Twin Research and Human Genetics (2025), 1–9
doi:10.1017/thg.2025.10008

CAMBRIDGE UNIVERSITY PRESS

Article

A Seven-Gene Signature for the Diagnosis of Parkinson's Disease and Immune Infiltration Analysis

Chengqun Wei[1*], Rui Xue[2*], Zhan Gao[2], Hongyan Zhu[2] and Xiuzhi Xu[2]

[1]Department of Neurology, Huaian Hospital of Huaian city-Huaian Cancer Hospital), Huai'an, China and [2]Department of General Practice, Heilongjiang Provincial Hospital, Harbin, China

Abstract

The objective was to identify the predictive markers and develop a diagnostic model with predictive markers for Parkinson's disease (PD) and investigate the roles of immune cells in the disease pathology. Microarray datasets of PD and control samples were obtained from the Gene Expression Omnibus (GEO) database. We then performed a comprehensive analysis of differentially expressed genes (DEGs), functional enrichment, and protein-protein interactions to pinpoint a set of promising candidate genes. To establish a diagnosis model for PD, we utilized machine learning algorithms and evaluated the corresponding diagnostic performance using the receiver operating characteristic (ROC) curve and the area under the ROC curve (AUC). Additionally, the differential abundance of immune cell subsets between PD and control samples was evaluated using the single-sample Gene Set Enrichment Analysis (ssGSEA) method. A total of 264 DEGs were identified in GSE72267. The PPI network ultimately identified 30 hub genes for model construction. Seven genes, namely *CD79B*, *CD40*, *CCR9*, *ADRA2A*, *SIGLEC1*, *FLT3LG*, and *THBD*, were identified as diagnostic markers for PD, with an AUC of 0.870. This seven-gene signature model was subsequently validated in an independent cohort (GSE22491), demonstrating an AUC of 0.825. Ultimately, the infiltration of 28 immune cells showed that activated B cells, natural killer T cells, and regulatory T cells may contribute to the occurrence and progression of PD. We also found complex associations between these genes and immune cells. *CD79B*, *CD40*, *CCR9*, *ADRA2A*, *SIGLEC1*, *FLT3LG*, and *THBD* were identified as diagnostic markers for PD, and the infiltration of immune cells may contribute to the pathogenesis of the disease.

Keywords: Parkinson's disease; Diagnosis; Gene signature; Immune infiltration

(Received 27 February 2025; revise received 22 April 2025; accepted 1 May 2025)

Parkinson's disease (PD) is one of the most common neurodegenerative disorders affecting older individuals. More than 6 million people worldwide suffer from PD, and this number is expected to increase significantly as the population ages (GBD 2016 Parkinson's Disease Collaborators, 2018). One of the major pathological features of PD is the accumulation of misfolded α-synuclein into Lewy bodies and Lewy neurites (Mahul-Mellier et al., 2020). This pathological hallmark causes motor symptoms, such as tremors, stiffness, and bradykinesia, as well as a range of nonmotor symptoms, including hyposmia, mood disorders, sleep disturbance, and cognitive impairment, which can lead to reduced mobility, function, and quality of life for PD patients (Armstrong & Okun, 2020; Mu et al., 2017).

Over the past decade, an emerging perspective has emphasized the significance of immune dysfunction and the immune system as a potentially pivotal component of PD etiology (Zhu et al., 2022). A substantial epidemiological study has revealed the risk of PD in patients with autoimmune diseases is 33% higher, indicating that autoimmune diseases may play a role in promoting or inhibiting the pathogenesis of PD (Li et al., 2012). Accumulated evidence indicates that both cellular and humoral immune responses are altered in PD patients (Kustrimovic et al., 2018; Wijeyekoon et al., 2018). Additionally, several PD-related genes, such as leucine-rich repeat kinase 2 (*LRRK2*) and parkin RBR E3 ubiquitin-protein ligase (*PRKN*), have been found to play functional roles in immune processes (Matheoud et al., 2016; Rivas et al., 2018). However, the mechanisms by which innate or adaptive immune systems regulate immune responses in PD remain poorly understood.

In recent years, high-throughput technology has not only provided more comprehensive genetic characterization for various diseases but has also been able to identify new and more accurate biomarkers (Chowdhary et al., 2022). Several diagnostic models for PD have been reported. Zhao et al. (2022) developed a PD diagnostic model using proptosis-related genes, with an area under the curve (AUC) of 0.752 in the corresponding ROC curve. Similarly, another study constructed a biomarker panel of 85 genes with decreased expression and increased DNA in PD using linear regression analysis, and developed a gene expression classifier with an AUC of 0.74 and two gene methylation classifiers with AUCs of 0.685 and 0.677 based on all or dominant methylation-altered region CpGs (Wang et al., 2019). All of these models show good predictive power for PD diagnosis. However, the diagnostic performance of these models still needs to be improved. In this study, we aimed to identify the predictive markers and develop a diagnostic model with predictive markers for PD and investigate the roles of immune cells in the disease pathology. Our study may

Corresponding author: Xiuzhi Xu; Email: 18576301387@163.com
*Chengqun Wei and Rui Xue contributed to this work equally.

Cite this article: Wei C, Xue R, Gao Z, Zhu H, Xu X. A Seven-Gene Signature for the Diagnosis of Parkinson's Disease and Immune Infiltration Analysis. *Twin Research and Human Genetics* https://doi.org/10.1017/thg.2025.10008

© The Author(s), 2025. Published by Cambridge University Press on behalf of International Society for Twin Studies.

https://doi.org/10.1017/thg.2025.10008 Published online by Cambridge University Press

contribute to improving the accuracy of PD diagnosis and elucidating the potential role of immune cells in PD pathology.

Materials and Methods

Data Preparation

We obtained microarray datasets from the public GEO database (https://www.ncbi.nlm.nih.gov/geo). The following were inclusion criteria for the database: (1) each dataset contained the human gene expression profiles of both PD patients and healthy controls; (2) there were at least five PD and control tissue samples in each dataset; and (3) the series matrix file was available. Using these criteria, we downloaded series matrix files from GSE72267 (Calligaris et al., 2015) containing blood samples of 40 PD patients and 19 healthy controls, and GSE22491 (Mutez et al., 2011) containing blood samples of 10 PD patients and 8 healthy controls from GEO and retrieved the group information and expression data using the 'GEOquery package' of R software. All datasets were preprocessed, normalized, and $\log^2$-transformed. We observed that GSE72267 (sporadic PD) and GSE22491 (LRRK2-related PD) represent distinct genetic subtypes of PD. This heterogeneity reflects the diversity of PD in etiology and clinical manifestations. Nonetheless, we acknowledge that such heterogeneity may have an impact on the generalization of the model.

DEGs Screening

DEGs between the two groups in GSE72267 were screened by the 'limma' package. The genes that met the criterion of '$p < .05$ and $|\text{fold-change}| > 1.2$' were considered statistically significant. A volcano plot displaying all DEGs was generated using the 'ggplot2' package, and we constructed a heatmap of the top 10 DEGs using the 'pheatmap' of the R package.

Functional Enrichment Analysis

Gene ontology (GO) enrichment analyses of upregulated and downregulated DEGs were performed using Bioconductor's 'clusterProfiler' package of the R package. For the reference gene set, we chose the 'h.all.v7.2.symbols' collection in the Molecular Signatures Database (MsigDB) and executed GSEA on all-gene profiling with the above package. Enrichment results with a nominal $p < .05$ and false discovery rate (FDR) < 0.25 were considered statistically significant. The leading edges analysis was performed by the 'Pi' package.

PPI Network Analysis

To illustrate the potential relation between the DEGs, we conducted a PPI network using STRING (https://string-db.org/), an online database of known and predicted protein interactions. The network was established with a threshold value of a combined score > 0.4 and visualized by Cytoscape (v3.7.2). All proteins or genes were scored by CytoHubba, a Cytoscape plugin, to identify the hub genes in the network (Clustering Coefficient > 0.25 and maximal clique centrality (MCC) > 30).

