

<https://doi.org/10.1038/s41531-025-00888-2>

# Early detection of Parkinson's disease through multiplex blood and urine biomarkers prior to clinical diagnosis

Check for updates

Shuo Gao<sup>1,2,3,6</sup>, Zheng Wang<sup>1,2,3,4,6</sup>, Yuanfeng Huang<sup>1,2,3</sup>, Guang yang<sup>1</sup>, Yijing Wang<sup>1,2</sup>, Yan Yi<sup>2</sup>, Qiao Zhou<sup>1</sup>, Xingxing Jian<sup>1,2</sup>, Guihu Zhao<sup>1,2,3</sup>, Bin Li<sup>1,2,3</sup>, Linyong Xu<sup>5</sup>, Kun Xia<sup>1</sup>, Beisha Tang<sup>1,3</sup> & Jinchun Li<sup>1,2,3</sup> ✉

Blood and urine biomarkers are commonly used to diagnose and monitor chronic diseases. We initially screened 67 biomarkers, including 4 urine biomarkers and 63 blood biomarkers, and identified 13 blood biomarkers significantly associated with Parkinson's disease (PD). Among these, we discovered three novel markers demonstrating strong associations: phosphate ( $P = 1.81 \times 10^{-3}$ ), AST/ALT ratio ( $P = 8.53 \times 10^{-6}$ ), and immature reticulocyte fraction (IRF) ( $P = 3.49 \times 10^{-20}$ ). We also substantiated eight well-studied biomarkers and elucidated the roles of two previously ambiguous biomarkers. Our analyses confirmed IGF-1 ( $P = 7.46 \times 10^{-29}$ ) as a risk factor, and C-reactive protein (CRP) ( $P = 1.43 \times 10^{-3}$ ) as protective against PD. Genetic analysis highlighted that IRF, CRP, and IGF-1 share significant genetic loci with PD, notably at *MAPT*, *SETD1A*, *HLA-DRB1*, and *HLA-DQA1*. Furthermore, Mendelian randomization (MR) analysis suggested potential causal associations between IGF-1, CRP, and PD. We identified several blood biomarkers that may be associated with the risk of developing PD, providing valuable insights for further exploration of PD-related biomarkers.

Parkinson's disease (PD) is the second most common neurodegenerative disorder, affecting approximately 10 million people worldwide, with an incidence of about 3% among those over 80 years old<sup>1</sup>. Despite extensive research into the mechanisms underlying PD over the past decade, diagnosis still relies on neurological examinations, primarily identifying motor symptoms such as bradykinesia, resting tremor, and rigidity, along with a range of non-motor symptoms (NMS) including reduced olfaction, constipation, urinary dysfunction, and orthostatic hypotension. Even with the proper application of these criteria, the misdiagnosis rate remains as high as 20% due to clinical overlap with other PD syndromes<sup>2–5</sup>. Meanwhile, studies have shown differences in diagnostic accuracy between movement disorder specialists and general clinicians, with an accuracy of 92.5% for the MDS clinically probable PD criteria and 90.3% for general clinicians<sup>6,7</sup>. Research indicates that by the time most patients are diagnosed with PD, 50% to 70% of dopaminergic neurons in the substantia nigra have already degenerated<sup>8</sup>. Furthermore, prodromal symptoms of PD, such as rapid eye movement sleep behavior disorder (RBD), may be present years before the disease

manifests clinically<sup>9,10</sup>. Consequently, there is an increasing necessity to identify biomarkers that can detect the initiation of pathology or identify the prodromal stage of the disease. Improving early diagnosis of PD can enhance patient outcomes, however, the challenge of accurately identifying the disease in its early stages remains significant<sup>11</sup>.

Various biomarkers that facilitate the early diagnosis of disease include imaging, cerebrospinal fluid analysis, oxidative stress markers, neuroprotective factors, and inflammatory indicators. Previous research on biomarkers for diagnosing PD has often concentrated on bodily fluids, particularly cerebrospinal fluid and blood samples. However, there are currently significant shortcomings in this area<sup>8,12</sup>. Investigations into cerebrospinal fluid have primarily concentrated on  $\alpha$ -synuclein<sup>13,14</sup>. Misfolded  $\alpha$ -synuclein aggregates are a defining feature of PD and other synucleinopathies.

In patients with PD, cerebrospinal fluid levels of total  $\alpha$ Syn are significantly lower compared to healthy controls<sup>15,16</sup>. Recently,  $\alpha$ Syn seed amplification assays have demonstrated significant potential as a diagnostic

<sup>1</sup>National Clinical Research Center for Geriatric Disorders, Department of Geriatrics, Xiangya Hospital and Hunan Key Laboratory of Medical Genetics, School of Life Sciences, Central South University, Changsha, Hunan, China. <sup>2</sup>Bioinformatics Center, Xiangya Hospital & Furong Laboratory, Central South University, Changsha, 410008 Hunan, China. <sup>3</sup>Key Laboratory of Hunan Province in Neurodegenerative Disorders, Department of Neurology, Xiangya Hospital, Central South University, Changsha, Hunan, 410008, China. <sup>4</sup>Hunan Key Laboratory of Molecular Precision Medicine, Xiangya Hospital, Central South University, Changsha, Hunan, 410008, China. <sup>5</sup>Hunan Provincial Key Laboratory of Oral Health Research & Xiangya Stomatological Hospital & Xiangya School of Stomatology, Central South University, Changsha, China. <sup>6</sup>These authors contributed equally: Shuo Gao, Zheng Wang.

✉ e-mail: [lijinchun@csu.edu.cn](mailto:lijinchun@csu.edu.cn)



tool for PD<sup>17</sup>. Furthermore, Increased concentrations of neurofilament light chains in CSF and blood have been associated with diseases of the central and peripheral nervous systems involving axonal damage or degeneration. NfL is released into extracellular fluid following axonal injury, making it a potential biomarker for such neurodegenerative conditions. Studies by Brit Mollenhauer and colleagues in two independent longitudinal cohorts have shown that baseline serum NfL levels are significantly higher in individuals with Parkinson's disease compared to control groups<sup>18</sup>. Although these studies have identified numerous biomarkers relevant to PD diagnosis and monitoring, the invasive nature of lumbar puncture for CSF collection, along with the high cost primarily associated with CSF analysis, limits its widespread use in PD diagnosis.

However, biomarkers based on blood and urine, due to their non-invasive nature, low cost, and capability for frequent monitoring, can be widely applied in clinical settings. Notably, blood biomarkers have demonstrated superior sensitivity and specificity over cerebrospinal fluid in accurately detecting Alzheimer's disease<sup>19,20</sup>. For instance, the Neurofilament light chain has been adopted as a biomarker for relapsing multiple sclerosis<sup>21</sup>. In the AMORIS study conducted in Sweden, Fang et al. explored the relationship between lipid-related biomarkers and PD in roughly 600,000 participants. They found that higher levels of total cholesterol, LDL cholesterol, and triglycerides correlated with a decreased risk of PD<sup>22</sup>. In a related analysis, Solmaz Yazdani et al. investigated immune-related biomarkers in 812,073 participants, focusing on the hazard ratios for ALS and PD associated with markers such as leukocytes and immunoglobulin G. This comprehensive research, tracking changes in biomarkers up to two decades prior to diagnosis, bolstered these findings<sup>23</sup>. However, these findings were often limited to a small number of biomarkers and utilized small clinical cohorts. More importantly, it was unclear whether pre-diagnostic levels of blood and urine biomarkers in epidemiological cohorts could predict the future risk of PD. To tackle these challenges, we conducted a comprehensive analysis using a large longitudinal cohort from the UK Biobank to investigate the association between baseline levels of multiple blood and urine biomarkers and the risk of incident PD. Subsequently, we performed extensive genetic analyses to explore the potential causal relationships linking these biomarkers to PD risk.

## Results

### Demographics

The UK Biobank recruited a total of 502,357 participants. We excluded non-British White participants ( $N = 59,885$ ), individuals with prevalent PD at baseline ( $N = 294$ ), and those with other types of neurological disorders at baseline ( $N = 15,284$ ). To handle missing data, we performed multiple imputations with chained equations. After imputation, all blood and urine biomarkers were standardized. Participants with values exceeding five standard deviations were identified as outliers and removed ( $N = 34,614$ ). Ultimately, 392,280 participants were included in the analysis. Over a median follow-up period of 9.5 years (Interquartile Range (IQR), 8.65–10.35 years), a total of 3,084 incident PD cases were observed. These cases had an average follow-up period of 7.89 years and comprised 1,931 males and 1,153 females. The mean age at onset was 63.01 years. Baseline characteristics data are shown in Table 1.

### Association of Biomarkers with Incident Parkinson's Disease

Firstly, We used a Lasso-Cox model adjusted for age and sex with the penalty parameter lambda determined by tenfold cross-validation to select biomarkers of predictivity, we preliminarily identified 46 potential blood and urine biomarkers associated with the onset of PD, predominantly related to blood cell count indicators (See Fig. 1).

We progressively included stricter covariates in our models to more accurately estimate the relationship between specific biomarkers and the onset of PD, while controlling for potential confounding variables. The detailed results of the 3 models are presented in Table 2. In Model 1, we adjusted for sex, age, smoking consumption, alcohol consumption, physical activity levels, and education. These basic demographic and lifestyle factors

are common determinants of health and allowed us to assess the relationship between biomarkers and PD risk in a population with similar lifestyle habits, resulting in the identification of 22 significant biomarkers. In Model 2, we added covariates such as diabetes, hypertension, and stroke, which are critical health conditions closely linked to neurological health. By controlling for these diseases, we could more precisely analyze the impact of biomarkers on PD risk independently of these severe health conditions, eliminating potential interference. This adjustment led to no significant changes in the associations for most biomarkers, except for total protein, which was no longer significantly related. Finally, in Model 3, we conducted a multivariable Cox proportional hazards regression analysis that included all covariates and significantly associated biomarkers identified in Model 2. This comprehensive approach reduced the number of significant biomarkers to 13, highlighting those independently associated with the occurrence of PD events. This step helped minimize the influence of uncontrolled variables, allowing us to identify independent and highly correlated biomarkers potentially associated with PD onset, which may provide valuable insights for better predicting the occurrence of PD. This progression from identifying 22 significant biomarkers in Model 1 to 13 in Model 3 illustrates the importance of rigorous adjustment of covariates in isolating the most robust biomarkers for PD onset.

