T. M. Molecular pathophysiology of Parkinson's disease. *Annu. Rev. Neurosci.* **28**, 57–87. <https://doi.org/10.1146/annurev.neuro.28.061604.135718> (2005).
- Spillantini, M. G. *et al.* Alpha-synuclein in Lewy bodies. *Nature* **388**, 839–840 (1997).
- Uéda, K. *et al.* Molecular cloning of cDNA encoding an unrecognized component of amyloid in Alzheimer disease. *Proc. Natl. Acad. Sci. U. S. A.* **90**, 11282–11286 (1993).
- Polymeropoulos, M. H. *et al.* Mutation in the alpha-synuclein gene identified in families with Parkinson's disease. *Science* **276**, 2045–2047 (1997).
- Singleton, A. B. *et al.* alpha-Synuclein locus triplication causes Parkinson's disease. *Science* **302**, 841. <https://doi.org/10.1126/science.1090278> (2003).
- Ross, O. A. *et al.* Genomic investigation of alpha-synuclein multiplication and Parkinsonism. *Ann. Neurol.* **63**, 743–750. <https://doi.org/10.1002/ana.21380> (2008).
- Atik, A., Stewart, T. & Zhang, J. Alpha-synuclein as a biomarker for Parkinson's disease. *Brain Pathol.* **26**, 410–418 (2016).
- Guella, I. *et al.* alpha-synuclein genetic variability: a biomarker for dementia in Parkinson disease. *Ann. Neurol.* **79**, 991–999. <https://doi.org/10.1002/ana.24664> (2016).
- Campelo, C. L. C. *et al.* Variants in SNCA gene are associated with Parkinson's disease risk and cognitive symptoms in a Brazilian sample. *Front. Aging Neurosci.* **9**, 198. <https://doi.org/10.3389/fnagi.2017.00198> (2017).
- Libbrecht, M. W. & Noble, W. S. Machine learning applications in genetics and genomics. *Nat. Rev. Genet.* **16**, 321–332. <https://doi.org/10.1038/nrg3920> (2015).
- Fuchs, J. *et al.* Phenotypic variation in a large Swedish pedigree due to SNCA duplication and triplication. *Neurology* **68**, 916–922. <https://doi.org/10.1212/01.wnl.0000254458.17630.c5> (2007).
- Nagano-Saito, A. *et al.* Cerebral atrophy and its relation to cognitive impairment in Parkinson disease. *Neurology* **64**, 224–229. <https://doi.org/10.1212/01.WNL.0000149510.41793.50> (2005).
- Beyer, M. K., Janvin, C. C., Larsen, J. P. & Aarsland, D. Amagnetic resonance imaging study of patients with Parkinson's disease with mild cognitive impairment and dementia using voxelbased morphometry. *J. Neurol.* **78**, 254–259 (2007).
- Weintraub, D. *et al.* Neurodegeneration across stages of cognitive decline in Parkinson disease. *Arch. Neurol.* **68**, 1562–1568 (2011).
- Weintraub, D. *et al.* Alzheimer's disease pattern of brain atrophy predicts cognitive decline in Parkinson's disease. *Brain* **135**, 170–180 (2012).
- Somme, J. H. *et al.* Initial neuropsychological impairments in patients with the E46K mutation of the a-synuclein gene (PARK 1). *J. Neurol. Sci.* **310**, 86–89 (2011).
- Camicioli, R. *et al.* Voxel-based morphometry reveals extra-nigral atrophy patterns associated with dopamine refractory cognitive and motor impairment in Parkinsonism. *Parkinsonism Relat. Disord.* **15**, 187–195 (2009).
- Hanganu, A. *et al.* Mild cognitive impairment in patients with Parkinson's disease is associated with increased cortical degeneration. *Mov. Disord.* **28**, 1360–1369 (2013).
- Ramezani, M. *et al.* Synuclein-alpha 's variant, rs 894280 shows association with memory and attention in PD patients. *MDS international Congress* **1725** (2019).
- Urbanowicz, R. J., Meeker, M., La Cava, W., Olson, R. S. & Moore, J. H. Relief-based feature selection: introduction and review. *J. Biomed. Inf.* **85**, 189–203 (2019).
- Beretta, L. & Santaniello, A. Implementing relief filters to extract meaningful features from genetic lifetime datasets. *J. Biomed. Inf.* **44**, 361–369 (2011).