LASSO Model Construction and Performance Assessment

We employed the LASSO logistic regression to screen the diagnostic markers for PD. The LASSO model was built using the 'glmnet' package by utilizing the expression matrix of hub genes. In this study, the dataset was randomly split into training and testing datasets (50% : 50%) with the 'caret' of the R package. We evaluated the model by fitting it to the training set and used the fitted model to predict the test set. We further validated the model in an independent validation dataset, GSE22491. Ultimately, we used the 'pROC' package to generate ROC curves and the 'ggplot2' package to visualize the performance of the model.

Immune Infiltration Analysis

The set of 28 immune cell marker genes used in this study was integrated from the validated marker genes in the TCGA Immunogenomic Program (Danaher et al., 2017) and the GEO single-cell database (Lu et al., 2022). We then calculated the ssGSEA scores for each sample and created boxplots using the 'ggplot2' package to visualize the differences in immune cell abundance. Finally, we generated a correlation matrix and heatmap to illustrate the correlations among the 28 immune cell subpopulations using the 'corrplot' R package. Furthermore, the differential expression of those diagnostic markers between the two groups was visualized using boxplots, while the correlation between these genes and immune subtypes in PD patients was represented using lollipop plots. Multiple testing correction was applied when assessing correlations between diagnostic genes and immune cell subtypes.

Results

DEG Screening Between the Control and PD Groups

As shown in Figure 1, we identified 264 DEGs in GSE72267 by employing the criteria of '$p < 0.05$ and fold-change > 1.2', with 104 upregulated and 164 downregulated in the PD group in comparison with the control group. The DEGs were visualized in a volcano map (Figure 2A). Furthermore, we visualized the top 10 DEGs in a heatmap (Figure 2B).

Functional Enrichment Analysis of the 264 DEGs

We performed Gene Ontology (GO) enrichment and Gene Set Enrichment Analysis (GSEA) to analyze the biological functions and pathways of those DEGs. GO enrichment indicated that the upregulated genes were primarily associated with the regulations of MAP kinase activity, interleukin-6 (IL-6) production, and response to cAMP terms (Figure 3A). The downregulated genes

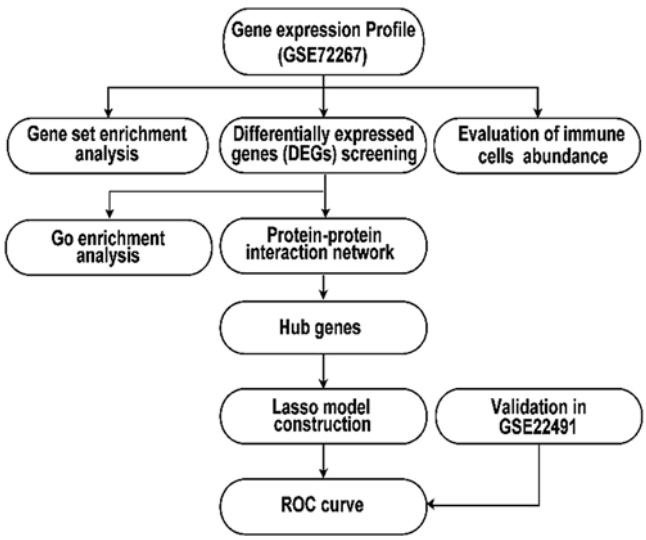

Figure 1. Graphical illustration of the workflow.

https://doi.org/10.1017/thg.2025.10008 Published online by Cambridge University Press

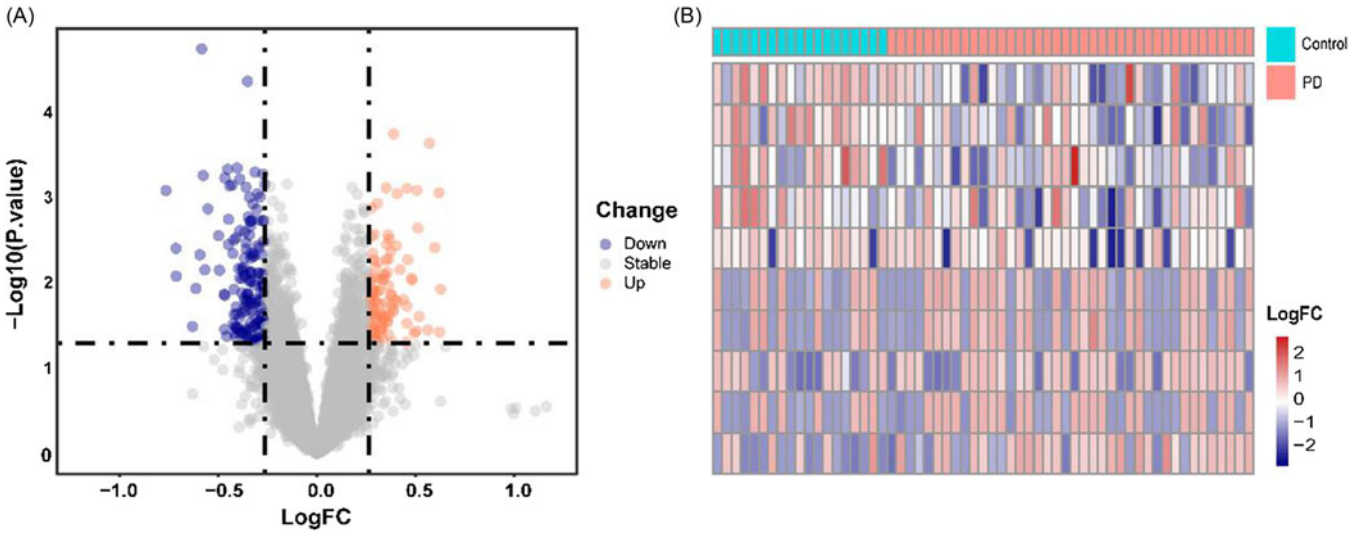

Figure 2. Differential expression genes (DEGs) visualizing. (A) Volcano plot of DEGs; orange represents upregulated and blue represents downregulated, while genes without differentially expression are indicated in grey. (B) Heatmap for the top 10 DEGs.

