International Journal of General Medicine

Dovepress

open access to scientific and medical research

Open Access Full Text Article

ORIGINAL RESEARCH

Integrated Bioinformatics Analysis for Revealing CBL is a Potential Diagnosing Biomarker and Related Immune Infiltration in Parkinson's Disease

Yanchen Chen[1,2], Yuqin Tu[2,3], Guiling Yan[2,3], Xinyao Ji[2,3], Shu Chen[2,3], Changchun Niu[2], Pu Liao[1,2]

[1]Department of Laboratory Medicine, North Sichuan Medical College, Nanchong, People's Republic of China; [2]Department of Clinical Laboratory, Chongqing General Hospital, Chongqing, People's Republic of China; [3]Chongqing Medical University, Chongqing, People's Republic of China

Correspondence: Pu Liao, Chongqing General Hospital, 118 Xingguang Avenue, Yubei District, Chongqing, People's Republic of China, Tel +86 13628358003, Fax +23-63519127, Email liaopu@sina.com

Purpose: There is growing evidence that the immune system plays an important role in the progression of Parkinson's disease, the second most common neurodegenerative disorder. This study aims to address the comprehensive understanding of the immunopathogenesis of Parkinson's disease and explore new inflammatory biomarkers.

Patients and Methods: In this study, Immune-related differential expressed genes (DEIRGs) were obtained from GEO database and Immport database. The hub gene was screened in DEIRGs using LASSO regression and random forest algorithm, and the mRNA expression of the identified hub gene was validated using clinical blood samples.

Results: We obtained a total of 157 DEIRGs that played an important role in the immune response. The results of immune cell infiltration analysis showed that the degree of memory B cells infiltration was higher in PD patients, while the degree of Monocytes, resting mast cells and M0 macrophages infiltration was lower ($p<0.05$). A total of 8 hub genes were screened by machine learning methods, and RT-PCR results showed that the expression level of CBL gene in PD was significantly increased ($p<0.05$).

Conclusion: Our findings suggest that CBL is a new potential diagnostic biomarker for PD and that abnormal immune cell infiltration may influence PD development.

Keywords: Parkinson's disease, immune cell infiltration, bioinformatics, immport database, gene expression omnibus database

Introduction

As a neurodegenerative disease, Parkinson's disease(PD) is characterized by degeneration of dopaminergic neurons in the substantia nigra and accumulation of large amounts of alpha-synuclein (α-syn), culminating in the formation of Lewy bodies and Lewy neurites.[1] Parkinson's disease is distinguished by motor manifestations such as bradykinesia, rigidity, and resting tremor. Furthermore, it encompasses a multitude of non-motor symptoms, such as constipation, depression, cognitive decline, and disturbances in sleep patterns.[2] Currently, the etiology of Parkinson's disease remains elusive, the early diagnosis of PD mainly depends on the evaluation of clinical signs (bradykinesia, static tremor or stiffness, etc.), and the misdiagnosis rate is high, and the comprehensive accuracy rate of PD diagnosis is only about 80%.[3,4] Therefore, the timely detection of molecular biomarkers is crucial for intervention prior to the manifestation of the condition.[5]

At present, Parkinson's disease is recognized as a complex condition affecting multiple bodily systems, characterized by pronounced neuroinflammation and impaired immune function.[1] Neuroinflammation is an important pathological feature of PD, promoted by both the innate immune system and the adaptive immune system. Among them, the innate immune system plays a crucial role in the neuroinflammation of PD.[6] A study found altered expressions of immune cells such as microglia and astrocytes in the brain of PD patients. The inflammatory mediators secreted by microglia stimulate astrocytes, jointly amplifying inflammatory signals and promoting neuroinflammation.[7] In Parkinson's disease, immune responses are not limited to alterations in brain immune cells, they also encompass modifications in the peripheral

Received: 1 February 2024
Accepted: 16 May 2024
Published: 22 May 2024

immune system.[8] Peripheral immune cells are difficult to detect in the central nervous system due to the presence of the blood-brain barrier(BBB) in healthy conditions.[9] The pathological advancement of Parkinson's disease can disrupt the integrity of the BBB, leading to increased penetration of peripheral immune cells into the central nervous system.[10] The increased infiltration of the peripheral immune system may be associated with increased neurodegeneration associated with Parkinson's disease, resulting in microgliosis, oxidative stress, and cytotoxicity.[11] There is evidence indicating that increased levels of CD3+, CD4+, and CD8+ T cells, which are part of the adaptive immune system, were observed in the substantia nigra pars compacta (SNpc) of the brain tissue of PD patients.[12] Furthermore, changes in peripheral immune cells have also been observed in the blood of PD patients, with significant alterations in T cells, B cells, and NK cells compared to the control group.[13] Hence, it is plausible that the peripheral immune system plays a significant role in the progression of Parkinson's disease. Recent studies have also shown that the component of immune cells in the blood can embody the early stages of Parkinson's disease progression, which can be identified to detect and diagnose the disease earlier.[14] This provides an opportunity to explore new immunotherapeutic strategies.[15]

