

## ARTICLE OPEN



# Biomarker characterization of clinical subtypes of Parkinson Disease

Xiao Deng<sup>1,2</sup>, Seyed Ehsan Saffari<sup>3</sup>, Nan Liu<sup>4</sup>, Bin Xiao<sup>1,2</sup>, John Carson Allen<sup>3</sup>, Samuel Yong Ern Ng<sup>1</sup>, Nicole Chia<sup>1</sup>, Yi Jayne Tan<sup>1</sup>, Xinyi Choi<sup>1</sup>, Dede Liana Heng<sup>1</sup>, Yew-long Lo<sup>1</sup>, Zheyu Xu<sup>1</sup>, Kay-Yaw Tay<sup>1</sup>, Wing-Lok Au<sup>1,2</sup>, Adeline Ng<sup>1,2</sup>, Eng-King Tan<sup>1,2,5</sup> and Louis C. S. Tan<sup>1,2,5</sup>✉

The biological underpinnings of the PD clusters remain unknown as the existing PD clusters lacks biomarker characterization. We try to identify clinical subtypes of Parkinson Disease (PD) in an Asian cohort and characterize them by comparing clinical assessments, genetic status and blood biochemical markers. A total of 206 PD patients were included from a multi-centre Asian cohort. Hierarchical clustering was performed to generate PD subtypes. Clinical and biological characterization of the subtypes were performed by comparing clinical assessments, allelic distributions of Asian related PD gene (SNCA, LRRK2, Park16, ITPKB, SV2C) and blood biochemical markers. Hierarchical clustering method identified three clusters: cluster A (severe subtype in motor, non-motor and cognitive domains), cluster B (intermediate subtype with cognitive impairment and mild non-motor symptoms) and cluster C (mild subtype and young age of onset). The three clusters had significantly different allele frequencies in two SNPs (Park16 rs6679073 A allele carriers in cluster A B C: 67%, 74%, 89%,  $p = 0.015$ ; SV2C rs246814 T allele distribution: 7%, 12%, 25%,  $p = 0.026$ ). Serum homocysteine (Hcy) and C-reactive protein (CRP) levels were also significantly different among three clusters (Mean levels of Hcy and CRP among cluster A B C were:  $19.4 \pm 4.2$ ,  $18.4 \pm 5.7$ ,  $15.6 \pm 5.6$ , adjusted  $p = 0.005$ ;  $2.5 \pm 5.0$ ,  $1.5 \pm 2.4$ ,  $0.9 \pm 2.1$ , adjusted  $p < 0.0001$ , respectively). Of the 3 subtypes identified amongst early PD patients, the severe subtype was associated with significantly lower frequency of Park16 and SV2C alleles and higher levels of Hcy and CRP. These biomarkers may be useful to stratify PD subtypes and identify more severe subtypes.

npj Parkinson's Disease (2022)8:109; <https://doi.org/10.1038/s41531-022-00375-y>

## INTRODUCTION

Parkinson's disease (PD) is the most common hypokinetic movement disorder with significant heterogeneity in symptoms and outcomes. Non-Motor symptoms (NMS) resulting from various neurotransmitter pathway dysfunctions affects both the central and peripheral nervous systems, which contribute to PD heterogeneity<sup>1,2</sup>. Subtype identification has been established as one of the top three clinical research priorities in the field of PD<sup>3</sup>. Identification of PD subtype could be valuable in revealing the underlying etiology and understanding the disease course. More importantly, PD subtyping could guide the design of clinical trial and future personalized PD treatment.

Cluster analysis, a data-driven approach could help to define the disease phenotypes. Most studies use cluster analysis to stratify PD subtypes based on clinical data, such as motor, NMS and demographic features<sup>4,5</sup>. These studies have however been limited by the inclusion of PD patients from different disease stages; absence of genetic data that may influence clinical heterogeneity; and limited analysis of Asian cohorts.

The biological underpinnings of the PD clusters remain unknown as the existing multidimensional data-driven derivation of PD clusters lacks biomarker characterisation. PD biomarkers including clinical, blood, cerebrospinal fluid (CSF) and imaging biomarkers have been playing increasingly important roles in early diagnosis and disease prognostication<sup>6</sup>. Blood biomarkers may have wider implications than CSF and imaging biomarkers as they are more accessible at lower cost<sup>7</sup>.

Homocysteine (Hcy) and C-reactive protein (CRP) are two blood biochemical biomarkers that are associated with PD severity. Severe PD subtypes have been found to have significantly higher levels of CRP<sup>8</sup> while elevated plasma Hcy level was found in depressed and cognitively impaired PD patients<sup>9</sup>. Vitamin D, uric acid(UA) and lipids are thought to play important neuroprotective roles in PD. Studies have shown that higher serum UA levels were associated with a lower risk of PD development<sup>10</sup> and more benign prognosis in PD patients<sup>11</sup>. Vitamin D deficiency is common in PD patients<sup>12</sup> and lower vitamin D levels have been associated with worse prognosis in PD<sup>13</sup>. In addition, serum lipid biomarkers were reduced in PD patients compared to healthy controls<sup>14–17</sup>.

To understand the clinical heterogeneity of PD, we used cluster analysis to search for subtypes in a multi-centre, Asian early PD cohort. The aims of our study are: (1) To identify distinct PD clusters from a comprehensive dataset; (2) To provide clinical and biological features of the identified subtypes by comparing the clinical characteristics, allelic distributions of Asian related PD genes, and blood biochemical markers.