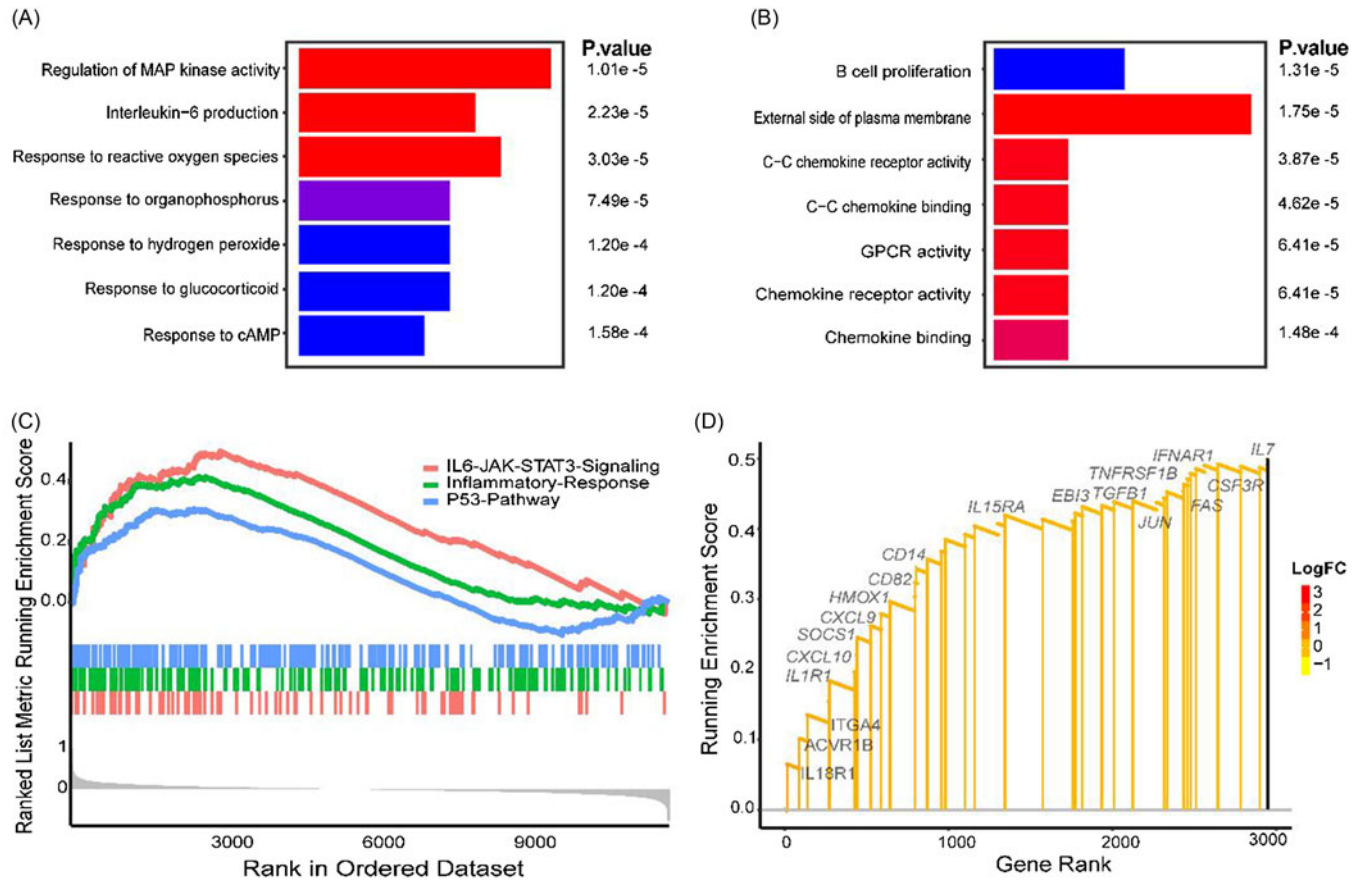

Figure 3. Functional enrichment analysis. (A) GO enrichment analysis of upregulated genes. (B) GO enrichment analysis of downregulated genes. (C) Gene set enrichment analysis (GSEA) of all genes. (D) Leading edges analysis of the above three pathways.

were mainly enriched for the B cell proliferation, G protein-coupled chemoattractant receptor (GPCR) activity and chemokine receptor activity functions terms (Figure 3B). GSEA revealed that IL6-JAK-STAT3 signaling, the inflammatory response, and P53 signaling were significantly activated in the PD group (Figure 3C). Notably, the leading-edge analysis highlighted the involvement of several key genes, including *CD14, CD82, CXCL9, IL7, FAS*, and *TGFB1*, in at least two of the three enriched pathways (Figure 3D).

https://doi.org/10.1017/thg.2025.10008 Published online by Cambridge University Press

Collectively, our results supported the notion of immune response playing a potentially important role in PD progression.

PPI Network Analysis and Hub Gene Calculation

A PPI network consisting of 173 nodes and 335 edges was successfully constructed. The connectivity degrees of nodes were differentiated by red, orange, and yellow colors, indicating high, high-intermediate, and intermediate connectivity respectively. Based on the scores of all proteins in the PPI network, 30 hub genes were identified using the network centrality algorithm (Figures 4A, 4B). Notably, *CD19*, *CD79B*, *CD40*, *CXCR5*, *CXCR3*, *CCR9*, and *SIGLEC* were among the hub genes identified.

LASSO Model Construction and Performance Assessment

To further investigate the potential diagnostic value of these hub genes, we applied the least absolute shrinkage and selection operator (LASSO) algorithm to the expression profiles of these 30 hub genes using a training dataset. Seven feature genes with nonzero coefficients were obtained in the LASSO regression model when the value of lambda.min was equal to 0.04157 (Figure 5A), including *CD79B, CD40*, *CCR9*, *ADRA2A*, *SIGLEC1*, *FLT3LG*, and *THBD*. A predictive index was then calculated based on the expression levels of these seven genes using a specific formula: index = *CD79B* × 0.13397380 + *CD40* × (-0.52086772) + *CCR9* × (-0.29455256) + *ADRA2A* × (-0.11868577) + *SIGLEC1* × 0.09235028 + *FLT3LG* × (-0.06840066) + *THBD* × 0.09333945. To evaluate the diagnostic performance of the model, a receiver operating characteristic (ROC) analysis was performed using an internal test set. The diagnostic model had a good diagnostic accuracy, with an AUC of 0.870 (Figure 5B). The diagnostic performance was confirmed in the validation set (GSE22491), and an AUC was 0.825 (Figure 5B). Collectively, our results suggested that the seven-gene model had the potential to serve as a biomarker for PD.

Immune Infiltration in PD

To further explore the immune infiltration in PD, ssGSEA was performed to estimate the relative proportions of various immune cell subtypes. There were striking differences between the PD and control groups. Specifically, regulatory T cells (Tregs) and natural killer T cells were elevated, while activated B cells were decreased in the PD group in comparison with the control group (Figure 6A).

To investigate the correlation among immune subsets, we constructed a correlation heatmap of the 28 immune subpopulations (Figure 6B). It showed that monocytes had a significant positive correlation with activated dendritic cells, central memory CD8 T cells, macrophages, T follicular helper cells, and natural killer cells respectively. Moreover, monocytes had a negative correlation with activated CD4 T cells. Additionally, type 2 T helper cells and effector memory CD4 T cells exhibited a significant positive correlation.

We then compared the diagnostic gene expression differences between the PD and control groups. There were five down-regulated genes (*CD79B*, *CD40*, *CCR9*, *ADRA2A*, and *FLT3LG*) and two upregulated genes (*SIGLEC1* and *THBD*) in the PD group (Figure 7A-7G). Additionally, we examined the association between diagnostic markers and immune cell subtypes in PD patients, which revealed broad associations between genes and immune cells. Our findings demonstrated that *CD79B* and *CD40* were significantly positively correlated with activated B cells while significantly negatively correlated with memory B cells and CD56 bright natural killer cells (Figure 8A, 8B). Interestingly, *CCR9* and *ADRA2A* were significantly correlated with all immune cells in PD patients (Figure 8C, 8D). *SIGLEC1* showed a positive correlation with type 2 T helper cells and a negative correlation with activated CD8 T cells (Figure 8E). *FLT3LG* was positively correlated with central memory CD4 T cells and negatively correlated with natural killer cells and immature dendritic cells (Figure 8F). *THBD* was positively correlated with natural killer T cells and neutrophils, while negatively correlated with gamma delta T cells (Figure 8G).