These biomarkers encompass categories such as enzymes, blood cell parameters, metabolites, and proteins. Among these, the AST/ALT ratio ( $HR = 1.14$ , 95% CI: 1.08–1.21,  $P = 8.53 \times 10^{-6}$ ), NLR ( $HR = 1.14$ , 95% CI: 1.08–1.19,  $P = 6.06 \times 10^{-7}$ ), glucose ( $HR = 1.13$ , 95% CI: 1.07–1.19,  $P = 5.02 \times 10^{-6}$ ), IGF-1 ( $HR = 1.25$ , 95% CI: 1.20–1.30,  $P = 7.46 \times 10^{-29}$ ), phosphate ( $HR = 1.08$ , 95% CI: 1.03–1.12,  $P = 1.81 \times 10^{-3}$ ), and total bilirubin ( $HR = 1.06$ , 95% CI: 1.02–1.11,  $P = 5.31 \times 10^{-3}$ ) were positively associated with PD risk. Conversely, platelet crit ( $HR = 0.76$ , 95% CI: 0.70–0.83,  $P = 5.97 \times 10^{-9}$ ), eosinophil count ( $HR = 0.92$ , 95% CI: 0.88–0.96,  $P = 4.79 \times 10^{-4}$ ), C-reactive protein (CRP) ( $HR = 0.89$ , 95% CI: 0.83–0.95,  $P = 1.43 \times 10^{-3}$ ), urate ( $HR = 0.85$ , 95% CI: 0.81–0.89,  $P = 1.13 \times 10^{-10}$ ), immature reticulocyte fraction (IRF) ( $HR = 0.75$ , 95% CI: 0.71–0.79,  $P = 3.49 \times 10^{-20}$ ), basophil count ( $HR = 1.31$ , 95% CI: 1.25–1.37,  $P = 2.98 \times 10^{-31}$ ), and platelet count ( $HR = 1.10$ , 95% CI: 1.01–1.20,  $P = 3.91 \times 10^{-2}$ ) were significantly associated with PD risk.

Among the biomarkers we studied, IRF, phosphate, and the AST/ALT ratio are newly identified concerning PD with minimal prior investigation. Established biomarkers like urate, total bilirubin, NLR, platelet crit, platelet count, eosinophil count, Glucose and basophil count have been examined primarily in experimental settings and are now being analyzed using large-scale longitudinal data from the UK Biobank for the first time<sup>24–28</sup>. Furthermore, IGF-1 and CRP, known for their roles in systemic inflammation, have demonstrated potential links with PD in clinical and epidemiological studies, though these findings have historically been inconsistent<sup>29–31</sup>. To address these discrepancies and deepen our understanding, our study has conducted extensive genetic analyses to further elucidate the relationships between these biomarkers and PD.

Finally, we conducted stratified analyses by sex and age. In males, NLR, glucose, phosphate, and total bilirubin were no longer significant, while in females, there were no differences (See Supplementary Table 1). In the age-stratified analysis, CRP and phosphate were no longer significant in individuals older than 65 years. After excluding individuals who developed diseases within three years from the baseline measurement, the results remained unchanged, confirming the reliability and stability of our findings. (See Supplementary Table 2).

### Genetic correlation analysis

To estimate the genetic correlation between 13 biomarkers and PD, we utilized LD Score Regression (LDSC) and Genome-wide Complex Trait Analysis (GNOVA), employing GWAS summary statistics. Our analysis covered 11 of these biomarkers, identifying significant genetic correlations with PD in six instances. Notably, both CRP and IGF-1 exhibited significant correlations in both LDSC and GNOVA analyses, with CRP showing a negative correlation (genetic correlation:  $-0.14$ ,  $P = 6.31 \times 10^{-5}$ ) and IGF-1 a

**Table 1 | Population characteristics included in the study**

Characteristics	Before imputation		After imputation		After exclusion	
	No PD ( <i>n</i> = 423,499)	Incident PD ( <i>n</i> = 3395)	No PD ( <i>n</i> = 423,499)	Incident PD ( <i>n</i> = 3395)	No PD ( <i>n</i> = 389,196)	Incident PD ( <i>n</i> = 3084)
Follow-up time (Years)	15.17 ± 0.90	15.23 ± 0.90	15.17 ± 0.90	15.23 ± 0.90	9.50 ± 0.85	7.89 ± 2.63
Age (Years)	56.77 ± 8.01	62.97 ± 5.18	56.77 ± 8.01	62.97 ± 5.18	56.74 ± 8.00	63.01 ± 5.13
Sex						
Male	193,324 (45.6%)	2132 (62.8%)	193,324 (45.6%)	2132 (62.8%)	175,913 (45.2%)	1931 (62.6%)
Female	230,175 (54.4%)	1263 (37.2%)	230,175 (54.4%)	1263 (37.2%)	213,283 (54.8%)	1153 (37.4%)
BMI (Kg/m <sup>2</sup> )	27.36 ± 4.73	27.63 ± 4.49	27.36 ± 4.73	27.63 ± 4.49	27.28 ± 4.64	27.49 ± 4.38
Townsend deprivation index	−1.54 ± 2.95	−1.61 ± 2.94	−1.54 ± 2.95	−1.61 ± 2.94	−1.58 ± 2.92	−1.66 ± 2.88
Physical activity (High)	64,398 (15.2%)	482 (14.2%)	64,398 (15.2%)	482 (14.2%)	59,762 (15.4%)	455 (14.8%)
Smoking status						
Never	230,774 (54.7%)	1,807 (53.5%)	231,500 (54.7%)	1,815 (53.5%)	214,296 (55.1%)	1,657 (53.7%)
Former	148,216 (35.1%)	1,367 (40.5%)	148,763 (35.1%)	1,374 (40.5%)	135,986 (34.9%)	1,243 (40.3%)
Current	43,050 (10.2%)	205 (6.1%)	432,36 (10.2%)	206 (6.1%)	38,914 (10.0%)	184 (6.0%)
Drinking status						
Never	131,15 (3.1%)	147 (4.3%)	13,133 (3.1%)	147 (4.3%)	11,772 (3.0%)	128 (4.2%)
Former	13,925 (3.3%)	190 (5.6%)	13,939 (3.3%)	191 (5.6%)	12,260 (3.2%)	166 (5.4%)
Current	396,112 (93.6%)	3,051 (90.1%)	396,427 (93.6%)	3,057 (90.0%)	365,164 (93.8%)	2,790 (90.5%)
Education (High)	68,890 (20.0%)	506 (20.5%)	85,616 (20.2%)	708 (20.9%)	78,771 (20.2%)	639 (20.7%)
High pressure (Yes)	167,101 (39.5%)	2,020 (59.5%)	167,101 (39.5%)	2,020 (59.5%)	150,156 (38.6%)	1,797 (58.3%)
Stroke (Yes)	12,826 (3.0%)	465 (13.7%)	12,827 (3.0%)	465 (13.7%)	11,162 (2.9%)	408 (13.2%)
Diabetes (Yes)	38,127 (9.0%)	568 (16.7%)	38,128 (9.0%)	568 (16.7%)	30,249 (7.8%)	458 (14.9%)

This table presents the characteristics of the study population, divided into three categories: before imputation, after imputation, and after exclusion. The characteristics analyzed include all covariates. Mean values (standard deviation) for continuous variables, and *n* (%) for categorical variables. Definition of abbreviations: *BMI* body mass index, *PD* Parkinson's disease.

positive correlation (genetic correlation: 0.12,  $P = 7.52 \times 10^{-6}$ ). This consistent significance across both methodologies underscores their robust association with PD (See Supplementary Table 3).

Additionally, the superior performance of GNOVA in single-trait analyses enabled the detection of four additional fluid biomarkers associated with PD that were not identified in the LDSC analysis. Specifically, eosinophil count (genetic correlation: -0.08,  $P = 3.34 \times 10^{-03}$ ), IRF (genetic correlation: -0.08,  $P = 3.45 \times 10^{-3}$ ), glucose (genetic correlation: -0.06,  $P = 3.51 \times 10^{-2}$ ), and total bilirubin (genetic correlation: 0.04,  $P = 3.18 \times 10^{-6}$ ) demonstrated notable correlations only in the GNOVA analysis.

These findings not only validate some of the previously identified biomarkers associated with PD risk but also confirm their genetic correlation with the disease.