## RESULTS

### Patient demographic and clinical characteristics

A total of 206 PD patients were enrolled in the study and 122 (59.2%) patients were male. Mean age of diagnosis was  $63.5 \pm 9.0$  years. MCI was presented in 108 (52.4%) patients at baseline in

<sup>1</sup>Department of Neurology, National Neuroscience Institute, Singapore, Singapore. <sup>2</sup>Duke-NUS Medical School, Singapore, 8 College Rd, Singapore 169857, Singapore. <sup>3</sup>Centre for Quantitative Medicine, Duke-NUS Medical School, Singapore, Singapore. <sup>4</sup>Programme in Health Services and Systems Research, Duke-NUS Medical School, Singapore, Singapore. <sup>5</sup>These authors jointly supervised this work: Eng-King Tan, Louis C. S. Tan. ✉email: louis.tan.c.s@singhealth.com.sg

**Table 1.** Patient demographics and clinical characteristics.

Variable	Whole cohort ( <i>n</i> = 206)
Sex: Male	122 (59.2%)
Ethnicity:	
Chinese	177 (85.9%)
Malay	11 (5.3%)
Indian	14 (6.8%)
Others	4 (2%)
Education (Year)	10.6 ± 4.4
Age of diagnosis (year)	63.5 ± 9.0
PRS	0.7 (0.3–1.0)
Patients with RBD symptoms	51 (24.8%)
MCI	108 (52.4%)
Memory score	−1.0 (−1.9–0.2)
Visuospatial score	−0.5 (−1.2–0.2)
Attention score	−0.3 (−0.9–0.3)
Language score	−0.1 (−1.0–0.3)
Executive score	−0.2 (−0.8–0.4)
MDS-UPDRS part I score	3 (2–6)
MDS-UPDRS part II score	3 (1–6)
MDS-UPDRS part III score	20 (15–26)
PIGD score	1 (1–2)
Tremor score	3 (1–5)
HADS Anxiety Total score	2 (0–4)
HADS Depression Total score	2 (1–4)
NMSS Total score	14 (9–26)
LEDD	194.8 ± 138.0

Categorical variables reported as frequency (%); continuous variables reported as mean ± standard deviation or median and first and third quartile, where appropriate.

PRS polygenic risk score, RBD rapid eye movement sleep behaviour disorder, MCI mild cognitive impairment, MDS-UPDRS movement disorder society-unified Parkinson's disease rating scale, PIGD postural instability and gait disorder, HADS Hospital Anxiety Depression Scale, NMSS Non-motor symptom scale, LEDD levodopa equivalent daily dose.

our cohort. A summary of patient demographics and clinical characteristics was shown in Table 1. The comparison of comorbidities among three clusters can be found in Supplementary Table 1.

### Cluster analysis results

Three independent PD clusters were identified from hierarchical cluster analysis (Fig. 1). The features comparison among three clusters can be seen in Table 2. The three clusters had significant differences in age of diagnosis (mean age of diagnosis of cluster A, B, C was 69.6 ± 7.9, 63.6 ± 7.4, 59.4 ± 9.7 years,  $p < 0.001$ , respectively). The three clusters also differed significantly in all cognitive domain scores, most of the motor scores (MDS-UPDRS part II, III score, tremor score, PIGD score) and most NMS items (MDS-UPDRS part I score, systolic BP drop, ESS total score and HADS depression score). There were no significant differences in terms of PRS, HADS anxiety score and RBD1Q among the three clusters.

Cluster A was severe subtype in motor, NMS and cognition, which comprised 43 (20.9%) PD patients. Cluster A was the most severe in terms of motor, NMS and cognition domains as supported by the highest MDS-UPDRS part I, II, III scores, PIGD score, cognitive domain scores and depression score. Significant BP drop was the most common in cluster A (35% of the patients

in cluster A had significant BP drop vs 21% and 11% in clusters B and C,  $p = 0.010$ ).

The second cluster (cluster B) was the largest cluster with 98 subjects, consisting of 47.6% of PD patients. Cluster B was the intermediate subtype characterized by cognitive impairment with mild NMS. Cognitive domain scores in this cluster were moderate, ranging between cluster A and C. However, patients in this cluster had very mild NMS impairment supported by the lowest MDS-UPDRS part I score (2.5 in cluster B vs 6 and 4 in cluster A and C,  $p < 0.001$ ).

There were 65 patients in cluster C, accounting for 31.6% of the PD patients. Cluster C was a mild subtype with a younger age of onset. The mean age of diagnosis of cluster C was significantly younger than the other two clusters (59.4 ± 9.7 in cluster C vs 63.6 ± 7.4, 69.6 ± 7.9 in cluster B and A,  $p < 0.001$ ). Cluster C had good cognitive performance supported by the highest cognitive domain scores. Significant BP drop was not common in the cluster, with only 7 patients having significant systolic blood pressure drop. The other non-motor and motor profiles in cluster C were relatively mild.

### Characterization of PD subtypes using clinical biomarkers

Clinical variables that were not included in the clustering model were used for post-hoc comparison among the clusters (Table 3). The 3 clusters remained significantly different with regard to MCI rate, MoCA Score, and most of the NMSS domain scores, except domain 4 (perceptual problems) and domain 8 (sexual function), after correction for multiple comparison.

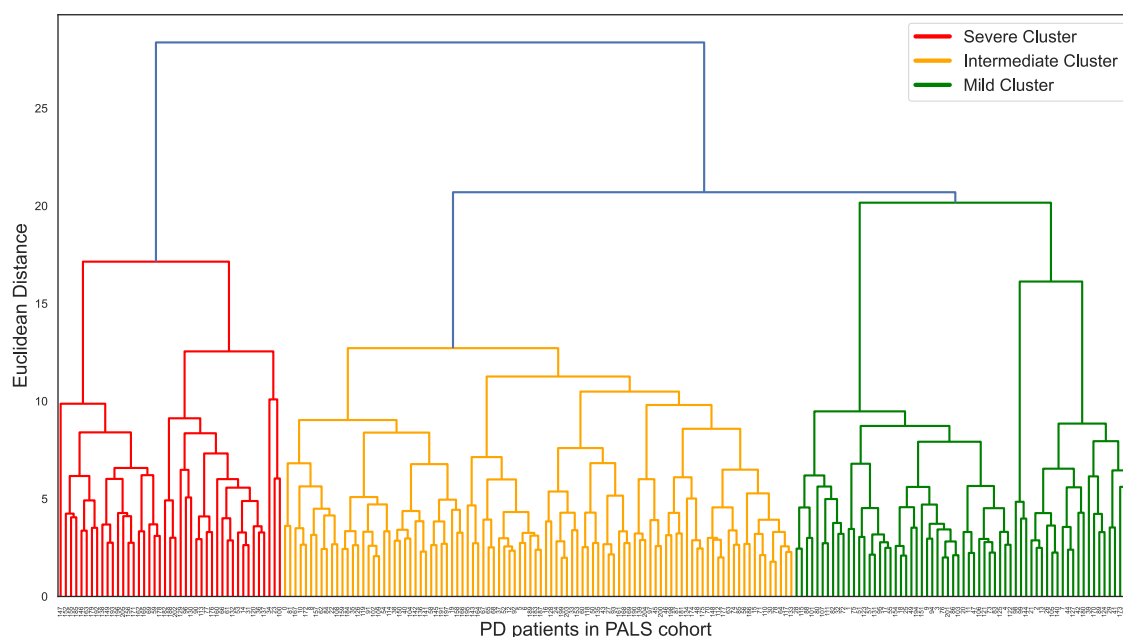
Cluster A consistently had significantly worse performance in all profiles, including highest MCI percentage (81%) and NMSS total score. Cluster B had obvious cognitive impairment with mild NMS and was characterized by having a moderate percentage of MCI (64%) and the lowest NMSS total score (11 vs 26 and 16 in cluster A and C,  $p < 0.001$ ,  $q < 0.0019$ ). Cluster C was a mild subtype and was characterized by having the lowest MCI percentage (15%, 64%, 81%, for cluster C, B, A respectively,  $p < 0.001$ ,  $q < 0.0019$ ).