Discussion

PD is a neurodegenerative disorder characterized by the progressive loss of dopaminergic neurons in the substantia nigra (Dickson, 2012). PD diagnosis is based on medical history, clinical examinations, neurological scales, response to dopaminergic medications, and the exclusion of other diseases (Marsili et al., 2018). Although PD is a

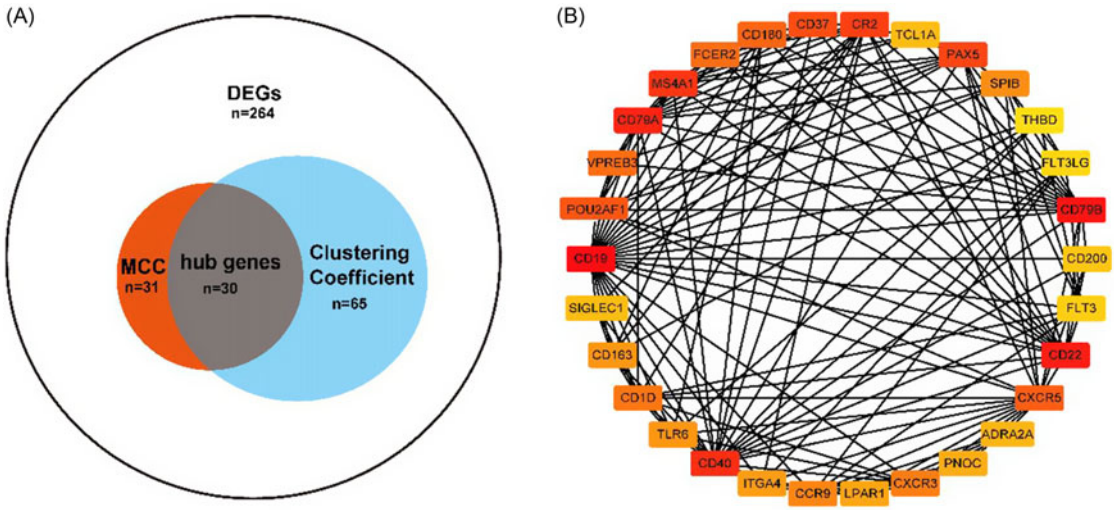

Figure 4. Hub gene calculation and protein-protein interaction (PPI) network. (A) Hub genes calculation; dark orange represents the MCC algorithm, and blue represents the Clustering Coefficient algorithm. (B) PPI network of 30 hub genes. Each node represents a gene and each edge indicates the interaction between genes. Red, orange, and yellow represent high, high-intermediate, and intermediate connective degrees respectively.

https://doi.org/10.1017/thg.2025.10008 Published online by Cambridge University Press

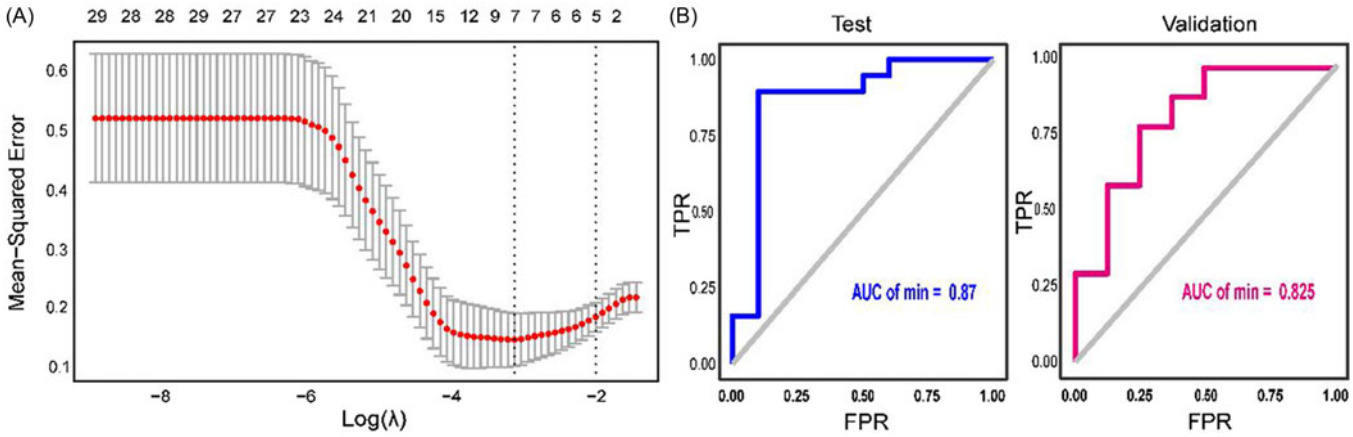

Figure 5. Building and verification of the LASSO diagnostic model. (A) LASSO logistic regression algorithm to screen potential diagnostic markers. There were seven feature genes with nonzero coefficients in the LASSO model when the value of lambda.min = 0.04157 (left dotted line). (B) The ROC (Receiver Operating Characteristic) curve of the diagnostic efficacy in the test dataset and validation dataset.

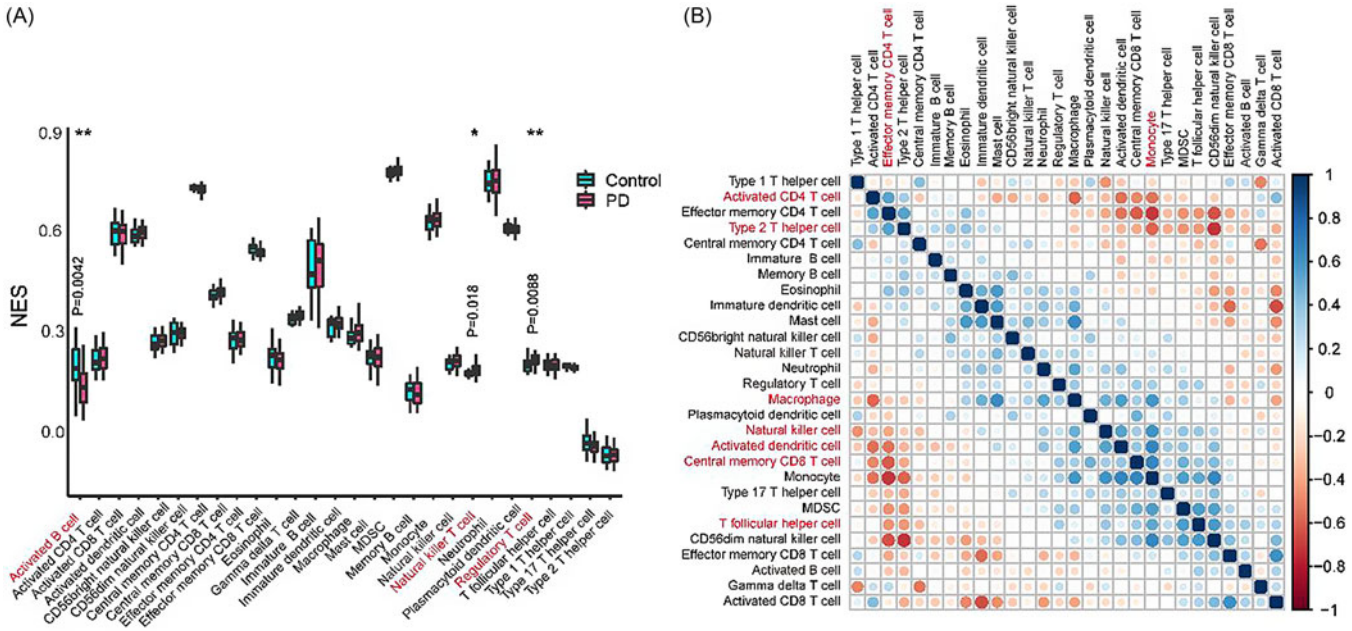

Figure 6. Immune cells abundance landscape. (A) Boxplot of the differential abundance of 28 immune cell subsets; the red marks indicate statistically significant differences between Parkinson's disease patients and controls. (B) Heatmap showing the correlation of 28 immune cells. Squares size represents the relative strength of immune cell relationships; blue (red) represents a positive (negative) correlation. A darker color indicates a stronger correlation.

multifactorial disease, both in vitro and in vivo studies have indicated that inflammatory processes and compensatory reactions are involved (Herrero et al., 2015). Several researches have demonstrated that myeloid and lymphoid populations (T and B cells) in the central nervous system may play a critical role in some neurodegenerative diseases, such as PD (Brochard et al., 2009; Doorn et al., 2014; Sommer et al., 2018) and Alzheimer's disease. In this study, we aimed to identify potential biomarkers for PD diagnosis and explore the roles of immune cells in the disease.

Recent advances in high-throughput technologies have enabled the identification of new and more accurate biomarkers for PD and other diseases through detailed genetic characterizations, and the growing availability of computationally intelligent techniques has been beneficial in the development of diagnostic models (Boutet et al., 2021; Kumar et al., 2022; Selvaraj et al., 2022). In this study, we used LASSO regression analysis to build a simpler and more interpretable prediction model for PD, which can improve the generalizability of the model and reduce overfitting. Additionally, LASSO regression analysis can handle high-dimensional data with a large number of predictors (Tibshirani, 1996).