### Estimation of pleiotropic enrichment

The stratified quantile-quantile (QQ) plots for PD conditioned on the *P*-values associated with IGF-1, CRP, and IRF reveal a clear pattern of successive enrichment, indicating that the proportion of non-null SNPs in PD increases with higher levels of association with these three biomarkers. This trend of successive enrichment is further supported from the opposite direction of analysis (see Supplementary Fig. 1). This observed enrichment in the QQ plots provides a compelling rationale to explore fold enrichment plots, which offer additional insights into the magnitude of these associations. Fold enrichment plots further quantify these relationships by assessing the degree of upward shift from the expected null line ( $P = 1$ ), thus providing a clearer measure of the strength of the genetic overlap and the extent of enrichment. In these plots, we observed significant levels of enrichment: CRP-conditioned PD enrichment exceeded 7-fold, IGF-1-conditioned enrichment was approximately 13-fold, and IRF-conditioned enrichment reached 4-fold. In the reverse direction, the enrichment levels

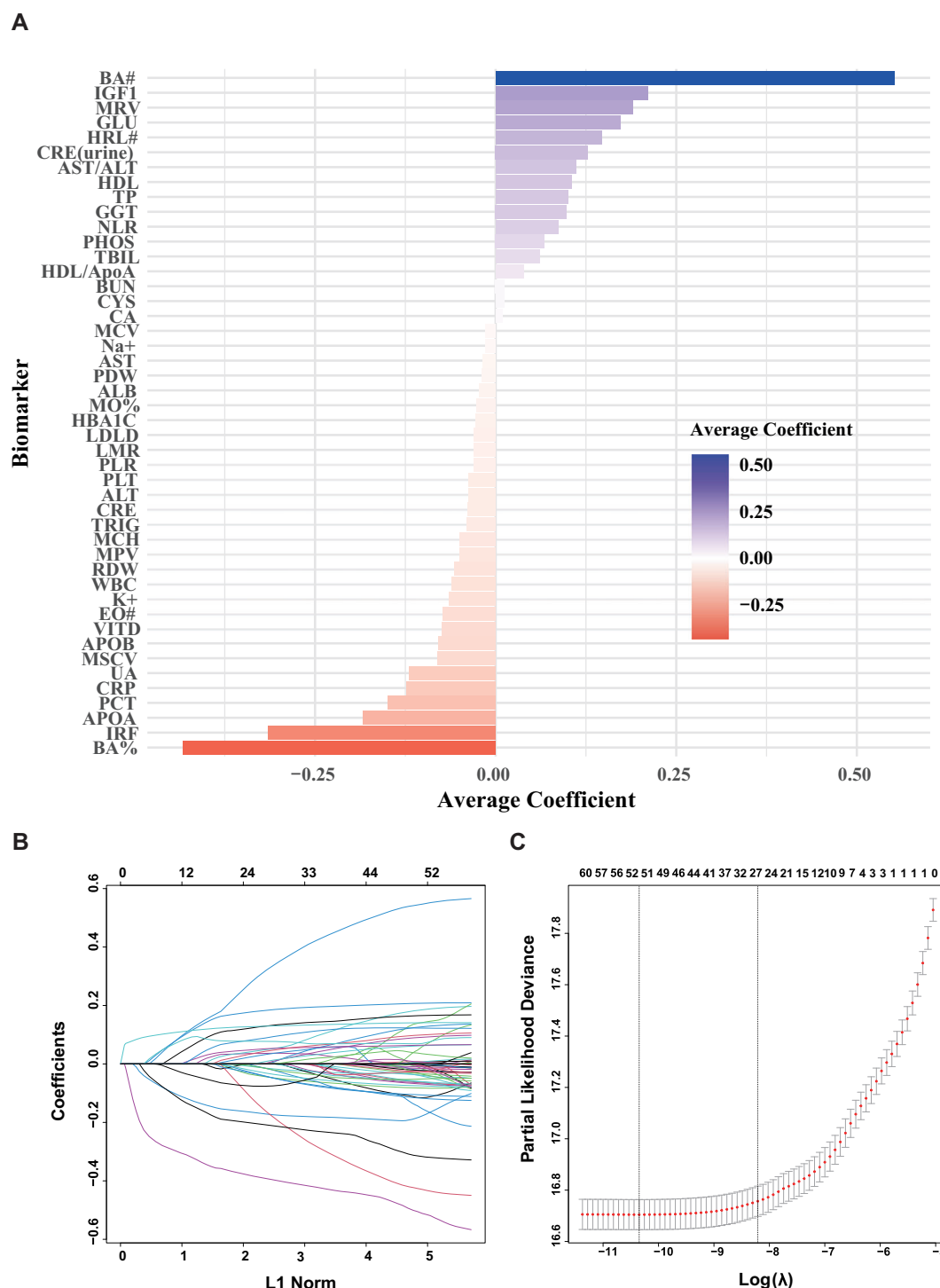
were even more pronounced: CRP-conditioned PD enrichment surpassed 20-fold, IGF-1-conditioned enrichment was around 4.5-fold, and IRF-conditioned enrichment approached 20-fold. These results not only confirm the trends indicated in the QQ plots but also highlight the significant pleiotropic effects, illustrating the substantial genetic interplay between these biomarkers and PD, thereby underscoring their potential roles in the pathogenesis and genetic architecture of PD (see Supplementary Fig. 2).

To discover genetic variants associated with biomarkers conditional on PD, we continued to perform conditional false discovery rate (conjFDR) analysis to identify pleiotropic variants between CRP and the risk of PD. We identified four pleiotropic loci with conjFDR values less than 0.01 (Table 3). The most significant locus was rs17573509 (*MAPT*). The *MAPT* H1 haplotype is widely recognized as a risk factor for various neurodegenerative diseases, such as PD, progressive supranuclear palsy, and corticobasal degeneration. Other significant loci include *HLA-DRB1* and *HLA-DQA1* located in the major histocompatibility complex (HLA), as well as *SETD1A*, which is crucial in hematopoiesis and development.

We leveraged the pleiotropic effects associated with various traits using a genetic pleiotropic condFDR method, employing GWAS summary data to investigate the genetic overlap among polygenic traits. This approach deepens our understanding of genetic pleiotropy and enhances statistical power, thereby elucidating the genetic mechanisms associated with multiple phenotypes. Through this method, we systematically assessed the shared genetic background between biomarkers and PD.

### Mendelian randomization analysis

In our study, we aimed to conduct MR analyses on three biomarkers: IRF, IGF-1, and CRP, all of which have been previously identified as having pleiotropic loci associated with PD. This approach helps minimize confounding biases and eliminate potential reverse causality, enabling us to assess the impact of these biomarkers on the risk of developing PD. By using



**Fig. 1 | Biomarker selection by LASSO regression model and biomarker's importance.** This figure illustrates the importance and direction of biomarkers selected through Lasso regression analysis. Panel A shows the average coefficients for each biomarker, with the color and length of the bars representing the magnitude and direction of their impact on the predictive outcome. Panel B illustrates the changes in coefficients of biomarkers during the LASSO regression process as a function of the L1 norm. Each colored line represents the coefficient trajectory of a different biomarker across varying levels of L1 norm regularization. The plot visualizes how coefficients are shrunk towards zero, highlighting which biomarkers remain significant as the regularization increases. Panel C illustrates the LASSO

coefficient profiles for the 35 variables as the log-transformed regularization parameter  $\lambda$  varies. The vertical axis shows the partial likelihood deviance, indicating the model's goodness of fit, while the horizontal axis displays the  $\log(\lambda)$  values. A vertical line within the graph marks the  $\lambda$  value selected by 10-fold cross-validation, optimized for balancing model complexity and predictive accuracy. As  $\lambda$  decreases, model compression intensifies, which enhances the model's ability to discern and retain only the most important variables. The sequence of numbers at the top indicates the count of variables retained in the model at each specific  $\lambda$ , helping to visualize how variable selection changes with increasing regularization.