- Eppstein, M. J., Payne, J. L., White, B. C. & Moore, J. H. Genomic mining for complex disease traits with random chemistry. *Genet. Prog. Evol. Mach.* **8**, 395–411 (2007).
- McKinney, B. A. *et al.* Reliefseq: a gene-wise adaptive-k nearest-neighbor feature selection tool for finding gene-gene interactions and main effects in mrna-seq gene expression data. *PLoS ONE* **8**, e81527 (2013).
- Lei, H. *et al.* Joint regression and classification via relational regularization for Parkinson's disease diagnosis. *Technol. Health Care* **26**, 19–30 (2018).
- Nilashi, M., Ibrahim, O. & Ahani, A. Accuracy Improvement for Predicting Parkinson's Disease Progression. *Nature* **6**, 1–18 (2016).
- Williams-Gray, C. H. *et al.* The distinct cognitive syndromes of Parkinson's disease: 5 year follow-up of the CamPaIGN cohort. *Brain* **132**, 2958–2969. <https://doi.org/10.1093/brain/awp245> (2009).
- Reekes, T. H. *et al.* Sex specific cognitive differences in Parkinson disease. *NPJ Parkinsons Dis.* **6**, 7. <https://doi.org/10.1038/s41531-020-0109-1> (2020).
- Cholerton, B. *et al.* Sex differences in progression to mild cognitive impairment and dementia in Parkinson's disease. *Parkinsonism Relat. Disord.* **50**, 29–36. <https://doi.org/10.1016/j.parkreldis.2018.02.007> (2018).
- Cammisuli, D. M., Cammisuli, S. M., Fusi, J., Franzoni, F. & Pruneti, C. Parkinson's disease-mild cognitive impairment (PD-MCI): a useful summary of update knowledge. *Front. Aging Neurosci.* **11**, 303. <https://doi.org/10.3389/fnagi.2019.00303> (2019).
- Oldfield, R. C. The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* **9**, 97–113. [https://doi.org/10.1016/0028-3932\(71\)90067-4](https://doi.org/10.1016/0028-3932(71)90067-4) (1971).
- Yust-Katz, S., Tesler, D., Treves, T. A., Melamed, E. & Djaldetti, R. Handedness as a predictor of side of onset of Parkinson's disease. *Parkinsonism Relat. Disord.* **14**, 633–635 (2008).
- van der Hoorn, A., Burger, H., Leenders, K. L. & de Jong, B. M. Handedness correlates with the dominant Parkinson side: a systematic review and meta-analysis. *Mov. Disord.* **27**, 206–210 (2012).

37. Prasad, S., Saini, J., Yadav, R. & Pal, P. K. Motor asymmetry and neuromelanin imaging: Concordance in Parkinson's disease. *Parkinsonism Relat. Disord.* **53**, 28–32 (2018).
38. Claassen, D. O. *et al.* Cortical asymmetry in Parkinson's disease: early susceptibility of the left hemisphere. *Brain Behav.* **6**, e00573 (2016).
39. Modestino, E. J., Amenechi, C., Reinhofer, A. & O'Toole, P. Side-of-onset of Parkinson's disease in relation to neuropsychological measures. *Brain Behav.* **19**, e00590 (2016).
40. Verbaan, D. *et al.* Cognitive impairment in Parkinson's disease. *J. Neurol. Neurosurg. Psychiatry* **78**, 1182–1187. <https://doi.org/10.1136/jnnp.2006.112367> (2007).
41. Tomer, R., Levin, B. E. & Weiner, W. J. Side of onset of motor symptoms influences cognition in Parkinson's disease. *Ann. Neurol.* **34**, 579–584 (1993).
42. Aarsland, D. *et al.* Frequency of dementia in Parkinson disease. *Arch. Neurol.* **53**, 538–542 (1996).
43. Bloem, B. R. *et al.* Measurement instruments to assess posture, gait, and balance in Parkinson's disease: critique and recommendations. *Mov. Disord.* **31**, 1342–1355 (2016).
44. Karr, J. E., Graham, R. B., Hofer, S. B. & Muniz-Terrera, G. When does cognitive decline begin? A systematic review of change point studies on accelerated decline in cognitive and neurological outcomes preceding mild cognitive impairment, dementia, and death. *Psychol. Aging* **33**, 195–218 (2019).