In this study, we obtained the PD-related gene expression datasets from the GEO database and identified 264 DEGs. GO enrichment revealed that these DEGs were significantly associated with the regulation of IL-6 production and the activities of MAP kinases, GPCR, and chemokine receptors. The above results suggest that these DEGs play important roles in immune processes and signal transduction in PD patients. Additionally, we discovered that some pathways, such as IL6-JAK-STAT3, inflammatory response, and P53, were enriched in the PD group by using GSEA. It has been reported that the IL6-JAK-STAT3 signaling pathway widely participates in physiological processes, including cell growth,

https://doi.org/10.1017/thg.2025.10008 Published online by Cambridge University Press

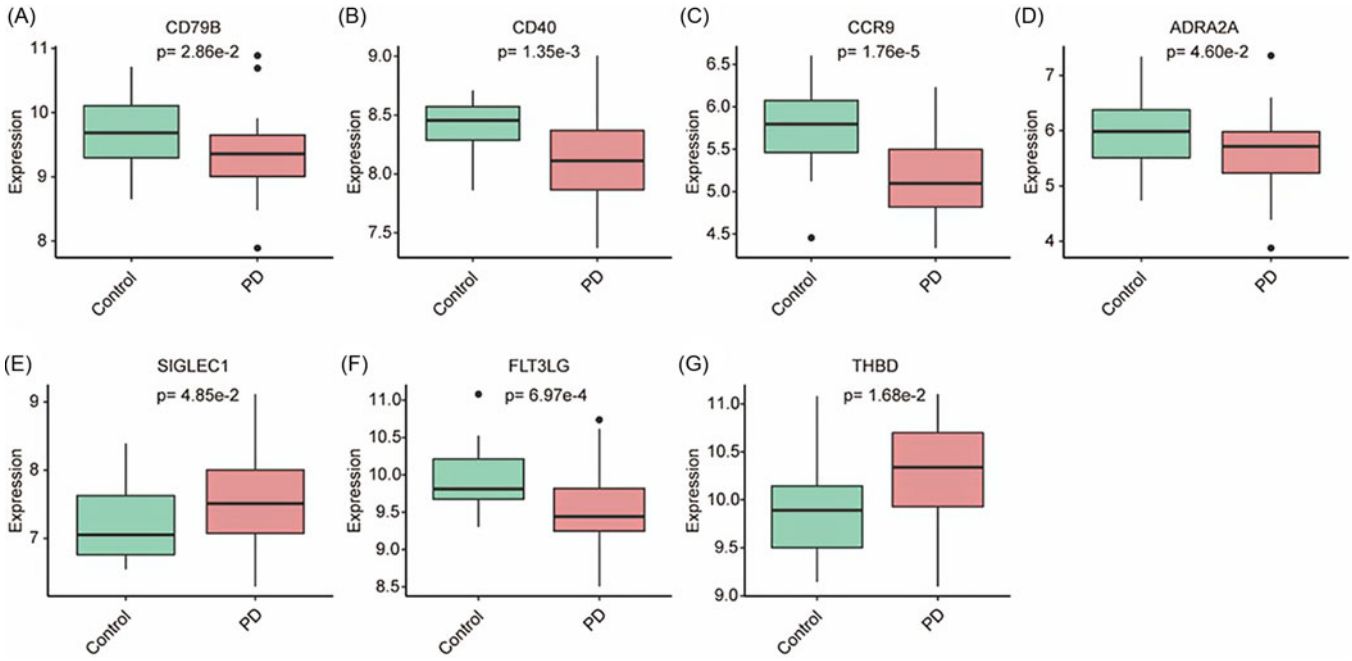

Figure 7. Differential expression of the diagnostic markers between Parkinson's disease and control groups in GSE72267. (A-G) The expression levels of *CD79B* (A), *CD40* (B), *CCR9* (C), *ADRA2A* (D), *SIGLEC1* (E), *FLT3LG* (F), and *THBD* (G) between the PD and control groups.

differentiation, and immune response (Niwa et al., 2005). Notably, increased levels of proinflammatory cytokines, including IL-6, have been observed in the brain and cerebrospinal fluid of PD patients, indicating that inflammation is involved in the disease progression (Lin et al., 2019). *P53* is involved in diverse biological pathways, and *P53* activation has been implicated in PD pathogenesis (Wolfrum et al., 2022). Its selective depletion in dopaminergic neurons improves their survival and behavioral motor deficits in a PD mouse model (Qi et al., 2016). Our results are consistent with the above previous studies.

To narrow down the list of candidate genes, we employed the MCC and clustering coefficient algorithms to identify 30 hub genes for model building. The LASSO regression analysis was then used to further shrink seven candidate genes, namely *CD79B*, *CD40*, *CCR9*, *ADRA2A*, *SIGLEC1*, *FLT3LG*, and *THBD*, as diagnostic markers for PD, with an AUC of 0.870 in test dataset and 0.825 in the validation set. According to our findings, the seven-gene model seems likely to be a biomarker for PD. Our results showed that their expression levels were significantly different in the PD group in comparison with the control group. Interestingly, some of these genes have previously been demonstrated to be associated with PD. For example, *CD79B* expression was significantly decreased in the peripheral blood of PD patients (Kedmi et al., 2011), and the variant of *ADRA2A* was associated with the frequency of insomnia in the PD group (Blauwendraat et al., 2019). Stimulation of microglia and astrocytes via CD40 signaling upregulates inducible nitric oxide synthase and cyclooxygenase-2, two known molecules involved in the pathogenesis of PD, causing selective loss of dopaminergic neurons in cell cultures (Okuno et al., 2005). Furthermore, *SIGLEC1* was related to alpha-synuclein regulation (Sarkar et al., 2020). However, the function of other candidates (*CCR9*, *FLT3LG*, and *THBD*) in PD has not been reported and deserves further study.

To evaluate the different patterns of immune cell abundance in PD individuals and healthy controls, ssGSEA was conducted. The results revealed that the increased abundance of natural killer T cells and Tregs accompanied by a decreased abundance of activated B cells may be relevant to the development of PD. Previous studies have shown that the activation of Tregs in the presence of antigen-presenting cells (APCs) can effectively inhibit the inflammatory response and the induction of effector T cells. Tregs can increase the expression level of neurotrophins released by astrocytes, thus effectively driving a neurotoxic environment to an opposite state (Schwab et al., 2020). Given that Tregs processes are often compromised in acute neuronal damage and chronic neurodegenerative diseases due to decreased numbers or dysfunction, Tregs may represent a promising therapeutic target for PD (Álvarez-Luquín et al., 2019). Horvath and Ritz (2015) found that compared to the healthy controls, PD patients display fewer B cells. Similarly, Stevens et al. (2012) reported a small reduction in the number of CD19+ B cells in the peripheral blood of PD patients. Interestingly, Kedmi et al. (2011) found a decreased expression of B cell-related genes only in female PD patients. Regrettably, to our knowledge, no previous study has addressed the association between natural killer T cells and PD. Natural killer T cells could have been underestimated due to their small population in peripheral blood. Our findings are largely in keeping with previous studies, illustrating that lymphoid cells may play important roles in PD etiology and should be further explored in later studies.