### Characterization of PD subtypes using blood biomarkers

*Allelic distributions of Asian related PD genes in three PD clusters.* A total of 206 PD patients were genotyped. The Park16 rs6679073 A allele frequency was 76.7% (158 A allele carriers, including 77 patients carried AA and 81 patients harboured AC), the SV2C rs246814 T allele frequency was 15.0% (31 T allele carriers, including 2 patient carried TT and 29 patients carried TC).

The three clusters had significantly different effect allele frequency in these two SNPs (distribution of Park16 rs6679073 A allele carriers in cluster A B C: 67%, 74%, 89%,  $p = 0.015$ ,  $q = 0.065$ ; SV2C rs246814 T allele distribution: 7%, 12%, 25%,  $p = 0.026$ ,  $q = 0.065$ ; Table 4). Cluster A (severe subtype in motor, NMS and cognitive domains) had the lowest percentage of both Park16 and SV2C effect allele, while cluster C (mild subtype and young age of onset) contained the largest number of the carriers of these two SNPs.

*Comparison of blood biochemical markers among three clusters.* We found significant differences in Hcy and CRP levels among three clusters in the generalized linear model after adjustment for age and sex. Highest levels of Hcy and CRP were present in Cluster A (severe subtype in motor, NMS and cognitive domains), while lowest levels were shown in Cluster C (mild subtype and young age of onset). The differences of Hcy and CRP levels among three cluster were remained significant after adjustment for multiple comparisons. Mean levels of Hcy among three clusters were: 19.4 ± 4.2, 18.4 ± 5.7, 15.6 ± 5.6,  $p = 0.001$ ,  $q = 0.005$ ; while the mean levels of CRP were: 2.5 ± 5.0, 1.5 ± 2.4, 0.9 ± 2.1,  $p = 0.000$ ,  $q < 0.0001$  (Table 5). The comparison of Hcy and CRP levels among



**Fig. 1** Dendrogram of the final hierarchical cluster solution in the PALS cohort. Reference: <https://www.analyticsvidhya.com/blog/2019/05/beginners-guide-hierarchical-clustering/>.

three clusters remained significant after adjustment for age of diagnosis, sex and significant comorbidities including hypertension, hyperlipidemia, lipid medication and hypertension medication. Please refer to Supplementary Table 2.

## DISCUSSION

In this study, 206 early PD patients who were recruited within 1 year from diagnosis were assigned to three clusters by an unbiased data-driven hierarchical cluster analysis: cluster A (severe subtype in motor, NMS and cognitive domains), cluster B (intermediate subtype with cognitive impairment and mild NMS) and cluster C (mild subtype and young age of onset). Despite similar disease durations, the three clusters presented with substantially different clinical features and blood biomarker (genetic markers and biochemical markers) profiles. The significantly different allele frequencies in two SNPs (*Park16* rs6679073 A allele and *SV2C* rs246814 T allele), suggest that these may be important genetic biomarkers for PD subtypes. We also found Hcy and CRP to be promising biomarkers to identify the severe PD subtype. These findings contribute to our understanding of PD heterogeneity, especially among Asian PD.

Various PD subtypes have been identified through cluster analysis in previous studies. The diffuse malignant cluster previously reported by Fereshtehnejad in two different studies<sup>4,18</sup> is most akin to cluster A in our study. Cluster A was severe in all disease domains including motor, NMS and cognition. The underlying mechanism of this severe cluster most likely lies in simultaneous involvement of dopaminergic and non-dopaminergic pathways at an early disease stage<sup>18</sup>.

Previous studies reported that the most critical determinants of PD subtype were UPDRS, cognitive status, RBD, and orthostatic hypotension<sup>4,18</sup>. In our study, cluster A was best defined by MDS-UPDRS part I, II, III scores, cognitive impairment and significant BP drop, suggesting that the most critical drivers for PD subtype in our cohort are consistent with previous reports. However, RBD was not found to be an effective clinical determinant for PD subtyping in our cohort. Our study had a low RBD detection rate, which is likely attributable to the use of RBD1Q to detect RBD rather than use of gold standard overnight polysomnography assessment. In

addition, the MCI percentage in our cohort was higher than that in PPMI cohort reported by Weintraub<sup>19</sup>. Older age of diagnosis, lower education year and different ethnic population in our cohort may contribute to the difference. The identification of the severe cluster and its critical clinical drivers would enable clinicians to identify PD patients with a more severe subtype at an early disease stage.

Besides the severe cluster, there were two comparatively more benign PD clusters in our cohort. Cluster C was characterised by young onset with generally better performance in all disease domains. This finding is consistent with previous studies<sup>5,20</sup> that have identified a mild PD subtype with young onset.