45. Lin, Y. S., Lee, W. J., Wang, S. J. & Fuh, J. L. Levels of plasma neurofilament light chain and cognitive function in patients with Alzheimer or Parkinson disease. *Sci. Rep.* <https://doi.org/10.1038/s41598-018-35766-w> (2018).
46. Armstrong, M. J. *et al.* Roles of education and IQ in cognitive reserve in Parkinson's disease-mild cognitive impairment. *Dement. Geriatr. Cogn. Disord. Extra* **2**, 343–352 (2012).
47. Monchi, O., Hanganu, A. & Bellec, P. Markers of cognitive decline in PD: the case for heterogeneity. *Parkinsonism Relat. Disord.* **24**, 8–14. <https://doi.org/10.1016/j.parkreldis.2016.01.002> (2016).
48. Mateos-Perez, J. M. *et al.* Structural neuroimaging as clinical predictor: A review of machine learning applications. *Neuroimage Clin.* **20**, 506–522. <https://doi.org/10.1016/j.nicl.2018.08.019> (2018).
49. Christopher, L. *et al.* Saliency network and parahippocampal dopamine dysfunction in memory-impaired Parkinson disease. *Ann. Neurol.* **77**, 269–280 (2015).
50. Nguyen, A. A., Maia, P. D., Gao, X., Damasceno, P. F. & Raj, A. Dynamical role of pivotal brain regions in Parkinson symptomatology uncovered with deep learning. *Brain Sci.* <https://doi.org/10.3390/brainsci10020073> (2020).
51. Zhang, D. *et al.* Multimodal classification of Alzheimer's disease and mild cognitive impairment. *Neuroimage* **55**, 856–867 (2011).
52. Jia, X. *et al.* Entorhinal cortex atrophy in early, drug-naive Parkinson's disease with mild cognitive impairment. *Aging Dis.* **10**, 1221–1232 (2019).
53. Pagonabarraga, J. *et al.* Pattern of regional cortical thinning associated with cognitive deterioration in Parkinson's disease. *PLoS ONE* **8**, e54980 (2013).
54. Ekman, U. *et al.* Functional brain activity and presynaptic dopamine uptake in patients with Parkinson's disease and mild cognitive impairment: a cross-sectional study. *Lancet Neurol.* **8**, 679–687 (2012).
55. Monchi, O. & Stoessel, A. J. Imaging neural correlates of mild cognitive impairment in Parkinson's disease. *Lancet Neurol.* **11**, 653–655 (2012).
56. Katako, A. *et al.* Machine learning identified an Alzheimer's disease-related FDG-PET pattern which is also expressed in Lewy body dementia and Parkinson's disease dementia. *Sci. Rep.* **8**, 13236. <https://doi.org/10.1038/s41598-018-31653-6> (2018).
57. Monchi, O. *et al.* Neural bases of set-shifting deficits in Parkinson's disease. *J. Neurosci.* **24**, 702–710. <https://doi.org/10.1523/JNEUROSCI.4860-03.2004> (2004).
58. Halliday, G. M., Leverenz, J. B., Schneider, J. S. & Adler, C. H. The Neurobiological basis of cognitive impairment in Parkinson's disease. *Mov. Disord.* **29**, 634–650 (2014).
59. Palumbo, B. *et al.* Diagnostic accuracy of Parkinson disease by support vector machine (SVM) analysis of 123I-FP-CIT brain SPECT data: implications of putaminal findings and age. *Medicine (Baltimore)* **93**, 228. <https://doi.org/10.1097/MD.0000000000000228> (2014).
60. Habak, C. *et al.* Dopamine transporter SLC6A3 genotype affects cortico-striatal activity of set-shifts in Parkinson's disease. *Brain* **137**, 3025–3035. <https://doi.org/10.1093/brain/awu251> (2014).
61. Bras, J. *et al.* Genetic analysis implicates APOE, SNCA and suggests lysosomal dysfunction in the etiology of dementia with Lewy bodies. *Hum. Mol. Genet.* **23**, 6139–6146. <https://doi.org/10.1093/hmg/ddu334> (2014).
62. Ng, A. S. L. *et al.* SNCA Rep1 promoter variability influences cognition in Parkinson's disease. *Mov. Disord.* **34**, 1232–1236. <https://doi.org/10.1002/mds.27768> (2019).
63. Senkevich, K. *et al.* Associations of genetic variants in COMT, BDNF, SNCA, MAPT genes with cognitive impairment in Parkinson's disease. (P6.090). *Neurology* **90** (2018).