In addition, we also analyzed the correlations between different immune cell subsets. Our results revealed that activated dendritic cells, central memory CD8 T cells, macrophages, T follicular helper cells, and natural killer cells are closely related to monocytes, and type 2 T helper cells are closely related to effector memory CD4 T cells. Increasedα-synuclein levels were closely linked to reduced or impaired lysosomal function in monocytes (Chu et al., 2009; Stefanis et al., 2019). This may be due to reduced glucocerebrosidase activity in monocytes, which can lead to the failure of α-synuclein clearance and contribute to PD pathogenesis (Atashrazm et al., 2018). Furthermore, we found that the

https://doi.org/10.1017/thg.2025.10008 Published online by Cambridge University Press

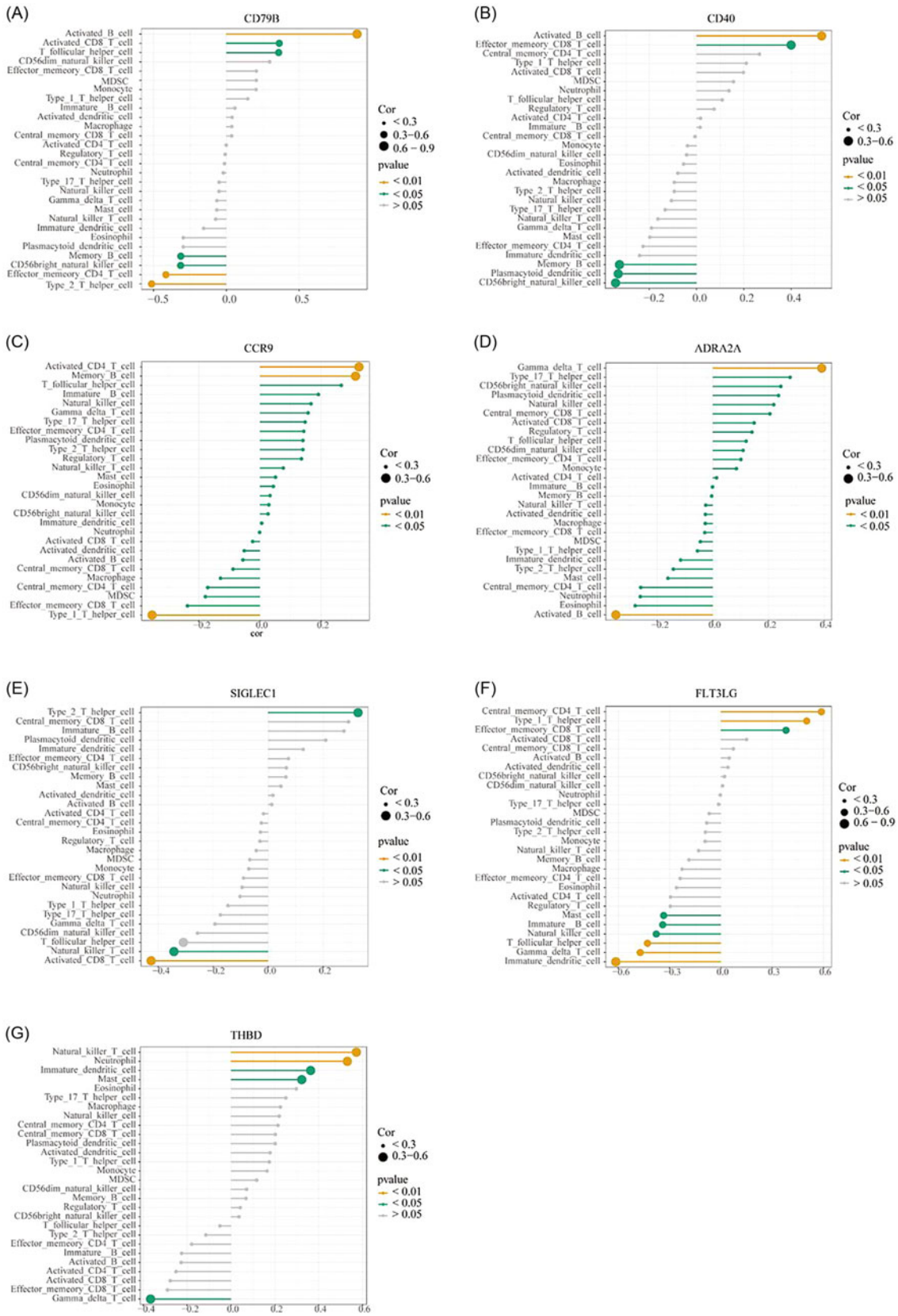

Figure 8. The correlation of the diagnostic markers and immune subtypes in Parkinson's disease (PD) patients. (A-G) Lollipop plots depicting person correlation between immune subtypes and diagnostic markers of *CD79B* (A), *CD40* (B), *CCR9* (C), *ADRA2A* (D), *SIGLEC1* (E), *FLT3LG* (F), and *THBD* (G) in PD patients.

https://doi.org/10.1017/thg.2025.10008 Published online by Cambridge University Press

correlation between these predictive genes and different immune subgroups was diverse, indicating that they may be involved in immune system regulation and possibly related to the onset and PD progression. The molecular mechanisms underlying these correlations warrant further investigation.

However, there are also some limitations in this study. First, the limited sample size in our model may lead to potential bias and limit the generalizability of our findings. Second, as a secondary analysis of previous data, our study is inherently limited by the quality and completeness of the original dataset. Moreover, there were the limited genetic subtypes covered by the dataset. Previous studies have indicated that *LRRK2* mutation-related PD may differ from sporadic PD in pathological processes (such as α-synuclein deposition patterns) and drug responsiveness (Dulski et al., 2022; Vinagre-Aragón et al., 2021). Nevertheless, our stratified analysis revealed that the seven-gene model is effective in both subtypes, which may be attributed to the involvement of these genes in common pathways across subtypes (such as endoplasmic reticulum stress and the ubiquitin-proteasome system). Future validation of the model's generalization in larger, multi-subtype cohorts is needed.

Conclusion

In conclusion, we identify a seven-gene signature as a candidate biomarker signature for PD. We also highlight the involvement of IL6-JAK-STAT3 signaling, the inflammatory response, and the P53 pathway in PD pathogenesis. Furthermore, our study provides new insights into the role of natural killer T cells, Tregs, and activated B cells in the development and progression of PD. Our findings underscore the importance of further studies on the immune subpopulations in PD, which may lead to the identification of novel therapeutic targets.

Authors' contributions

Wei C is the gurantor of the entire study. Xue R contributed to the study concept and study design. Wei C and Xu X defined the intellectual content. Wei C., Xue R, Gao Z. and Zhu H. acquired the data. Gao Z. and Zhu H. did the literature research and analyzed the data. Wei C prepared and edited the paper. Xue R, Gao Z and Xu X. reviewed the paper. All authors have read and approved the final paper.

Acknowledgments. This work has benefited from GEO. We thank the GEO network for its generous sharing of large amounts of data.

Data availability. The datasets presented in this study can be found in GEO repositories (accession number: GSE72267, GSE22491).

Funding. This study was supported by Huai'an Municipal Health Commission Project(No. HAWJZ2024004)

Conflicts of interest. All authors declare that they have no conflicts of interest in this work.

Ethics approval. Not applicable.

Informed consent. Not applicable.