**Table 2 | Associations of biomarkers and the risk of Parkinson's disease**

Biomarkers	Model I		Model II		Model III	
	HR (95% CI)	P-value	HR (95% CI)	P-value	HR (95% CI)	P-value
AST/ALT	1.18 (1.13, 1.23)	<b><math>3.34 \times 10^{-12}</math></b>	1.19 (1.13, 1.24)	<b><math>6.24 \times 10^{-13}</math></b>	1.14 (1.08, 1.21)	<b><math>8.53 \times 10^{-06}</math></b>
HDL/ApoA	1.10 (1.05, 1.15)	<b><math>1.18 \times 10^{-04}</math></b>	1.12 (1.07, 1.17)	<b><math>1.02 \times 10^{-05}</math></b>	1.06 (1.00, 1.12)	$6.74 \times 10^{-02}$
NLR	1.19 (1.13, 1.24)	<b><math>2.10 \times 10^{-12}</math></b>	1.16 (1.11, 1.22)	<b><math>3.93 \times 10^{-10}</math></b>	1.14 (1.08, 1.19)	<b><math>6.06 \times 10^{-07}</math></b>
Platelet count	0.86 (0.82, 0.90)	<b><math>2.88 \times 10^{-09}</math></b>	0.86 (0.82, 0.90)	<b><math>9.69 \times 10^{-10}</math></b>	1.10 (1.01, 1.20)	<b><math>3.91 \times 10^{-02}</math></b>
Platelet crit	0.83 (0.79, 0.87)	<b><math>1.28 \times 10^{-12}</math></b>	0.83 (0.79, 0.87)	<b><math>1.38 \times 10^{-13}</math></b>	0.76 (0.70, 0.83)	<b><math>5.97 \times 10^{-09}</math></b>
Eosinophil count	0.88 (0.85, 0.93)	<b><math>3.82 \times 10^{-07}</math></b>	0.88 (0.84, 0.92)	<b><math>5.52 \times 10^{-08}</math></b>	0.92 (0.88, 0.96)	<b><math>4.79 \times 10^{-04}</math></b>
Basophil count	1.24 (1.19, 1.30)	<b><math>5.41 \times 10^{-22}</math></b>	1.24 (1.19, 1.30)	<b><math>5.51 \times 10^{-22}</math></b>	1.31 (1.25, 1.37)	<b><math>2.98 \times 10^{-31}</math></b>
Monocyte percentage	0.92 (0.88, 0.96)	<b><math>9.49 \times 10^{-04}</math></b>	0.91 (0.87, 0.96)	<b><math>2.11 \times 10^{-04}</math></b>	0.95 (0.91, 1.00)	$5.94 \times 10^{-02}$
Immature reticulocyte fraction	0.75 (0.72, 0.78)	<b><math>4.01 \times 10^{-37}</math></b>	0.74 (0.71, 0.77)	<b><math>1.12 \times 10^{-40}</math></b>	0.75 (0.71, 0.79)	<b><math>3.49 \times 10^{-20}</math></b>
High light scatter reticulocyte count	0.84 (0.80, 0.88)	<b><math>6.06 \times 10^{-12}</math></b>	0.81 (0.78, 0.85)	<b><math>9.01 \times 10^{-16}</math></b>	1.07 (1.00, 1.14)	$6.78 \times 10^{-02}$
Sodium in urine	0.94 (0.90, 0.98)	<b><math>3.84 \times 10^{-03}</math></b>	0.94 (0.91, 0.98)	<b><math>4.75 \times 10^{-03}</math></b>	0.99 (0.95, 1.03)	$6.41 \times 10^{-01}$
Alanine aminotransferase	0.83 (0.78, 0.89)	<b><math>1.76 \times 10^{-08}</math></b>	0.82 (0.77, 0.87)	<b><math>3.93 \times 10^{-10}</math></b>	0.98 (0.91, 1.05)	$5.70 \times 10^{-01}$
Apolipoprotein B	0.88 (0.84, 0.92)	<b><math>3.20 \times 10^{-09}</math></b>	0.92 (0.89, 0.96)	<b><math>1.46 \times 10^{-04}</math></b>	1.06 (0.96, 1.17)	$3.01 \times 10^{-01}$
C-reactive protein	0.83 (0.77, 0.89)	<b><math>3.52 \times 10^{-07}</math></b>	0.82 (0.77, 0.88)	<b><math>8.33 \times 10^{-08}</math></b>	0.89 (0.83, 0.95)	<b><math>1.43 \times 10^{-03}</math></b>
Glucose	1.16 (1.11, 1.21)	<b><math>2.54 \times 10^{-09}</math></b>	1.12 (1.06, 1.17)	<b><math>1.64 \times 10^{-05}</math></b>	1.13 (1.07, 1.19)	<b><math>5.02 \times 10^{-06}</math></b>
IGF-1	1.28 (1.23, 1.33)	<b><math>4.17 \times 10^{-36}</math></b>	1.28 (1.23, 1.33)	<b><math>2.63 \times 10^{-36}</math></b>	1.25 (1.20, 1.30)	<b><math>7.46 \times 10^{-29}</math></b>
LDL direct	0.88 (0.84, 0.91)	<b><math>1.49 \times 10^{-08}</math></b>	0.92 (0.88, 0.96)	<b><math>1.78 \times 10^{-04}</math></b>	0.91 (0.82, 1.01)	$9.95 \times 10^{-02}$
Phosphate	1.06 (1.02, 1.11)	<b><math>6.62 \times 10^{-03}</math></b>	1.06 (1.02, 1.10)	<b><math>7.67 \times 10^{-03}</math></b>	1.08 (1.03, 1.12)	<b><math>1.81 \times 10^{-03}</math></b>
Total bilirubin	1.10 (1.06, 1.15)	<b><math>1.68 \times 10^{-06}</math></b>	1.10 (1.06, 1.14)	<b><math>4.08 \times 10^{-06}</math></b>	1.06 (1.02, 1.11)	<b><math>5.31 \times 10^{-03}</math></b>
Total protein	1.05 (1.01, 1.09)	<b><math>4.94 \times 10^{-02}</math></b>	1.04 (1.00, 1.08)	$8.84 \times 10^{-02}$	-	-
Triglycerides	0.86 (0.82, 0.90)	<b><math>1.15 \times 10^{-08}</math></b>	0.86 (0.82, 0.90)	<b><math>4.73 \times 10^{-09}</math></b>	0.85 (0.81, 0.89)	$2.29 \times 10^{-01}$
Urate	0.82 (0.78, 0.86)	<b><math>2.53 \times 10^{-15}</math></b>	0.81 (0.77, 0.85)	<b><math>8.63 \times 10^{-19}</math></b>	0.85 (0.81, 0.89)	<b><math>1.13 \times 10^{-10}</math></b>

HR hazard ratio, CI confidence interval, ns, not significance, AST aspartate transaminase, ALT alanine transaminase, HDL High-density lipoprotein, ApoA Apolipoprotein A, NLR neutrophil-to-lymphocyte ratio. Model I: Adjusted for age, sex, sociodemographic factors, and lifestyle domains. Model II: Adjusted for variables in Model I plus cardiovascular diseases. Model III: Includes only the statistically significant biomarkers ( $P < 0.05$ ) in Model II. These significant factors were further analyzed in a multivariate survival analysis.

**Table 3 | Shared risk loci between biomarkers and Parkinson's disease**

SNP	Genomic position	Closest gene	FDR-value	P-value in PD GWAS	P-value in Biomarker GWAS
<b>Immature reticulocyte fraction</b>					
rs1870293	16:30970941	SETD1A	<b><math>3.09 \times 10^{-6}</math></b>	<b><math>1.13 \times 10^{-9}</math></b>	<b><math>8.20 \times 10^{-9}</math></b>
rs17573509	17:44081224	MAPT	<b><math>2.67 \times 10^{-6}</math></b>	<b><math>1.58 \times 10^{-12}</math></b>	<b><math>6.87 \times 10^{-9}</math></b>
<b>IGF-1</b>					
rs7760841	6:32574868	HLA-DRB1; HLA-DQA1	<b><math>2.66 \times 10^{-4}</math></b>	<b><math>1.41 \times 10^{-6}</math></b>	<b><math>6.31 \times 10^{-18}</math></b>
<b>C-reactive protein</b>					
rs7760841	6:32574868	HLA-DRB1; HLA-DQA1	<b><math>1.27 \times 10^{-4}</math></b>	<b><math>1.41 \times 10^{-6}</math></b>	<b><math>5.44 \times 10^{-8}</math></b>

SNP, single nucleotide polymorphism; FDR, false discovery rate. The genomic position was based on GRCh37.

genetic variants as instrumental variables, we can explore potential causal relationships between these biomarkers and PD with greater precision.

We analyzed the causal effects of CRP and IGF-1 on PD risk using MR analyses. For CRP, 179 independent SNPs served as instrumental variables, showing robust instrument strength (median F-statistic of 65.7) and no evidence of weak instrument bias or horizontal pleiotropy (MR-Egger intercept  $P = 0.78$ ). The inverse variance weighted (IVW) method indicated a negative causal relationship between CRP levels and PD risk, consistent across leave-one-out analysis. In contrast, 289 SNPs were used for IGF-1, also demonstrating strong instrument validity (median F-statistic of 55.3) and no pleiotropic effects (MR-Egger intercept  $P = 0.74$ ). The IVW results suggested a positive association between IGF-1 levels and PD risk. For IRF, no significant causal association was found in the analysis.

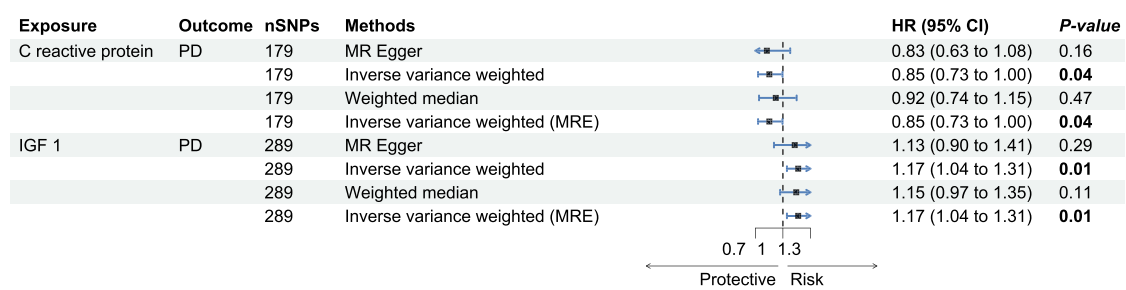
The MR analyses underscore the causal relationships between specific inflammatory biomarkers and PD. Specifically, our findings indicate that

CRP may have a protective role against PD, suggesting a possible anti-inflammatory mechanism that mitigates disease progression. Conversely, the positive association between IGF-1 levels and increased PD risk highlights the complex role of inflammatory processes in neurodegeneration (See Fig. 2).

## Discussion

This study leveraged the UK Biobank's extensive dataset of blood and urine biomarkers to advance our understanding of PD through case-control studies, genetic analysis, and MR. We identified 13 biomarkers associated with PD, including newly explored markers such as the AST/ALT ratio, IRF, and phosphates, alongside IGF-1 and CRP, which had previously shown inconsistent results in research. Genetic correlation analysis revealed significant inverse genome-wide correlations between PD and several biomarkers like eosinophil count, IRF, CRP, glucose, and total bilirubin, while showing a positive correlation with IGF-1. Among these, the relationships





**Fig. 2 | Mendelian Randomization Analysis of CRP and IGF-1 Associations.** This figure displays MR results for the association between CRP and IGF-1 with PD, focusing on the IVW method. We primarily examine the *p*-values and hazard ratio ranges from the IVW results.

with IGF-1 and CRP were the most pronounced. Our pleiotropy analysis confirmed four significant shared genetic loci linked to PD, including three previously reported (*HLA-DRB1*, *HLA-DQA1*, *MAPT*) and one less commonly reported locus, *SETD1A*. Ultimately, MR studies supported the nature of these associations, indicating potential causal relationships between the immunoinflammatory biomarkers IGF-1 and CRP and PD. These MR findings help resolve inconsistencies reported in earlier studies and provide robust support for these biomarkers' roles in PD.