Another comparatively benign PD cluster in our cohort was cluster B, comprising 47.6% of the PD patients, with the key features of cognitive impairment and mild NMS. Cluster B is a unique subtype in our cohort with the cognitive performance and NMS scores found to be in opposing directions from each other. The mechanism of cognitive impairment in PD is not fully understood. Acetylcholine neurotransmitter dysfunction is one of the possible pathways<sup>21</sup>. Muller et al reported that cognitive impairment alone in PD patients was related to isolated cortical cholinergic deficits, while a combination of cognitive decline, falls and RBD correlated with thalamic and cortical cholinergic deficiency<sup>22</sup>. The features of cluster B in our cohort suggest that the underlying affected brain areas of cognitive impairment in PD patients might be heterogeneous.

When characterizing genetic markers in the three PD subtypes, we found that they had significantly different allele frequencies in two SNPs even though the composite genetic score was not significantly different among three clusters. The mild cluster had significantly higher frequencies of the *Park16* rs6679073 A allele and *SV2C* rs246814 T allele, indicating that these two SNPs may have potential neuroprotective effects in our Asian cohort.

*Park16* SNPs has been consistently reported to play a protective role of PD development in different populations<sup>23,24</sup>. However, there is little information about the clinical characteristics of *Park16* carriers. Our study found that the number of *Park16* carriers was highest in the mild cluster (*Park16* rs6679073 A allele frequency in mild cluster was similar with

**Table 2.** Cluster features comparison.

Variables included for clustering	Cluster A Severe cluster <i>n</i> = 43	Cluster B Intermediate <i>n</i> = 98	Cluster C Mild cluster <i>n</i> = 65	<i>p</i> value*
Age of diagnosis (year)	69.6 ± 7.9	63.6 ± 7.4	59.4 ± 9.7	<0.001
PRS	0.7 (0.3–0.8)	0.7 (0.5–0.8)	0.7 (0.5–1.0)	0.100
Standardized Memory score	−1.8 (−2.6 to −0.9)	−1.3 (−2.0 to −0.7)	−0.1 (−0.6 to 0.3)	<0.001
Standardized Visuospatial score	−1.1 (−1.9 to 0.1)	−0.6 (−1.3 to −0.0)	−0.0 (−0.5 to 0.5)	<0.001
Standardized Attention score	−0.7 (−1.1 to −0.1)	−0.6 (−0.9 to 0.0)	0.3 (−0.4 to 0.9)	<0.001
Standardized Language score	−1.0 (−1.6 to −0.1)	−0.4 (−1.2 to 0.1)	0.3 (−0.0 to 0.8)	<0.001
Standardized Executive score	−0.7 (−1.5 to −0.2)	−0.4 (−0.9 to 0.2)	0.5 (0.1–1.0)	<0.001
MDS-UPDRS part I score	6 (3–8)	2.5 (1–4)	4 (2–7)	<0.001
MDS-UPDRS part II score	6 (4–9)	2 (1–4)	3 (1–6)	<0.001
MDS-UPDRS part III score	32 (26–37)	19.5 (15–24)	18 (14–20)	<0.001
PIGD score	3 (2–4)	1 (1–2)	1 (1–2)	<0.001
Tremor score	5 (3–8)	2 (0–4)	2 (0–3)	<0.001
Orthostatic SBP drop >10 mmHg (%)	15 (35%)	21 (21%)	7 (11%)	0.010
ESS Total Score	9 (5–12)	5 (2–8)	5 (3–7)	<0.001
HADS Anxiety score	2 (1–4)	1 (0–3)	2 (0–6)	0.055
HADS Depression score	4 (2–7)	2 (1–3)	2 (1–4)	<0.001
RBD1Q	0 (0–1)	0 (0–0)	0 (0–1)	0.520

All the variables were standardized before cluster analysis.

Categorical variables reported as frequency (%); continuous variables reported as mean ± standard deviation or median and first and third quartile (where appropriate)

\*Chi-square or Fisher exact test (where appropriate) for categorical variables, One-way ANOVA or Kruskal Wallis test for continuous variables (depends on normality assumption).

PRS polygenic risk score, MDS-UPDRS Movement Disorder Society-Unified Parkinson's Disease Rating Scale, PIGD postural instability and gait disorder, SBP systolic blood pressure, ESS Epworth sleepiness scale, HADS Hospital Anxiety Depression Scale, NMSS non-motor symptom scale, RBD1Q rapid eye movement sleep behaviour disorder single-question screen.

**Table 3.** Post hoc comparison of the baseline demography and clinical features among the three PD clusters.

Post-hoc comparison of other variables	Cluster A severe cluster <i>n</i> = 43	Cluster B intermediate <i>n</i> = 98	Cluster C mild cluster <i>n</i> = 65	<i>p</i> value*	<i>q</i> value**
Sex: male	29 (67%)	56 (57%)	37 (57%)	0.46	0.4600
MCI	35 (81%)	63 (64%)	10 (15%)	<0.001	<0.0019
MoCA Score	23 (19–26)	25 (22–27)	28 (26–29)	<0.001	<0.0019
NMSS Total Score	26 (14–43)	11 (6–18)	16 (9–32)	<0.001	<0.0019
NMSSD1Score (cardiovascular)	0 (0–4)	0 (0–0)	0 (0–0)	0.003	0.0050
NMSSD2Score (sleep/fatigue)	4 (0–8)	0 (0–3)	1 (0–5)	<0.001	<0.0019
NMSSD3Score (mood/apathy)	1 (0–5)	0 (0–0)	0 (0–2)	<0.001	<0.0019
NMSSD4Score (perceptual problems)	0 (0–0)	0 (0–0)	0 (0–0)	0.27	0.2893
NMSSD5Score (attention/memory)	1 (0–4)	0 (0–1)	0 (0–2)	0.027	0.0338
NMSSD6Score (gastrointestinal)	2 (0–7)	0 (0–2)	0 (0–3)	0.004	0.006
NMSSD7Score (urinary)	5 (4–12)	4 (0–5)	2 (0–5)	<0.001	<0.0019
NMSSD8Score (sexual function)	0 (0–2)	0 (0–0)	0 (0–0)	0.26	0.2893
NMSSD9Score (miscellaneous)	0 (0–3)	0 (0–3)	1 (0–4)	0.024	0.0327
LEDD	205.5 ± 103.5	211.3 ± 142.3	162.5 ± 147.8	0.089	0.1458

Categorical variables reported as frequency (%); continuous variables reported as mean ± standard deviation or median and first and third quartile (where appropriate)

\*Chi-square or Fisher exact test (where appropriate) for categorical variables, One-way ANOVA or Kruskal Wallis test for continuous variables (depends on normality assumption)

\*\*False discovery rate (FDR) method was performed and *q* values were calculated to control for multiple testing and the threshold of *q* values was set as 0.1. MCI mild cognitive impairment, H&Y Modified Hoehn and Yahr (H&Y) staging scale, NMSS non-motor symptom scale, LEDD levodopa equivalent daily dose.