References

Álvarez-Luquín, D. D, Arce-Sillas, A., Leyva-Hernández, J., Sevilla-Reyes, E., Boll, M. C., Montes-Moratilla, E., Vivas-Almazán, V., Pérez-Correa, C., Rodríguez-Ortiz, U., Espinoza-Cárdenas, R., Fragoso, G., Sciutto, E., & Adalid-Peralta, L. (2019). Regulatory impairment in untreated Parkinson's disease is not restricted to Tregs: Other regulatory populations are also involved. *Journal of Neuroinflammation*, *16*, 212. https://doi.org/10.1186/s12974-019-1606-1

Armstrong, M. J., & Okun, M. S. (2020) Diagnosis and treatment of Parkinson disease: A review. *JAMA*, *32*, 548–560. https://doi.org/10.1001/jama.2019.22360

Atashrazm, F., Hammond, D., Perera, G., Dobson-Stone, C., Mueller, N., Pickford, R., Kim, W. S., Kwok, J. B., Lewis, S. J. G., Halliday, G. M., & Dzamko, N. (2018). Reduced glucocerebrosidase activity in monocytes from patients with Parkinson's disease. *Scientific Reports*, *8*, 15446. https://doi.org/10.1038/s41598-018-33921-x

Blauwendraat, C., Nalls, M. A., & Singleton, A. B. (2020). The genetic architecture of Parkinson's disease. *Lancet Neurology*, *19*, 170–178. https://doi.org/10.1016/S1474-4422(19)30287-X

Boutet, A., Madhavan, R., Elias, G. J. B., Joel, S. E., Gramer, R., Ranjan, M., Paramanandam, V., Xu, D., Germann, J., Loh, A., Kalia, S. K., Hodaie, M., Li, B., Prasad, S., Coblentz, A., Munhoz, R. P., Ashe, J., Kucharczyk, W., Fasano, A., & Lozano, A. M. (2021). Predicting optimal deep brain stimulation parameters for Parkinson's disease using functional MRI and machine learning. *Nature Communications*, *2,* 3043. https://doi.org/10.1038/s41467-021-23311-9

Brochard, V., Combadière, B., Prigent, A., Laouar, Y., Perrin, A., Beray-Berthat, V., Bonduelle, O., Alvarez-Fischer, D., Callebert, J., Launay, J. M., Duyckaerts, C., Flavell, R. A., Hirsch, E. C., & Hunot, S. (2009). Infiltration of CD4+ lymphocytes into the brain contributes to neurodegeneration in a mouse model of Parkinson disease. *Journal of Clinical Investigation*, *119*, 182–192. https://doi.org/10.1172/JCI36470

Calligaris, R., Banica, M., Roncaglia, P., Robotti, E., Finaurini, S., Vlachouli, C., Antonutti, L., Iorio, F., Carissimo, A., Cattaruzza, T., Ceiner, A., Lazarevic, D., Cucca, A., Pangher, N., Marengo, E., di Bernardo, D., Pizzolato, G., & Gustincich, S. (2015). Blood transcriptomics of drug-naive sporadic Parkinson's disease patients. *BMC Genomics*, *16*, 876. https://doi.org/10.1186/s12864-015-2058-3

Chowdhary, C. L., Khare, N., Patel, H., Koppu, S, Kaluri, R., & Rajput, D. S. (2022). Past, present and future of gene feature selection for breast cancer classification ¾ A survey. *International Journal of Engineering Systems Modelling and Simulation*, *13,* 140–153. https://doi.org/10.1504/IJESMS.2022.10047821

Chu, Y., Dodiya, H., Aebischer, P., Olanow, C. W., & Kordower, J. H. (2009). Alterations in lysosomal and proteasomal markers in Parkinson's disease: Relationship to alpha-synuclein inclusions. *Neurobiology of Disease*, *35*, 385–398. https://doi.org/10.1016/j.nbd.2009.05.023.

Danaher, P., Warren, S., Dennis, L., D'Amico, L., White, A., Disis, M. L., Geller, M. A., Odunsi, K., Beechem, J., & Fling, S. P. (2017). Gene expression markers of tumor infiltrating leukocytes. *Journal for ImmunoTherapy of Cancer*, *5,* 18. https://doi.org/10.1186/s40425-017-0215-8

Dickson, D. W. (2012). Parkinson's disease and parkinsonism: Neuropathology. *Cold Spring Harbor Perspectives in Medicine*, *2*, a009258. https://doi.org/10.1101/cshperspect.a009258.

Doorn, K. J., Moors, T., Drukarch, B., van de Berg, W. D. J., Lucassen, P. J., & van Dam, A. M. (2014). Microglial phenotypes and toll-like receptor 2 in the substantia nigra and hippocampus of incidental Lewy body disease cases and Parkinson's disease patients. *Acta Neuropathologica Communications*, *2,* 90. https://doi.org/10.1186/s40478-014-0090-1

Dulski, J., Uitti, R. J., Ross, O. A., & Wszolek, Z. K. (2022). Genetic architecture of Parkinson's disease subtypes ¾ Review of the literature. *Frontiers in Aging Neuroscience*, *14*, 1023574. https://doi.org/10.3389/fnagi.2022.1023574

GBD 2016 Parkinson's Disease Collaborators. (2018). Global, regional, and national burden of Parkinson's disease, 1990-2016: A systematic analysis for the Global Burden of Disease Study 2016. *Lancet Neurology*, *17*, 939–953. https://doi.org/10.1016/S1474-4422(18)30295-3

Herrero, M. T., Estrada, C., Maatouk, L., & Vyas, S. (2015). Inflammation in Parkinson's disease: Role of glucocorticoids. *Frontiers in Neuroanatomy*, *9,* 32. https://doi.org/10.3389/fnana.2015.00032

Horvath, S., & Ritz B. R. (2015). Increased epigenetic age and granulocyte counts in the blood of Parkinson's disease patients. *Aging*, *7,* 1130–1142. https://doi.org/10.18632/aging.100859

Kedmi, M., Bar-Shira, A., Gurevich, T., Giladi, N., & Orr-Urtreger, A. (2011). Decreased expression of B cell related genes in leukocytes of women with Parkinson's disease. *Molecular Neurodegeneration*, *6,* 66. https://doi.org/10.1186/1750-1326-6-66

Kumar, V., Lalotra, G. S., Sasikala, P., Rajput, D. S., Kaluri, R., Lakshmanna, K., Shorfuzzaman, M., Alsufyani, A., & Uddin, M. (2022). Addressing binary classification over class imbalanced clinical datasets using computationally intelligent techniques. *Healthcare*, *10*, 1293. https://doi.org/10.3390/healthcare10071293

Kustrimovic, N., Comi, C., Magistrelli, L., Rasini, E., Legnaro, M., Bombelli, R., Aleksic, I., Blandini, F., Minafra, B., Riboldazzi, G., Sturchio, A., Mauri, M., Bono, G., Marino, F., & Cosentino, M. (2018). Parkinson's disease patients have a complex phenotypic and functional Th1 bias: Cross-sectional studies of CD4+ Th1/Th2/T17 and Treg in drug-naive and drug-treated patients. *Journal of Neuroinflammation*, *15*, 205. https://doi.org/10.1186/s12974-018-1248-8

Li, X., Sundquist, J., & Sundquist, K. (2012). Subsequent risks of Parkinson disease in patients with autoimmune and related disorders: A nationwide epidemiological study from Sweden. *Neurodegenerative Diseases*, *10,* 277–284. https://doi.org/10.1159/000333222

Lin, C. H., Chen, C. C., Chiang, H. L., Liou, J. M., Chang, C. M., Lu, T. P., Chuang, E. Y., Tai, Y. C., Cheng, C., Lin, H Y., & Wu, M. S. (2019). Altered gut microbiota and inflammatory cytokine responses in patients with Parkinson's disease. *Journal of Neuroinflammation*, *16,* 129. https://doi.org/10.1186/s12974-019-1528-y

Lu, X., Li, L., Suo, L., Huang, P., Wang, H., Han, S., & Cao, M. (2022). Single-cell RNA sequencing profiles identify important pathophysiologic factors in the progression of diabetic nephropathy. *Frontiers in Cell and Developmental Biology*, *10*, 798316. https://doi.org/10.3389/fcell.2022.798316

Mahul-Mellier, A. L., Burtscher, J., Maharjan, N., Weerens, L., Croisier, M., Kuttler, F., Leleu, M., Knott, G. W., & Lashuel, H. A. (2020). The process of Lewy body formation, rather than simply alpha-synuclein fibrillization, is one of the major drivers of neurodegeneration. *Proceedings of the National Academy of Sciences of the United States of America*, *117*, 4971–4982. https://doi.org/10.1073/pnas.1913904117

Marsili, L., Rizzo, G., & Colosimo, C. (2018). Diagnostic criteria for Parkinson's disease: From James Parkinson to the concept of prodromal disease. *Frontiers in Neurology*, *9,* 156. https://doi.org/10.3389/fneur.2018.00156