Our study identified several biomarkers, including the AST/ALT ratio, which has not been previously confirmed as a significant indicator. Commonly associated with liver damage, the AST/ALT ratio is also linked to various metabolic syndromes. Moreover, research has shown that this ratio has a negative correlation with grip strength and walking speed, both of which are relevant to PD<sup>32,33</sup>. This suggests that the AST/ALT ratio may serve as a valuable biomarker in understanding the broader physiological impacts associated with PD. Phosphates, typically associated with the musculoskeletal system, can cause metabolic dysregulation, leading to an accumulation of excessive inorganic phosphates. Excessive phosphate levels are thought to impair the function of Complex I in the mitochondrial electron transport chain, a key component in PD pathology<sup>34,35</sup>. This impairment is associated with the activation of the mitochondrial permeability transition pore, which leads to increased production of reactive oxygen species, heightened inflammation, DNA damage, and depolarization of the mitochondrial membrane<sup>36</sup>. Consequently, additional research is essential to fully understand the potential impact of phosphate dysregulation in PD. The IRF reflects the erythropoietic activity of the bone marrow and is commonly used in the laboratory diagnosis of anemia<sup>37</sup>. A follow-up study from the Korean National Screening Program found that anemia was associated with a lower risk of PD, particularly in patients with moderate to severe anemia<sup>38</sup>. Additionally, some studies have shown that newly diagnosed anemia increases the risk of PD<sup>39</sup>. Therefore, further research is needed to explore the complex relationship between IRF and PD.

To further explore whether there are pleiotropic risk variants shared between biomarkers and PD, we systematically assessed the shared genetic background between biomarkers and PD. We subsequently utilized summary statistics from genome-wide association studies to analyze the genetic correlations between statistically significant biomarkers and PD. Using the condFDR method, we analyzed their genetic enrichment and successfully identified four common genetic loci. Among these loci, the *MAPT* identified between IRF and PD encodes the microtubule-associated protein tau. Research indicates that tau protein may influence alpha-synuclein and could be a key target for treating diseases related to tau and alpha-synuclein. Dysfunctions in *SETD1A* could lead to schizophrenia, and its reduced function may disrupt the normal development of cortical axons, dendrites, and spines, leading to cognitive deficits. Re-expression of *SETD1A* in adulthood may help reverse these cognitive impairments<sup>40,41</sup>. The genes *HLA-DRB1* and *HLA-DQA1*, which are shared among IGF-1, CRP, and PD, are implicated in immune regulation. These findings suggest a significant role of these genetic regions in modulating immune responses, which may influence the pathogenesis of PD. PD and Alzheimer's disease (AD) show genome-wide associations within the HLA class II region, which includes

the closely linked *HLA-DR* and *HLA-DQ* genes<sup>42</sup>. The *HLA-DRB1\*04* subtype can selectively bind the PHF6 sequence of tau protein, influencing the progression of neurodegenerative diseases<sup>43</sup>.

The significance of IGF-1 and CRP in PD research has been unclear, prompting us to conduct MR analyses to resolve these uncertainties. We explored the causal relationships between these genetically correlated biomarkers and PD, aiming to provide definitive insights into their roles in the disease's pathogenesis. In previous studies, IGF-1 is known for its neuro-protective functions, playing a crucial role in the development of the nervous system and the survival and differentiation of neurons<sup>44</sup>. Research indicates that IGF-1 levels decrease with age, correlating with PD, which predominantly affects the elderly<sup>45</sup>. However, several cohort studies suggest that PD patients exhibit higher levels of IGF-1 compared to healthy controls, potentially as a compensatory mechanism in response to the disease<sup>46</sup>. Our findings corroborate these cohort studies, indicating that elevated levels of IGF-1 are associated with an increased risk of PD. CRP is a biomarker that reflects the body's inflammatory state, with plasma CRP levels capable of rising rapidly and significantly in response to acute inflammatory stimuli<sup>47</sup>. It has been reported to be associated with the risk and progression of PD<sup>48</sup>. A previous large meta-analysis indicated that peripheral CRP levels were elevated in PD patients, suggesting that inflammation may play a role in the pathogenesis and symptoms of PD<sup>49</sup>. Additionally, a study based on the Oxford Discovery prospective cohort found that CRP might serve as a blood biomarker for predicting PD prognosis<sup>50</sup>. However, another study that included 12,642 participants found no significant association between CRP and the risk of developing PD<sup>51</sup>. These studies are often limited by sample size, highlighting the need for further research to explore the relationship between CRP and PD.

Our findings indicate that elevated levels of IGF-1 are associated with an increased risk of PD, while higher baseline levels of CRP are linked to a reduced risk of PD. This study represents the first demonstration of a causal relationship between IGF-1 and PD using MR analysis. These findings enhance our understanding of the association between these two biomarkers and PD risk, providing partial evidence for further exploration of their underlying pathogenesis.

The main strength of this study is the use of diverse pre-diagnostic blood and urine biomarkers to predict the risk of PD, addressing the limitations of small sample sizes and potential confounding factors in previous studies. Additionally, we utilized the large sample size from the UK Biobank, which enhances the statistical power and robustness of our findings, making the conclusions more reliable and generalizable. In our sensitivity analysis, we strategically excluded patients diagnosed with PD within three years prior to the study to minimize potential reverse causation—a common challenge in longitudinal studies. Furthermore, we employed a robust series of genetic validations that not only effectively addressed reverse causation but also mitigated the impact of residual confounding factors in epidemiological studies. These measures have significantly strengthened the validity of our results, solidly corroborating our epidemiological findings and enhancing our understanding of PD's pathogenesis. Additionally, our integration of genetic analysis with MR provided a novel perspective on the causal relationships between identified biomarkers and PD, confirming

findings that may have otherwise been obscured by biases. This study focused primarily on white British participants, which minimizes potential confounding from ethnic differences in plasma concentrations. However, this limitation reduces the generalizability of the findings to other demographic groups. Notably, white British individuals make up approximately 94.1% of the UK Biobank cohort, providing a sufficiently large sample size to ensure robust statistical power and valid conclusions. Despite this, future research should include more diverse populations to better understand the role of blood and urine biomarkers in PD risk. Using ICD codes to define PD might lead to misdiagnosis or misclassification issues, and as a result latent incident cases of the disease associated with less severe clinical symptoms might not have been captured, further in-depth studies are needed to validate these findings. Furthermore, biomarkers derived from body fluids, notably blood and urine, are vulnerable to fluctuations caused by physiological conditions and pharmacological interventions. Finally, given that our study was observational and the UK Biobank does not provide detailed information on medication use or certain pre-diagnostic characteristics, we recognize that, although MR was used to validate causal associations, the interpretation of the relationship between biomarkers and Parkinson's disease should still be approached with caution.

## Methods

### Study Population Data Availability

The UK Biobank, a community-based longitudinal cohort, includes over half a million individuals aged 40–69 years, recruited between 2006 and 2010. Participants attended one of 22 assessment centers across England, Scotland, or Wales, where they completed questionnaires and participated in face-to-face interviews. Trained staff conducted phenotypic measurements and collected biosamples. The process began with self-administered questionnaires on touch screens, followed by computer-assisted personal interviews when necessary. The UK Biobank questionnaire, developed with input from international experts, covered key topics including socio-demographic information, lifestyle factors (such as smoking and alcohol use), early life exposures, psychological state, cognitive function, family and medical history, and overall health<sup>52</sup>. This rich resource enables us to conduct detailed and extensive research. Utilizing these detailed biomarker profiles, our study seeks to deepen the understanding of the relationship between these biomarkers and the onset of PD, enhancing our capabilities for early detection and better understanding of the disease's mechanisms<sup>53</sup>.

### Biomarkers Measurement

The UK Biobank collected blood and urine samples during baseline recruitment, measuring pertinent blood biochemistry, blood count, and urine-related biomarkers. The specific Field IDs for the UK Biobank can be found in Supplementary Table 4. For detailed information on missing values, refer to Supplementary Fig. 3. Detailed information regarding the experimental procedures and quality control protocols can be found on the UK Biobank website.

Previous studies have indicated that various lipid trait ratios and immune-related markers are associated with dementia and cognitive function<sup>51,54–56</sup>. We calculated various ratios, including AST/ALT, HDL/ApoA, LDL/ApoB, and ApoB/ApoA ratios, NLR (Neutrophil-to-lymphocyte ratio), PLR (platelet-lymphocyte ratio), LMR (Lymphocyte-to-monocyte ratio), SII (systemic immune inflammation index). Before division, all components of the ratios were converted into the same units. Subsequently, standardization transformations were applied to all fluid biomarkers. Samples deviating more than 5 standard deviations from the mean were considered outliers and were excluded from the analysis involving relevant participants.

### Parkinson's Disease Outcome Definitions

PD diagnoses and onset dates are documented within hospital admission records across the UK, encompassing diagnostic data derived from England's Hospital Episode Statistics, Scotland's Morbidity Records, and Wales's Patient Episode Database, all classified using the 10th Revision of

the International Classification of Diseases (ICD-10)<sup>52</sup>. The G20 code encompasses PD, including conditions such as hemiparkinsonism, paralysis agitans, and parkinsonism that is not otherwise specified (NOS), idiopathic, or primary. Participants were considered at risk for PD from the date of enrollment, with follow-up continuing until either the initial diagnosis or September 2023, whichever came first. To minimize the potential for reverse causality, individuals with prevalent PD were excluded from the study prior to enrollment.