**Table 4.** Allelic distributions of Asian related PD genes among the three PD clusters.

SNP	Effect allele	Cluster A severe cluster <i>n</i> = 43	Cluster B intermediate <i>n</i> = 98	Cluster C mild cluster <i>n</i> = 65	<i>p</i> value*	<i>q</i> value**
<i>Park16</i> rs6679073	A	29 (67%)	72 (74%)	57 (89%)	0.015	0.065
<i>SV2C</i> rs246814	T	3 (7%)	12 (12%)	16 (25%)	0.026	0.065
<i>SNCA</i> rs6826785	C	36 (86%)	77 (86%)	51 (82%)	0.87	0.880
<i>Lrrk2</i> rs141336855	T	1 (2%)	3 (3%)	3 (5%)	0.88	0.880
<i>ITPKB</i> rs16846351	G	6 (15%)	10 (11%)	5 (9%)	0.60	0.880

\*Fisher's exact test was carried out to compare the Gene allelic distributions among different clusters.

\*\*False discovery rate (FDR) method was performed and *q* values were calculated to control for multiple testing and the threshold of *q* values was set as 0.1.

**Table 5.** Comparison of blood biomarkers among the three PD clusters.

Blood biochemical markers	Cluster A severe cluster <i>n</i> = 43	Cluster B intermediate <i>n</i> = 98	Cluster C mild cluster <i>n</i> = 65	<i>p</i> value*	<i>q</i> value**
Hcy (u/molL)	19.4 ± 4.2	18.4 ± 5.7	15.6 ± 5.6	0.001	0.005
CRP (mg/L)	2.5 ± 5.0	1.5 ± 2.4	0.9 ± 2.1	0.000	<0.0001
Vit D3(ng/mL)	22.8 ± 7.5	23.2 ± 7.3	21.7 ± 6.7	0.522	0.522
UA(mg/dL)	5.3 ± 1.2	5.0 ± 1.4	4.8 ± 1.5	0.200	0.286
TC(mg/dL)	176.0 ± 37.8	195.1 ± 36.2	189.4 ± 31.5	0.045	0.113
TG(mg/dL)	103.5 ± 29.6	110.5 ± 53.6	93.3 ± 34.1	0.039	0.100
HDL(mg/dL)	56.9 ± 13.6	61.3 ± 15.5	61.4 ± 15.8	0.254	0.300
LDL(mg/dL)	98.6 ± 30.8	111.9 ± 31.7	109.4 ± 26.3	0.189	0.286
APOA1(g/L)	1.3 ± 0.2	1.3 ± 0.2	1.3 ± 0.2	0.270	0.300
APOB(g/L)	0.8 ± 0.2	0.9 ± 0.2	0.8 ± 0.2	0.141	0.282

Mean ± standard deviation

\*Generalized linear model was applied to compare the biomarkers against different clusters and adjusted for age of diagnosis, sex.

\*\*False discovery rate (FDR) method was performed and *q* values were calculated to control for multiple testing and the threshold of *q* values was set as 0.1.

Hcy: Homocysteine, CRP C-reactive protein, Vit D3 Vitamin D3, UA Uric acid; TC Cholesterol; TG Triglyceride; HDL-C high-density lipoprotein cholesterol, Apo A1 apolipoprotein A1, LDL-C low-density lipoprotein cholesterol, Apo B apolipoprotein B.

the healthy control group, results not shown), indicating that *Park16* carriers were likely to have mild symptoms, which supports the SNP's neuroprotective effects in PD. The underlying mechanism is not entirely clear. One possible explanation may lie in the interaction between *Park16* and *LRRK2*<sup>25</sup>. MacLeod et al. reported that deficiency of the *Park16* locus gene *RAB7L1*(*RAB29*) resulted in neurodegeneration in *LRRK2* mutant neurons while overexpression of *RAB7L1* restored the function of neurons with *LRRK2* mutation in an animal model<sup>25</sup>. Future clinical studies are needed to further elucidate the interaction between *Park16* and *LRRK2*.

Recently, Foo et al.<sup>9</sup> reported that synaptic vesicle glycoprotein 2C (*SV2C*) was a novel gene having robust association with PD development in various populations. It was also reported that *SV2C* was a functional PD candidate gene and an important mediator of dopamine homeostasis. Genetic deletion of *SV2C* caused a reduction of dopamine release, resulting in a decrease in motor activity<sup>26</sup>. Our findings corroborate the possible neuroprotective effects of *SV2C*, as the severe cluster had the lowest percentage of *SV2C*, while the mild cluster had the highest number of the *SV2C* carriers.

Our characterization of blood biochemical markers and PD clusters found Hcy to be a promising biomarker for the severe PD subtype after adjusting for confounders and multiple comparisons. Previous evidence have found elevated blood

Hcy levels to be associated with cognitive impairment in PD patients<sup>9,27</sup>. However, to our best knowledge, Hcy has not been previously reported to be associated with PD severity. The robust relationship between elevated Hcy levels and severe PD subtype may open new strategies for PD treatment. Since the accelerated rate of brain atrophy in the elderly with MCI have been found to be slowed by treatment with homocysteine-lowering B vitamins<sup>28</sup>, it is worth investigating whether PD severity can be ameliorated by adding vitamin supplementation to lower the Hcy levels. In addition to Hcy, we found CRP to be another reliable biomarker for the severe PD subtype, a finding that is in agreeable with a previous report<sup>8</sup>. These findings lend evidence for the existence of an heightened inflammatory state in severe PD subtypes.