Matheoud, D., Sugiura, A., Bellemare-Pelletier, A., Laplante, A., Rondeau, C., Chemali, M., Fazel, A., Bergeron, J. J., Trudeau, L. E., Burelle, Y., Gagnon, E., McBride, H. M., & Desjardins, M. (2016). Parkinson's disease-related proteins PINK1 and parkin repress mitochondrial antigen presentation. *Cell*, *166*, 314–327. https://doi.org/10.1016/j.cell.2016.05.039

Mu, J., Chaudhuri, K. R., Bielza, C., de Pedro-Cuesta, J., Larrañaga, P., & Martinez-Martin, P. (2017). Parkinson's disease subtypes identified from cluster analysis of motor and non-motor symptoms. *Frontiers in Aging Neuroscience*, *9,* 301. https://doi.org/10.3389/fnagi.2017.00301

Mutez, E., Larvor, L., Leprêtre, F., Mouroux, V., Hamalek, D., Kerckaert, J. P., Pérez-Tur, J., Waucquier, N., Vanbesien-Mailliot, C., Duflot, A., Devos, D., Defebvre, L., Kreisler, A., Frigard, B., Destée, A., & Chartier-Harlin, M. C. (2011). Transcriptional profile of Parkinson blood mononuclear cells with LRRK2 mutation. *Neurobiology of Aging*, *32,* 1839–1848. https://doi.org/10.1016/j.neurobiolaging.2009.10.016

Niwa, Y., Kanda, H., Shikauchi, Y., Saiura, A., Matsubara, K., Kitagawa, T., Yamamoto, J., Kubo, T., & Yoshikawa, H. (2005). Methylation silencing of SOCS-3 promotes cell growth and migration by enhancing JAK/STAT and FAK signalings in human hepatocellular carcinoma. *Oncogene*, *24,* 6406–6417. https://doi.org/10.1038/sj.onc.1208788

Okuno, T., Nakatsuji, Y., Kumanogoh, A., Moriya, M., Ichinose, H., Sumi, H., Fujimura, H., Kikutani, H., & Sakoda, S. (2005). Loss of dopaminergic neurons by the induction of inducible nitric oxide synthase and cyclo-oxygenase-2 via CD 40: Relevance to Parkinson's disease. *Journal of Neuroscience Research*, *81,* 874–882. https://doi.org/10.1002/jnr.20599

Qi, X., Davis, B., Chiang, Y. H., Filichia, E., Barnett, A., Greig, NH., Hoffer, B., & Luo, Y. (2016). Dopaminergic neuron-specific deletion of p53 gene is neuroprotective in an experimental Parkinson's disease model. *Journal of Neurochemistry*, *138*, 746–757. https://doi.org/10.1111/jnc.13706

Rivas, M. A., Avila, B. E., Koskela, J., Huang, H., Stevens, C., Pirinen, M., Haritunians, T., Neale, B. M., Kurki, M., Ganna, A., Graham, D., Glaser, B., Peter, I., Atzmon, G., Barzilai, N., Levine, A. P., Schiff, E., Pontikos, N., Weisburd, B., … Daly, M. J. (2018). Insights into the genetic epidemiology of Crohn's and rare diseases in the Ashkenazi Jewish population. *PLoS Genetics*, *14,* e1007329. https://doi.org/10.1371/journal.pgen.1007329

Sarkar, S., Dammer, E. B., Malovic, E., Olsen, A. L., Raza, S. A., Gao, T., Xiao, H., Oliver, D. L., Duong, D., Joers, V., Seyfried, N., Huang, M, Kukar, T, Tansey, M. G., Kanthasamy, A. G., & Rangaraju, S. (2020). Molecular signatures of neuroinflammation induced by alphaSynuclein aggregates in microglial cells. *Frontiers in Immunology*, *11,* 33. https://doi.org/10.3389/fimmu.2020.00033

Schwab, A. D., Thurston, M. J., Machhi, J., Olson, K. E., Namminga, K. L., Gendelman, H. E., & Mosley, R. L. (2020). Immunotherapy for Parkinson's disease. *Neurobiology of Disease*, *137*, 104760. https://doi.org/10.1016/j.nbd.2020.104760

Selvaraj, N. P., Paulraj, S., Ramadass, P., Kaluri, R., Shorfuzzaman, M., Alsufyani, A., & Uddin, M. (2022). Exposure of botnets in cloud environment by expending trust model with CANFES classification approach. *Electronics*, *11,* 2350. 10.3390/electronics11152350

Stefanis, L., Emmanouilidou, E., Pantazopoulou, M., Kirik, D., Vekrellis, K., & Tofaris, G. K. (2019). How is alpha-synuclein cleared from the cell? *Journal of Neurochemistry*, *150,* 577–590. https://doi.org/10.1111/jnc.14704

Stevens, C. H., Rowe, D., Morel-Kopp, M. C., Orr, C., Russell, T., Ranola, M., Ward, C., & Halliday, G. M. (2012). Reduced T helper and B lymphocytes in Parkinson's disease. *Journal of Neuroimmunology*, *252,* 95–99. https://doi.org/10.1016/j.jneuroim.2012.07.015

Sommer, A., Marxreiter, F., Krach, F., Fadler, T., Grosch, J., Maroni, M., Graef, D., Eberhardt, E., Riemenschneider, M. J., Yeo, G. W., Kohl, Z., Xiang, W., Gage, F. H., Winkler, J., Prots, I., & Winner, B. (2018). Th17 lymphocytes induce neuronal cell death in a human iPSC-based model of Parkinson's disease. *Cell Stem Cell*, *23*, 123–131.e6. https://doi.org/10.1016/j.stem.2018.06.015

Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society Series B ¾ Methodological*, *58*, 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Vinagre-Aragón, A., Campo-Caballero, D., Mondragón-Rezola, E., Pardina-Vilella, L., Hernandez Eguiazu, H., Gorostidi, A., Croitoru, I., Bergareche, A., & Ruiz-Martinez, J. (2021). A more homogeneous phenotype in Parkinson's disease related to R1441G mutation in the LRRK2 gene. *Frontiers in Neurology*, *12*, 635396. https://doi.org/10.3389/fneur.2021.635396

Wang, C., Chen, L., Yang, Y., Zhang, M., &Wong, G. (2019). Identification of potential blood biomarkers for Parkinson's disease by gene expression and DNA methylation data integration analysis. *Clinical Epigenetics*, *11*, 24. https://doi.org/10.1186/s13148-019-0621-5

Wijeyekoon, R. S., Kronenberg-Versteeg, D., Scott, K. M., Hayat, S., Jones, J. L., Clatworthy, M. R., Floto, R. A., Barker, R. A., & Williams-Gray, C. H. (2018). Monocyte function in Parkinson's disease and the impact of autologous serum on phagocytosis. *Frontiers in Neurology*, *9,* 870. https://doi.org/10.3389/fneur.2018.00870

Wolfrum, P., Fietz, A., Schnichels, S., & Hurst, J. (2022). The function of p53 and its role in Alzheimer's and Parkinson's disease compared to age-related macular degeneration. *Frontiers in Neuroscience*, *16*, 1029473. https://doi.org/10.3389/fnins.2022.1029473

Zhao, S., Zhang, L., Ji, W., Shi, Y., Lai, G., Chi, H., Huang, W., & Cheng, C. (2022). Machine learning-based characterization of cuprotosis-related biomarkers and immune infiltration in Parkinson's disease. *Frontiers in Genetics*, *13,* 1010361. https://doi.org/10.3389/fgene.2022.1010361

Zhu, B., Yin, D., Zhao, H., & Zhang, L. (2022). The immunology of Parkinson's disease. *Seminars in Immunopathology*, *44*, 659–672. https://doi.org/10.1007/s00281-022-00947-3

https://doi.org/10.1017/thg.2025.10008 Published online by Cambridge University Press