### Analytical Cohort

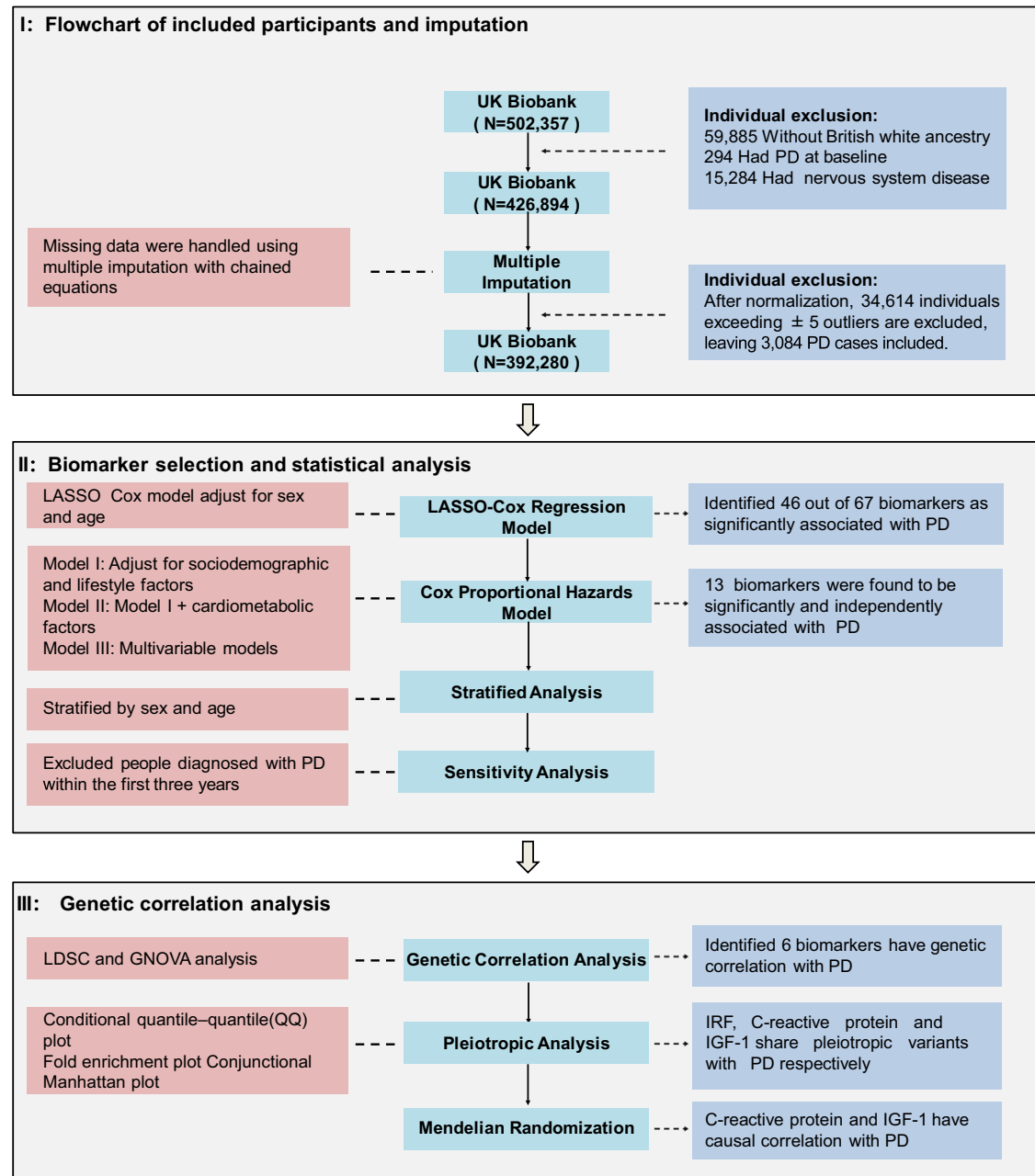
Figure 3 shows the process for constructing the analytical cohort. From the initial 502,357 participants in the UK Biobank, we excluded 59,885 individuals without British white ancestry, 294 individuals with PD at baseline, and 15,284 individuals with pre-existing nervous system diseases. After applying these criteria, we assessed the impact of missing trait values, particularly for the multivariable analysis, by performing multiple imputations using chained equations. This process was carried out over 10 iterations, generating 10 complete datasets for subsequent analysis<sup>57</sup>. Following normalization, we excluded 34,614 outliers exceeding  $\pm 5$  standard deviations. Consequently, the final sample size for the PD and biomarker analyses was 392,280 participants.

### Statistical Analyses

LASSO (Least Absolute Shrinkage and Selection Operator) is a form of regularization that reduces coefficients to enhance prediction accuracy. To preliminarily screen blood and urine biomarkers associated with PD onset, we performed gender and age adjusted LASSO Cox regression analysis, using a penalty parameter ( $\lambda$ ) optimized through tenfold cross-validation. This approach helped identify biomarkers most strongly associated with PD onset risk while minimizing the risk of overfitting<sup>58</sup>. The LASSO model was constructed using the “glmnet” package in R. This variable selection process was repeated across each imputed dataset, resulting in ten models. Subsequently, we set a threshold for inclusion in subsequent analytical models, requiring variables to retain nonzero coefficients in at least seven imputations of the LASSO model. Variables maintaining nonzero coefficients in fewer than seven LASSO models were excluded from further analysis<sup>57</sup>. We used the Rubin rule in all subsequent analyses to generate the final combined effect estimates<sup>59</sup>.

The confounding factors included in our study are age, sex, BMI, Townsend deprivation index, physical activity, smoking status, drinking status, education level, hypertension, stroke, and diabetes. Age refers to the age at enrollment, while BMI was measured using the Omron HEM-7015IT digital blood pressure monitor as the average of two sitting measurements of systolic blood pressure. BMI was calculated by dividing the individual's weight (in kilograms), measured using the Tanita BC-418 MA body composition analyzer, by the square of their height (in meters). The Townsend deprivation index assesses area deprivation based on four domains: unemployment, car ownership, household overcrowding, and owner occupation information, with higher scores indicating greater social deprivation. Smoking and drinking statuses were classified as never, previous, or current. Education level was categorized as whether the individual held a college/university degree or other professional qualifications. Stroke, hypertension, and diabetes were defined based on self-reported status and whether participants were taking antihypertensive medication or insulin.

Building upon the preliminary feature selection conducted using LASSO-Cox regression, we employed Cox proportional hazards regression models to further evaluate the relationship between blood and urine biomarkers and the onset of PD. By iteratively refining the models and incorporating progressively stringent covariates, we aimed to minimize potential confounding and obtain more precise estimates of associations. Specifically, we started with Model 1, which included adjustments for key confounders such as gender, age, Townsend deprivation index, smoking, alcohol consumption, BMI, physical activity, and education. Biomarkers with statistical significance ( $P < 0.05$ , Bonferroni corrected) in Model 1 were included in Model 2, which further accounted for cardiovascular and



**Fig. 3 | Flowchart of the study.** The study involves three main parts: participant selection, biomarker selection and statistical analysis, and genetic correlation analysis. In Part I, we started with 502,357 participants from the UK Biobank, applying exclusion criteria and multiple imputation, resulting in a final sample of 392,280

participants. In Part II, we used LASSO-Cox regression and a progressively stringent covariate-adjusted Cox proportional hazards mode. In Part III, we conducted genetic correlation analysis, identifying genetic links and causal relationships for key biomarkers. PD Parkinson's disease, CRP C-reactive Protein, QQ quantile–quantile.

metabolic factors, including hypertension, diabetes, and stroke. Ultimately, Model 3 integrated the significant biomarkers identified in Model 2, while adjusting for all potential confounders, allowing us to identify biomarkers with independent associations with PD onset. We regard these findings as a starting point for a more in-depth investigation to explore their further significance and potential implications.

### Genetic Correlation

The LDSC sumstats format files used in genetic correlation analysis are derived from the genome-wide association study (GWAS) round 2 data from the Neale Lab and are available for free on their website (<https://www.nealelab.is/uk-biobank>). We utilized default parameters in the Gnova and LDSC methods to estimate the genetic correlation between significant biomarkers and PD risk within the Cox proportional hazards regression model.

LDSC estimates the correlation between increased genetic susceptibility to diseases and quantitative trait values by analyzing single nucleotide polymorphisms (SNPs). By examining the relationship between test statistics and LDSC, it allows for the quantification of the contribution of each SNP to the overall genetic risk. This method can be used to assess whether different blood biomarkers share a common genetic basis with PD, providing insight into the potential genetic overlap between these biomarkers and PD susceptibility.

In the results of genetic correlation analysis between different phenotype pairs, a positive correlation indicates that the increased risk for one condition or higher values of a continuous trait is associated with increased risk or higher values of another trait. Conversely, a negative correlation means that an increased risk for one condition or higher values of a trait is associated with a decreased risk or lower values of another trait<sup>60,61</sup>. GNOVA



estimates genetic covariance using summary data of shared genetic variations between two GWAS and then calculates genetic correlation based on genetic covariance and variance-based heritability. Both LDSC and GNOVA analyses can statistically explain the possibility of sample overlap between PD and biomarkers.

### Pleiotropic Enrichment Analysis

For biomarkers with significant genetic associations with PD, we further investigated whether the genetic overlap between these biomarkers and PD is mediated by shared regulatory mechanisms in gene expression.

To evaluate the pleiotropic enrichment between two traits, stratified Q-Q plots were employed. Such plots typically depict nominal P-values ( $-\log_{10}(p)$ ) on the y-axis against empirical quantiles ( $-\log_{10}(q)$ ) on the x-axis. Pleiotropic enrichment can be discerned by the leftward deviation from the expected null line, as the primary trait successively depends on different significance levels of blood biomarkers. If indeed pleiotropic enrichment exists, an early departure from the null line towards the left will be evident. Larger gaps between the stratified Q-Q curves visually indicate higher levels of pleiotropic enrichment between the two traits.

To confirm these enrichment effects, fold-enrichment plots were generated. These plots illustrate the fold-enrichment of nominal  $-\log_{10}(P)$  values for PD SNPs below the conventional GWAS threshold of  $P < 5 \times 10^{-8}$ , as well as for subsets of SNPs categorized based on the significance of their association with biomarkers and vice versa. Fold-enrichment is evaluated by the extent of upward deviation from the null line.