Previous studies have reported that lipids have a neuroprotective effect on PD development. However, it is still controversial if lipid markers are associated with specific PD subtypes. We found that clusters with more than 60% MCI incidence (both clusters A and B) had significantly higher TG level, consistent with our previous finding that higher TG levels were related to cognitive impairment<sup>29</sup>. Our results also showed that the severe cluster (cluster A) had lowest TC levels. However, these associations were not significant after adjusting for multiple comparisons. Lawton et al recently reported that the severe motor disease phenotype, poor psychological well-being, and poor sleep subtype was

associated with reduced Apo A1 levels<sup>8</sup>. Our study, however, failed to reproduce this association. We were also unable to find out any significant correlation between our PD subtypes and Vitamin D or UA levels.

Our PD clusters were generated from an Asian cohort with all PD patients recruited within 1 year from diagnosis, which ensures that the cluster features were not driven by different disease durations and stages. In addition, we performed cluster analysis by including genetic status, that enabled us to investigate PD heterogeneity at the genetic level. Our study also tries to assess the association between a broad list of blood biomarkers (genetic markers and serum biochemical markers) and PD clusters, which provides comprehensive biological characterization for the newly generated clusters. However, some limitations of the study should be noted. The current study was cross-sectional and blood biomarkers were not monitored overtime. Longitudinal follow-up of these PD subtypes to monitor their biomarkers and disease progression will be needed. Furthermore, this was a single cohort study with limited sample size that requires further validation in other populations.

In summary, we introduced three subtypes of early PD patients in a multi-centre Asian cohort: 'severe', 'intermediate' and 'mild young-onset' subtypes. The severe subtype was associated with significantly lower frequency of Park16 and SV2C alleles; and had significantly higher levels of serum Hcy and CRP. Park16, SV2C, Hcy and CRP may be useful biomarkers to stratify PD patients into disease subtypes. Our findings also shed light on the possible underlying mechanisms that account for PD heterogeneity. This will improve the stratification of PD patients into disease subtypes that will enable more targeted personalised treatment strategies. Further validation of the genetic and biochemical differences between subtypes in larger cohorts and evaluation of their impact on PD progression is warranted.

## METHODS

### Participants and enrolment

**Study population.** A total of 206 idiopathic early PD patients defined by National Institute of Neurological Disorders and Stroke (NINDS) diagnostic criteria have been recruited from Early Parkinson's disease Longitudinal Singapore (PALS) cohort based on the inclusion and exclusion criteria of PALS study protocol<sup>30</sup>. PALS is an ongoing prospective cohort study undertaken to investigate the disease course of early PD patients who were recruited within 1 year of diagnosis.

**Enrolment.** Our study was conducted at two movement disorder outpatient clinics (Singapore General Hospital and Tan Tock Seng Hospital) in Singapore. Our study has been approved by SingHealth Centralized Institutional Review Board (CIRB) with Ref 2019/2433 and written informed consent was provided by all participants.

### Data collection

Comprehensive clinical features (motor, NMS and cognitive domains) and blood biomarkers were collected and used in the study. All clinical assessments were performed while patients were on their PD medications.

#### Clinical assessments

**Motor manifestations:** Movement Disorder Society-Unified Parkinson's Disease Rating Scale (MDS-UPDRS) Part II score (Motor Aspects of Experiences of Daily Living), part III motor score, tremor score, Postural Instability and Gait Disorder (PIGD) score<sup>31,32</sup> were used to assess motor performance. The calculation of tremor score was based on MDS-UPDRS items: 2.10, 3.15a, 3.15b, 3.16a, 3.16b, 3.17a, 3.17b, 3.17c, 3.17d, 3.17e, and 3.18, while the mean of MDS-UPDRS items 2.12, 2.13, 3.10, 3.11, and 3.12 was PIGD score<sup>32</sup>.

**NMS:** MDS-UPDRS Part I score (Non-Motor Aspects of Experiences of Daily Living) and Non-motor symptom scale (NMSS) total score were used to assess NMS burden; NMSS<sup>33</sup> consists of 30 items which are grouped into nine domains (cardiovascular domain, sleep/fatigue, mood/apathy, perceptual problems/hallucinations, attention/memory,

gastrointestinal, urinary, sexual function and miscellaneous). Olfaction impairment and autonomic failures are included in the miscellaneous domain. Rapid Eye Movement (REM) Sleep Behavior Disorder (RBD), Daytime sleepiness and sleep quality were evaluated by the RBD Single-Question Screen (RBD1Q)<sup>34</sup> and Epworth sleepiness scale (ESS) respectively. Patient depression and anxiety were assessed by Hospital Anxiety Depression Scale (HADS).

**Cognitive impairment:** Montreal Cognitive Assessment (MoCA) was performed to monitor the overall cognitive change. Comprehensive neuropsychological tests were performed and 5 cognitive domain scores (executive, visuospatial, memory, attention and working memory, language) were calculated by using the average of the standardized score of two neuropsychological tests from the same domain. Specifically, the following cognitive tests were administered to evaluate the cognitive status of the 5 domains: (1) Executive: Frontal Assessment Battery (FAB) total score and Fruit Fluency; (2) Visuospatial: Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) judgment of Line Orientation and Rey-Osterrieth Complex Figure Test (ROCF) copy total score; (3) Memory: Alzheimer Disease Assessment Scale (ADAS)-cog delayed recall score and ROCF delayed recall total score; (4) Attention and working memory: Digit Span Backward and Symbol Span total score; (5) Language: Boston Naming Test (BNT) total score and Wechsler Adult Intelligence Scale | Fourth Edition (WAIS-IV)-Similarities. Mild Cognitive Impairment (MCI) diagnosis was based on International Parkinson and Movement Disorder Society (MDS) level II criteria<sup>35</sup>, in which cognitive impairment should be present in at least two neuropsychological tests with 1.5 standard deviations (SDs) worse than norms as cut offs, either within a single cognitive domain or across different cognitive domains.

**Others:** Blood pressure was measured both in the supine position and after 3 min of standing. Orthostatic drop in Systolic Blood Pressure (SBP) greater than 10 mmHg was considered significant BP drop and viewed as an objective measure of autonomic dysfunction<sup>4</sup>. We also collected demographic data including sex, age of diagnosis, ethnicity.