64. Goldman, J. G., Williams-Gray, C., Barker, R. A., Duda, J. E. & Galvin, J. E. The spectrum of cognitive impairment in Lewy body diseases. *Mov. Disord.* **29**, 608–621. <https://doi.org/10.1002/mds.25866> (2014).
65. Mata, I. F. *et al.* Large-scale exploratory genetic analysis of cognitive impairment in Parkinson's disease. *Neurobiol. Aging* **56**(211), e211–e217. <https://doi.org/10.1016/j.neurobiolaging.2017.04.009> (2017).
66. Pfeiffer, H. C., Lokkegaard, A., Zoetmulder, M., Friberg, L. & Werdelin, L. Cognitive impairment in early-stage non-demented Parkinson's disease patients. *Acta Neurol. Scand.* **129**, 307–318. <https://doi.org/10.1111/ane.12189> (2014).
67. Cammisuli, D. M. & Crowe, S. Spatial disorientation and executive dysfunction in elderly nondemented patients with Parkinson's disease. *Neuropsychiatr. Dis. Treat.* **14**, 2531–2539 (2018).
68. Donolato, E., Giofre, D. & Mammarella, I. C. Differences in verbal and visuospatial forward and backward order recall: a review of the literature. *Front. Psychol.* **8**, 663. <https://doi.org/10.3389/fpsyg.2017.00663> (2017).
69. Shinohara, M. *et al.* Eating disorders with binge-eating behaviour are associated with the s allele of the 3'-UTR VNTR polymorphism of the dopamine transporter gene. *J. Psychiatr. Neurosci.* **29**, 134–137 (2004).
70. Dale, A. M., Fischl, B. & Sereno, M. I. Cortical surface-based analysis. I. Segmentation and surface reconstruction. *Neuroimage* **9**, 179–194 (1999).
71. Fischl, B., Serenob, M. I. & Dalea, A. M. Cortical surface-based analysis: II: inflation, flattening, and a surface-based coordinate system. *Neuroimage* **9**, 195–207 (1999).
72. Yoon, E. J. *et al.* Mild behavioral impairment is linked to worse cognition and brain atrophy in Parkinson disease. *Neurology* **98**, e766–e777 (2019).
73. Desikan, R. S. *et al.* An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* **31**, 968–980 (2006).
74. Kira, K. & Rendell, L. A. The feature selection problem: traditional methods and a new algorithm. *AAAI-92 Proceedings* (1992).
75. Vapnik, V., Golowich, S. E. & Smola, A. J. Support vector method for function approximation, regression estimation and signal processing. *Adv. Neural Inf. Process. Syst.* **9**, 281–287 (1997).

Acknowledgments

This work was funded by a project grant from the Canadian Institutes of Health Research (CIHR) (PJT-166123), the Tourmaline Oil Chair in Parkinson's Disease, the Canada Research Chair in non-motor symptoms of Parkinson's disease to O.M. MR received a scholarship from the Parkinson Association of Alberta. NDF is supported by the Canada Research Chair program and the River Fund at Calgary Foundation. This work was financially supported by Parkinson Canada, the Michael J. Fox Foundation, the Canadian Consortium on Neurodegeneration in Aging (CCNA) and the Canada First Research Excellence Fund (CFREF), awarded to McGill University for the Healthy Brains for Healthy Lives (HBHL) program. ZGO is supported by the Fonds de recherche du Québec-Santé (FRQS) Chercheurs-boursiers award and is a Parkinson's Disease Canada New Investigator awardee. Infrastructure for work performed in the GP lab was supported by a grant from the Canada Foundation for Innovation John R Evans Leaders Fund.

Author contributions

M.R. has prepared the data, performed the statistical analysis, interpreted the results, wrote and edited the manuscript. P.M. and D.R. performed the machine learning analysis and helped with Methods section. PM also helped with manuscript edition. E.Y. prepared the structural neuroimaging data and edited the manuscript. J.A.R., E.L. and K.M. prepared the genetic data. M.K. administered, scored and prepared the neuropsychological data. TH calculated and collected the clinical data. I.K. and N.M. recruited the participants and collected clinical data. JS and DM were involved in data collection and manuscript edition. G.P. and Z.G.O. were involved in genetic data collection. N.D.F. supervised the machine learning analysis and manuscript writing and editing. O.M. was involved in study design, supervision of the project and manuscript writing and edition.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to O.M.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021