We employed the conditional false discovery (condFDR) rate method to identify shared risk loci between biomarkers and PD. Specifically, condFDR was used to detect PD-associated SNPs related to the biomarker, and vice versa. The condFDR was utilized to identify SNPs jointly associated with both PD and the biomarker. After applying the condFDR procedure to both traits, the conjFDR analysis reported loci that simultaneously exceeded the significance threshold for condFDR in both traits, where the threshold is defined as the maximum of the condFDR values for the two traits, with a significance level set at 0.01<sup>62</sup>.

### GWAS Summary Statistics

In this study, the exposure and outcome GWAS data were obtained from the open-access database (<https://gwas.mrcieu.ac.uk/>)<sup>63,64</sup>. We utilized GWAS data for IGF-1 and CRP from the IEU database (ebi-a-GCST90014008 and ebi-a-GCST90014002, respectively). These data were derived from a genome-wide association study (GWAS) conducted on up to 407,746 participants of White British ancestry from the UK Biobank. Analyses were performed using the REGENIE software, which employed a rank-based inverse normal transformation to convert quantitative phenotypes into Z-scores. Regarding the IRF, we utilized publicly available GWAS results collected by the Neale Lab (ukb-d-30280\_irnt), which include summary statistics for both the raw and rank-normalized versions of these biomarker measurements. Additionally, the statistical model adjusted for population stratification by incorporating age, age squared, sex, an age-by-sex interaction term, and the top ten principal components provided by the UK Biobank<sup>65</sup>. GWAS data for PD outcomes were obtained from the International Parkinson's Disease Genomics Consortium (IPDGC; ieu-b-7). This analysis integrated data from three previously published studies, 13 new datasets, and UKD. Their case included clinician-identified PD cases and 23andMe self-reported cases and UK Biobank proxy cases<sup>66</sup>.

The instrumental variables (SNPs) must satisfy three core assumptions: First, relevance: SNPs must be robustly associated with the exposure. Second, independence: SNPs should not be influenced by confounders of the exposure-outcome relationship, such as population stratification or selection bias. Third, exclusion restriction: SNPs should affect the outcome only through the exposure pathway, implying no horizontal pleiotropy.

For Assumption 1, SNPs with  $P < 5 \times 10^{-8}$  served as instrumental variables for MR analysis, and excluded variants with an F-value  $< 10$  to minimize weak instrument bias<sup>67</sup>. In addition, We excluded SNPs located within a 5000-kb range upstream and downstream of the most significant SNP to minimize linkage disequilibrium and avoid including SNPs that might be correlated ( $r^2 > 0.01$  in the 1000 Genomes European data). For hypothesis 2, we excluded the same SNPs in the GWAS data of the covariates in our Cox proportional hazard regression model I to rule out potential mixing and found two SNPs rs11642090 and rs7259070 that existed between BMI and IGF-1. For Assumption 3, We used the MR-egger test to rule out pleiotropy.

### Bidirectional Mendelian Randomization

In this study, we employed four methods to investigate the potential causal relationship between biomarkers and PD: inverse-variance weighted (IVW), weighted mode, weighted median, and MR-Egger. When all selected SNPs are valid instruments, the IVW method is considered the most accurate and efficient for estimating causal effects, particularly suitable for using multiple genetic variants as instruments<sup>68</sup>. The weighted mode method is similar to the IVW method but allows for the consideration of correlation between genetic instruments and is used when employing a set of conservative genetic instruments. The weighted median method is a robust MR approach that estimates causal effects by calculating the median of the ratio estimates from genetic variants and is resilient to outliers. The MR-Egger method accommodates horizontal pleiotropy in the instruments but requires that the pleiotropy is independent of the variant-exposure association.

To test the robustness of the MR results, we conducted additional sensitivity analyses, including using Cochran's Q statistic to measure heterogeneity, MR-Egger regression to assess potential pleiotropy in the SNPs used as instruments, and leave-one-out analysis by sequentially removing genetic variants from the analysis and re-estimating the causal effect to evaluate the dependence on specific variants.

### Data availability

After the health-related research is approved by the UK Biobank Access Team, the data and materials are available online at <https://www.ukbiobank.ac.uk/>.

Received: 30 August 2024; Accepted: 6 February 2025;

Published online: 25 February 2025

### References

1. Global, regional, and national burden of neurological disorders, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Neurol.* **18**, 459–480 (2019).
2. Tolosa, E., Garrido, A., Scholz, S. W. & Poewe, W. Challenges in the diagnosis of Parkinson's disease. *Lancet Neurol.* **20**, 385–397 (2021).
3. Leite, S. A. et al. Premotor, nonmotor and motor symptoms of Parkinson's Disease: A new clinical state of the art. *Ageing Res Rev.* **84**, 101834 (2023).
4. Hughes, A. J., Daniel, S. E., Kilford, L. & Lees, A. J. Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. *J. Neurol. Neurosurg. Psychiatry* **55**, 181–184 (1992).
5. Rizzo, G. et al. Accuracy of clinical diagnosis of Parkinson disease. *Neurology* **86**, 566–576 (2016).
6. Virameteekul, S., Revesz, T., Jaunmuktane, Z., Warner, T. T. & De Pablo-Fernandez, E. Clinical Diagnostic Accuracy of Parkinson's Disease: Where Do We Stand? *Mov. Disord.* **38**, 558–566 (2023).
7. Khan, A. Z., Lavu, D. & Neal, R. D. Parkinson's disease: a scoping review of the quantitative and qualitative evidence of its diagnostic accuracy in primary care. *Br. J. Gen. Pract.* **74**, e227–e232 (2024).