**Blood biomarkers assessments.** We genotyped variants of SNCA, LRRK2, Park16, ITPKB, SV2C using Illumina Infinium Global Screening Array – 24 v2.0. The PRS was defined as the sum of the number of risk alleles per individual weighted by their effect size estimate corresponding to the logarithm of the odds ratio<sup>36</sup>. In the current study we calculated PRS by comprising 5 Asian GWAS SNPs (*SNCA*, *LRRK2*, *Park16*, *ITPKB*, *SV2C*) with the highest effect size and p level less than the genome wide significant association level ( $5 \times 10^{-8}$ ) from the latest Asian GWAS meta-analysis<sup>37</sup> to provide quantitative data of genetic burden individually. The SNPs data being used for PRS calculation can be found in supplementary data.

We tested 10 commercially available blood biomarkers. They are homocysteine (Hcy), C-reactive protein (CRP), vitamin D, uric acid (UA) and lipid markers including Triglyceride (TG), total cholesterol (TC), high-density lipoprotein cholesterol (HDL-C), low-density lipoprotein cholesterol (LDL-C) Apolipoprotein A1 (Apo A1) and Apolipoprotein B (Apo B). Blood biomarkers were measured using overnight fasting venous serum sample and were determined by enzymatic assay in a professional medical laboratory (Quest Laboratories Pte Ltd, Singapore).

### Statistical methods

**Cluster analysis.** Cluster analysis was performed in Python Software version 3 (<http://www.python.org>). Seventeen variables (Age of diagnosis, PRS, Number of patients having significant BP drop, MDS-UPDRS Part II score, MDS-UPDRS Part III score, tremor score, PIGD score, MDS-UPDRS Part I score, ESS Total Score, HADS Anxiety Total score, HADS Depression Total score, RBD1Q, Memory score, Visuospatial score, Attention score, Language score, Executive score) were selected by expert opinion and contemporary evidence<sup>18</sup>. All variable measurements were standardized by using the Z-scores for the cluster analysis. Agglomerative hierarchical clustering with Euclidean distance calculation was applied. We selected the three-cluster solution due to more balanced data distribution and better clinical interpretation. Missing value pattern was identified as missing by random. Hence, single imputation approach was used to impute 92 (2.6%) missing values in the baseline variables.

**Post-hoc comparisons of clinical characteristics and allelic distributions of related genes.** Post-hoc comparisons were performed in Stata software (Stata/SE 16.1 Stata Corp. 2019. Stata Statistical Software: Release 16.

College Station, TX: Stata Corp LLC) and SAS OnDemand for Academics (SAS Institute Inc. 2014. SAS® OnDemand for Academics: User's Guide. Cary, NC: SAS Institute Inc.). Continuous variables were summarized using mean with standard deviation (SD) or median with first and third quartile. Categorical variables were summarized by frequencies and percentages. Demographics, clinical characteristics not included in cluster analysis and allelic distributions of related PD genes were compared among clusters. Fisher's exact test or Pearson Chi square test (where appropriate) was carried out to compare the categorical variables among different clusters; while one-way ANOVA or Kruskal-Wallis tests (depends whether normality assumption was tenable) was performed to compare continuous variables among different clusters.

**Blood biochemical markers comparisons among clusters.** Blood biochemical markers comparisons were performed in SAS OnDemand for Academics (SAS Institute Inc. 2014. SAS® OnDemand for Academics: User's Guide. Cary, NC: SAS Institute Inc.). All blood biochemical markers except CRP were log-transformed to reduce the right-skewness. Generalized linear model was performed to compare the biomarkers among different clusters and adjusted for age of diagnosis, sex, using normal distribution assumption for the outcome variable. Gamma distribution was assumed for CRP due to the skewed distribution even after log-transformation. False discovery rate (FDR) method<sup>38</sup> was performed to control for multiple testing comparison and q value was calculated. We set the threshold of q values as 0.1.

## DATA AVAILABILITY

The data collected during this study are available from the corresponding author upon reasonable request from qualified individuals.

## CODE AVAILABILITY

No previously unreported custom computer code or algorithm was used to generate results in this study.

Received: 2 March 2022; Accepted: 5 August 2022;

Published online: 29 August 2022

## REFERENCES

- Jellinger, K. A. Neuropathology of sporadic Parkinson's disease: evaluation and changes of concepts. *Mov. Disord.* **27**, 8–30 (2012).
- Gjerloff, T. et al. Imaging acetylcholinesterase density in peripheral organs in Parkinson's disease with 11C-donepezil PET. *Brain* **138**, 653–663 (2015).
- Sieber, B. A. et al. Prioritized research recommendations from the National Institute of Neurological Disorders and Stroke Parkinson's Disease 2014 conference. *Ann. Neurol.* **76**, 469–472 (2014).
- Fereshtehnejad, S. M. et al. New clinical subtypes of Parkinson disease and their longitudinal progression: a prospective cohort comparison with other phenotypes. *JAMA Neurol.* **72**, 863–873 (2015).
- van Rooden, S. M. et al. The identification of Parkinson's disease subtypes using cluster analysis: a systematic review. *Mov. Disord.* **25**, 969–978 (2010).
- Delenclos, M., Jones, D. R., McLean, P. J. & Uitti, R. J. Biomarkers in Parkinson's disease: Advances and strategies. *Parkinsonism Relat. Disord.* **22**(Suppl.), S106–S110 (2016).
- Parnetti, L. et al. CSF and blood biomarkers for Parkinson's disease. *Lancet Neurol.* **18**, 573–586 (2019).
- Lawton, M. et al. Blood biomarkers with Parkinson's disease clusters and prodromosis: The Oxford discovery cohort. *Mov. Disord.* **35**, 279–287 (2020).
- O'Suilleabhain, P. E. et al. Elevated plasma homocysteine level in patients with Parkinson disease: motor, affective, and cognitive associations. *Arch. Neurol.* **61**, 865–868 (2004).
- de Lau, L. M., Koudstaal, P. J., Hofman, A. & Breteler, M. M. Serum uric acid levels and the risk of Parkinson disease. *Ann. Neurol.* **58**, 797–800 (2005).
- Simon, K. C. et al. Mendelian randomization of serum urate and parkinson disease progression. *Ann. Neurol.* **76**, 862–868 (2014).
- Ding, H. et al. Unrecognized vitamin D3 deficiency is common in Parkinson disease: Harvard Biomarker Study. *Neurology* **81**, 1531–1537 (2013).
- Yeshokumar, A. K., Saylor, D., Kornberg, M. D. & Mowry, E. M. Evidence for the Importance of Vitamin D Status in Neurologic Conditions. *Curr. Treat. Options Neurol.* **17**, 51 (2015).