8. Cocco, C., Manai, A. L., Manca, E. & Noli, B. Brain-Biomarker Changes in Body Fluids of Patients with Parkinson's Disease. *Int J. Mol. Sci.* **24**, 10932 (2023).
9. Tekriwal, A. et al. REM sleep behaviour disorder: prodromal and mechanistic insights for Parkinson's disease. *J. Neurol., Neurosurg. Psychiatry* **88**, 445–451 (2017).
10. Postuma, R. B. et al. Identifying prodromal Parkinson's disease: pre-motor disorders in Parkinson's disease. *Mov. Disord.* **27**, 617–626 (2012).
11. Espay, A. J. et al. Biomarker-driven phenotyping in Parkinson's disease: A translational missing link in disease-modifying clinical trials. *Mov. Disord.* **32**, 319–324 (2017).
12. Vijjaratnam, N. & Foltynie, T. How should we be using biomarkers in trials of disease modification in Parkinson's disease? *Brain* **146**, 4845–4869 (2023).
13. Simuni, T. et al. A biological definition of neuronal  $\alpha$ -synuclein disease: towards an integrated staging system for research. *Lancet Neurol.* **23**, 178–190 (2024).
14. Parnetti, L. et al. CSF and blood biomarkers for Parkinson's disease. *Lancet Neurol.* **18**, 573–586 (2019).
15. Ganguly, U. et al. Alpha-Synuclein as a Biomarker of Parkinson's Disease: Good, but Not Good Enough. *Front Aging Neurosci.* **13**, 702639 (2021).
16. Atik, A., Stewart, T. & Zhang, J. Alpha-Synuclein as a Biomarker for Parkinson's Disease. *Brain Pathol.* **26**, 410–418 (2016).
17. Concha-Marambio, L., Pritzkow, S., Shahnawaz, M., Farris, C. M. & Soto, C. Seed amplification assay for the detection of pathologic alpha-synuclein aggregates in cerebrospinal fluid. *Nat. Protoc.* **18**, 1179–1196 (2023).
18. Mollenhauer, B. et al. Validation of Serum Neurofilament Light Chain as a Biomarker of Parkinson's Disease Progression. *Mov. Disord.* **35**, 1999–2008 (2020).
19. Benedet, A. L. et al. Differences Between Plasma and Cerebrospinal Fluid Glial Fibrillary Acidic Protein Levels Across the Alzheimer Disease Continuum. *JAMA Neurol.* **78**, 1471 (2021).
20. Teunissen, C. E. et al. Blood-based biomarkers for Alzheimer's disease: towards clinical implementation. *Lancet Neurol.* **21**, 66–77 (2022).
21. Benkert, P. et al. Serum neurofilament light chain for individual prognostication of disease activity in people with multiple sclerosis: a retrospective modelling and validation study. *Lancet Neurol.* **21**, 246–257 (2022).
22. Fang, F. et al. Lipids, Apolipoproteins, and the Risk of Parkinson Disease. *Circ. Res.* **125**, 643–652 (2019).
23. Yazdani, S. et al. Peripheral immune biomarkers and neurodegenerative diseases: A prospective cohort study with 20 years of follow-up. *Ann. Neurol.* **86**, 913–926 (2019).
24. Schwarzschild, M. A. et al. Effect of Urate-Elevating Inosine on Early Parkinson Disease Progression: The SURE-PD3 Randomized Clinical Trial. *JAMA* **326**, 926–939 (2021).
25. Moccia, M. et al. Increased bilirubin levels in de novo Parkinson's disease. *Eur. J. Neurol.* **22**, 954–959 (2015).
26. Kim, R., Kang, N., Byun, K., Park, K. & Jun, J. Prognostic significance of peripheral neutrophils and lymphocytes in early untreated Parkinson's disease: an 8-year follow-up study. *J. Neurol., Neurosurg. psychiatry* **94**, 1040–1046 (2023).
27. Beura, S. K., Panigrahi, A. R., Yadav, P. & Singh, S. K. Role of platelet in Parkinson's disease: Insights into pathophysiology & theranostic solutions. *Ageing Res Rev.* **80**, 101681–101681 (2022).
28. Chung, H. S. et al. Fasting plasma glucose variability in midlife and risk of Parkinson's disease: A nationwide population-based study. *Diab. Metab.* **47**, 101195–101195 (2021).
29. Stuart, C. A., Meehan, R. T., Neale, L. S., Cintron, N. M. & Furlanetto, R. W. Insulin-Like Growth Factor-I Binds Selectively to Human Peripheral Blood Monocytes and B-Lymphocytes\*. *J. Clin. Endocrinol. Metab.* **72**, 1117–1122 (1991).
30. Qin, X., Zhang, S., Cao, C., Loh, Y. P. & Cheng, Y. Aberrations in Peripheral Inflammatory Cytokine Levels in Parkinson Disease: A Systematic Review and Meta-analysis. *JAMA Neurol.* **73**, 1316–1316 (2016).
31. Lewis, N. A. & Knight, J. E. Longitudinal associations between C-reactive protein and cognitive performance in normative cognitive ageing and dementia. *Age Ageing* **50**, 2199–2205 (2021).
32. Wu, C. et al. Association of AST/ALT (De Ritis) ratio with sarcopenia in a Chinese population of community-dwelling elderly. *Heliyon* **9**, e20427 (2023).
33. Maeda, D. et al. Aspartate aminotransferase to alanine aminotransferase ratio is associated with frailty and mortality in older patients with heart failure. *Sci. Rep.* **11**, 11957 (2021).
34. Elustondo, P. A. et al. Mitochondrial permeability transition pore induction is linked to formation of the complex of ATPase C-subunit, polyhydroxybutyrate and inorganic polyphosphate. *Cell Death Discov.* **2**, 16070 (2016).
35. Dionísio, P. A., Amaral, J. D. & Rodrigues, C. M. P. Oxidative stress and regulated cell death in Parkinson's disease. *Ageing Res Rev.* **67**, 101263 (2021).
36. Phadwal, K., Vrahnas, C., Ganley, I. G. & MacRae, V. E. Mitochondrial Dysfunction: Cause or Consequence of Vascular Calcification? *Front Cell Dev. Biol.* **9**, 611922 (2021).
37. Piva, E., Brugnara, C., Chiandetti, L. & Plebani, M. Automated reticulocyte counting: state of the art and clinical applications in the evaluation of erythropoiesis. *cclm* **48**, 1369–1380 (2010).
38. Cho, I. Y. et al. Anemia and the risk of Parkinson's disease in Korean older adults: A nationwide population-based study. *Sci. Rep.* **10**, 4268 (2020).
39. Hong, C. T. et al. Newly Diagnosed Anemia Increases Risk of Parkinson's disease: A Population-Based Cohort Study. *Sci. Rep.* **6**, 29651 (2016).
40. Kummeling, J. et al. Characterization of SETD1A haploinsufficiency in humans and Drosophila defines a novel neurodevelopmental syndrome. *Mol. Psychiatry* **26**, 2013–2024 (2021).
41. Mukai, J. et al. Recapitulation and Reversal of Schizophrenia-Related Phenotypes in Setd1a-Deficient Mice. *Neuron* **104**, 471–487.e12 (2019).
42. Yu, E. et al. Fine mapping of the HLA locus in Parkinson's disease in Europeans. *NPJ Parkinsons Dis.* **7**, 84 (2021).
43. Le Guen, Y. et al. Multi-ancestry analysis of the HLA locus in Alzheimer's and Parkinson's diseases uncovers a shared adaptive immune response mediated by HLA-DRB1\*04 subtypes. *Proc. Natl Acad. Sci. USA* **120**, e2302720120 (2023).
44. Bernhard, F. P. et al. Insulin-Like Growth Factor 1 (IGF-1) in Parkinson's Disease: Potential as Trait-, Progression- and Prediction Marker and Confounding Factors. *PLoS One* **11**, e0150552 (2016).
45. Castilla-Cortázar, I., Aguirre, G. A., Femat-Roldán, G., Martín-Estal, I. & Espinosa, L. Is insulin-like growth factor-1 involved in Parkinson's disease development? *J. Transl. Med.* **18**, 70 (2020).
46. Allwright, M., Mundell, H., Sutherland, G., Austin, P. & Guennewig, B. Machine learning analysis of the UK Biobank reveals IGF-1 and inflammatory biomarkers predict Parkinson's disease risk. *PLoS One* **18**, e0285416 (2023).
47. Black, S., Kushner, I. & Samols, D. C-reactive Protein. *J. Biol. Chem.* **279**, 48487–48490 (2004).
48. Mehta, N., Luthra, N. S., Corcos, D. M. & Fantuzzi, G. C-reactive protein as the biomarker of choice to monitor the effects of exercise on inflammation in Parkinson's disease. *Front Immunol.* **14**, 1178448 (2023).
49. Qin, X., Zhang, S., Cao, C., Loh, Y. P. & Cheng, Y. Aberrations in Peripheral Inflammatory Cytokine Levels in Parkinson Disease. *JAMA Neurol.* **73**, 1316 (2016).
50. Lawton, M. et al. Blood Biomarkers With Parkinson's Disease Clusters and Prognosis: The Oxford Discovery Cohort. *Mov. Disord.* **35**, 279–287 (2020).

51. Dommershuijsen, L. J. et al. Peripheral Immune Cell Numbers and C-Reactive Protein in Parkinson's Disease: Results from a Population-Based Study. *J. Parkinson's. Dis.* **12**, 667–678 (2022).
52. Sudlow, C. et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* **12**, e1001779 (2015).
53. Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
54. Gong, J., Harris, K., Peters, S. & Woodward, M. Serum lipid traits and the risk of dementia: A cohort study of 254,575 women and 214,891 men in the UK Biobank. *EClinicalMedicine* **54**, 101695 (2022).
55. Rhee, E. J., Byrne, C. D. & Sung, K. C. The HDL cholesterol/apolipoprotein A-I ratio: an indicator of cardiovascular disease. *Curr. Opin. Endocrinol. Diab. Obes.* **24**, 148–153 (2017).
56. Song, F. et al. Plasma apolipoprotein levels are associated with cognitive status and decline in a community cohort of older individuals. *PLoS One* **7**, e34078 (2012).
57. Payne, B. A. et al. Development and internal validation of the multivariable CIPHER (Collaborative Integrated Pregnancy High-dependency Estimate of Risk) clinical risk prediction model. *Crit. Care.* **22**, 278 (2018).
58. Tibshirani, R. The lasso method for variable selection in the Cox model. *Stat. Med.* **16**, 385–395 (1997).
59. Moons, K. G., Donders, R. A., Stijnen, T. & Harrell, F. J. Using the outcome for imputation of missing predictor values was preferred. *J. Clin. Epidemiol.* **59**, 1092–1101 (2006).
60. Bulik-Sullivan, B. K. et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
61. Bulik-Sullivan, B. et al. An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).
62. Andreassen, O. A. et al. Improved detection of common variants associated with schizophrenia and bipolar disorder using pleiotropy-informed conditional false discovery rate. *PLoS Genet.* **9**, e1003455 (2013).
63. Zheng, H., Shi, Y. Z., Liang, J. T., Lu, L. L. & Chen, M. Modifiable factors for migraine prophylaxis: A mendelian randomization analysis. *Front Pharmacol.* **14**, 1010996 (2023).
64. Hemani, G. et al. The MR-Base platform supports systematic causal inference across the human phenome. *Elife* **7**, e34408 (2018).
65. Mbatchou, J. et al. Computationally efficient whole-genome regression for quantitative and binary traits. *Nat. Genet.* **53**, 1097–1103 (2021).
66. Nalls, M. A. et al. Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet Neurol.* **18**, 1091–1102 (2019).
67. Burgess, S. & Thompson, S. G. Avoiding bias from weak instruments in Mendelian randomization studies. *Int J. Epidemiol.* **40**, 755–764 (2011).
68. Burgess, S., Butterworth, A. & Thompson, S. G. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet Epidemiol.* **37**, 658–665 (2013).

## Acknowledgements

We express our gratitude to all the participants and their families, as well as all the investigators and members of the UK Biobank. And we are grateful for resources from the High-Performance Computing Center of Central South University. This study was supported by the National Key R&D Program of China (Grant No. 2021YFC2502100), the Hunan Innovative Province Construction Project (Grant No. 2021SK1010), the Central South University Research Programme of Advanced Interdisciplinary Study (Grant No.

2023QYJC010), the Fundamental Research Funds for the Central Universities of Central South University (Grants No. 2023ZZTS0831), the 2023 Key R&D Projects in Hunan Province (Grant No. 2023SK2005) and the Scientific Research Program of FuRong laboratory (Grant No. 2023SK2093-1).

## Author contributions

Shuo Gao: Conceptualisation, formal analysis, Methodology, Validation, Writing of the original draft. Zheng Wang: Methodology, Writing, review, and editing. Yuanfeng Huang: Methodology, Writing, review, and editing. Guang Yang: Methodology, Writing, review, and editing. Yijing Wang: Methodology, Writing, review, and editing. Yan Yi: Methodology, Writing, review, and editing. Qiao Zhao: Methodology, Writing, review, and editing. Xingxing Jian: Methodology, Writing, review and editing. Bin Li: Methodology, Writing, review and editing. Guihu Zhao: Conceptualisation, Writing, review, and editing. Bin Li: Methodology, Writing, review and editing. Linyong Xu: Conceptualisation, Writing, review, and editing. Kun Xia: Conceptualisation and formal analysis. Beisha Tang: Conceptualisation and formal analysis. Jinchun Li: conceptualisation, formal analysis, writing, reviewing, and editing. All authors contributed to the article and approved the submitted version.

## Competing interests

The authors declare no competing interests.

## Ethics approval and consent to participate

The UK Biobank study has been approved by the North West Multi-centre Research Ethics Committee as a Research Tissue Bank and all UK Biobank participants provided written informed consent.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41531-025-00888-2>.

**Correspondence** and requests for materials should be addressed to Jinchun Li.

**Reprints and permissions 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. Consent for publication Not applicable.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. 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-nc-nd/4.0/>.

© The Author(s) 2025