- Yan, L. Y. et al. Association between carotid plaque and Parkinson's disease. *Ann. Transl. Med.* **7**, 94 (2019).
- Swanson, C. R. et al. Lower plasma apolipoprotein A1 levels are found in Parkinson's disease and associate with apolipoprotein A1 genotype. *Mov. Disord.* **30**, 805–812 (2015).
- Li, J. et al. Correlations between blood lipid, serum cystatin C, and homocysteine levels in patients with Parkinson's disease. *Psychogeriatrics* **20**, 180–188 (2020).
- Park, J. H. et al. Association of high-density lipoprotein cholesterol variability and the risk of developing Parkinson disease. *Neurology* **96**, e1391–e1401 (2021).
- Fereshtehnejad, S. M., Zeighami, Y., Dagher, A. & Postuma, R. B. Clinical criteria for subtyping Parkinson's disease: biomarkers and longitudinal progression. *Brain* **140**, 1959–1976 (2017).
- Weintraub, D. et al. Cognitive performance and neuropsychiatric symptoms in early, untreated Parkinson's disease. *Mov. Disord.* **30**, 919–927 (2015).
- Erro, R. et al. The heterogeneity of early Parkinson's disease: a cluster analysis on newly diagnosed untreated patients. *PLoS One* **8**, e70244 (2013).
- 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).
- Muller, M. L. et al. Clinical markers for identifying cholinergic deficits in Parkinson's disease. *Mov. Disord.* **30**, 269–273 (2015).
- Gopalai, A. A. et al. PARK16 is associated with PD in the Malaysian population. *Am. J. Med Genet B Neuropsychiatr. Genet* **171**, 839–847 (2016).
- He, T. et al. Association between PARK16 and Parkinson's disease: A meta-analysis. *Neurosci. Lett.* **657**, 179–188 (2017).
- MacLeod, D. A. et al. RAB7L1 interacts with LRRK2 to modify intraneuronal protein sorting and Parkinson's disease risk. *Neuron* **77**, 425–439 (2013).
- Dunn, A. R. et al. Synaptic vesicle glycoprotein 2C (SV2C) modulates dopamine release and is disrupted in Parkinson disease. *Proc. Natl. Acad. Sci. USA* **114**, E2253–E2262 (2017).
- de Jager, C. A., Oulhaj, A., Jacoby, R., Refsum, H. & Smith, A. D. Cognitive and clinical outcomes of homocysteine-lowering B-vitamin treatment in mild cognitive impairment: a randomized controlled trial. *Int. J. Geriatr. Psychiatry* **27**, 592–600 (2012).
- Smith, A. D. et al. Homocysteine-lowering by B vitamins slows the rate of accelerated brain atrophy in mild cognitive impairment: a randomized controlled trial. *PLoS One* **5**, e12244 (2010).
- Huang, X. et al. Higher serum triglyceride levels are associated with Parkinson's disease mild cognitive impairment. *Mov. Disord.* **33**, 1970–1971 (2018).
- Huang, X. et al. Serum uric acid level and its association with motor subtypes and non-motor symptoms in early Parkinson's disease: PALS study. *Parkinsonism Relat. Disord.* **55**, 50–54 (2018).
- Jankovic, J. et al. Variable expression of Parkinson's disease: a base-line analysis of the DATATOP cohort. The Parkinson Study Group. *Neurology* **40**, 1529–1534 (1990).
- Stebbins, G. T. et al. How to identify tremor dominant and postural instability/gait difficulty groups with the movement disorder society unified Parkinson's disease rating scale: comparison with the unified Parkinson's disease rating scale. *Mov. Disord.* **28**, 668–670 (2013).
- Chaudhuri, K. R. et al. The metric properties of a novel non-motor symptoms scale for Parkinson's disease: Results from an international pilot study. *Mov. Disord.* **22**, 1901–1911 (2007).
- Postuma, R. B. et al. A single-question screen for rapid eye movement sleep behavior disorder: a multicenter validation study. *Mov. Disord.* **27**, 913–916 (2012).
- 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 (2012).
- Nalls, M. A. et al. Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. *Nat. Genet* **46**, 989–993 (2014).
- Foo, J. N. et al. Identification of risk loci for Parkinson Disease in Asians and comparison of risk between Asians and Europeans: A Genome-Wide Association Study. *JAMA Neurol.* **77**, 746–754 (2020).
- Glickman, M. E., Rao, S. R. & Schultz, M. R. False discovery rate control is a recommended alternative to Bonferroni-type adjustments in health studies. *J. Clin. Epidemiol.* **67**, 850–857 (2014).

## ACKNOWLEDGEMENTS

We would like to thank all participants and their families for their support of the PALS study, and also neurologists from the National Neuroscience Institute, Singapore for referring their patients for the study. We also thank the National Medical Research Council (LCG Funding: MOH-OFLCG18May-0002 and Translational Clinical Research program in PD). The authors thank Niu Chenglin for Python coding support.

## AUTHOR CONTRIBUTIONS

All authors have contributed significantly to this study, have read the manuscript, attest to the validity. E.-K.T. and L.C.S.T.: study concept and design, revising the manuscript, study supervision, obtaining funding. X.D.: statistical analysis and interpretation of data, drafting/revising the manuscript. S.E.S., N.L., and J.C.A.: methodology support, statistical analysis, revising the manuscript. X.B.: interpretation of data and revising the manuscript. S.N.Y.E., N.C., J.T., X.Y.C., and D.H.: acquisition of data. Y.-L.L., Z.X., K.-Y.T., W.-L.A., A.N.: revising the manuscript.

## COMPETING INTERESTS

The authors declare no competing interests.

## ADDITIONAL INFORMATION

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41531-022-00375-y>.

**Correspondence** and requests for materials should be addressed to Louis C. S. Tan.

**Reprints and permission information** is available at <http://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 license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license 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 license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022