Changes in gene expression levels can indicate various disease progressions and high-throughput sequencing served as a useful tool for studying disease-related gene variations and identifying potential disease-associated genes, thus contributing to the discovery of new diagnostic and therapeutic methods.[16] Machine learning can be employed to evaluate high-dimensional transcriptomic data and locate genes with biological significance. It is often combined with high-throughput microarray analysis to identify novel diagnostic biomarkers.[17]

In this study, we obtained relevant pertinent datasets from the GEO database and Immport database, used bioinformatics to identify immune-related differential expressed genes (DIREGs) associated with PD and performed functional analysis on them. Furthermore, machine learning was applied to identify immune-related biomarkers. We used Cibersortx algorithm to perform immune infiltration analysis on the data set, found the difference between PD and HC immune cell infiltration, and discussed the correlation between central genes and immune cells. Ultimately, clinical samples of PD and HC were collected for RT-PCR to verify hub gene expression, not only highlighted the importance of immune cells in Parkinson's disease but also provided potential insights for the discovery of new therapeutic targets.

Materials and Methods

Data Collection

We used GSE22491 and GSE49126 in the GEO database as training sets and validation sets. GSE22491 dataset was a transcriptomic analysis of peripheral blood mononuclear cells (PBMCs) from 10 PD patients and 8 healthy controls. GSE49126 dataset also was a transcriptomic profiling of PBMCs from 30 PD patients and 20 healthy controls. The Immunology Database and Analysis Portal database (ImmPort) provided a list of 2483 immune-related genes (IRGs) to help us conduct subsequent immune-related gene analysis.

Identification of Immune-Related Differential Genes

Principal component analysis (PCA) is a commonly used method for data dimension reduction, often applied to analyze gene expression microarray data. It facilitates a preliminary understanding of the overall gene expression differences among samples of different groups, as well as the magnitude of variation among samples within each group. We used PCA to observe the distribution between PD and HC by FactoMineR package in R. limma package is an R software package developed for microarray data and RNA-Seq data. It provided a set of statistical models and tools to identify differentially expressed genes. We used limma package (version 3.52.2) to perform statistical analysis on the data set and evaluate the differential expression of the two groups of genes. We defined differentially expressed genes (DEGs) as those with p-values less than 0.05 and those with absolute values of logFC greater than 0.5. The ggplot2 package (version 3.3.6) was used to plot volcano plots to visually represent the differential genes of the two groups, and the heat map package (version 1.0.12) was used to plot the difference in the expression content of differential genes between the two groups. The DEGs were compared with IRGs using a Venn Diagram to identify immune-related differential genes (DEIRGs), which were subsequently utilized for further analysis.

Gene Set Enrichment Analysis

GSEA is an enrichment method based on gene sets. It first sorts genes according to the degree of their association with phenotypic data. Then, it determines whether the genes within each gene set are enriched in the upper or lower part of the gene list sorted by phenotypic correlation, thereby assessing the impact of coordinated changes in the genes within the set on phenotypic variation.[18] WebGestalt (WEB-based Gene SeT AnaLysis Toolkit) is an online website that can perform functional enrichment analysis.[19] We inputed the gene names of DEGs and their corresponding logFC values into this website to perform GSEA analysis. We ranked the categories based on FDR, and then selected the top 10 most important categories from each positively and negatively correlated group for presentation.

Functional and Pathway Enrichment Analysis of DEIRGs

We used GO and KEGG enrichment analysis to integrate functional information related to each DEIRG, and the ClusterProfiler package in R (version 4.4.4) was used for analysis to explore the functions and molecular mechanisms of DEIRGs in biology. p value<0.05 was considered a meaningful enrichment pathway. GO functional enrichment analysis displayed the top 5 most significantly enriched functions for each category presented, while KEGG analysis exhibited the top 10 most significantly enriched pathways.

Identification of Key Genes with Machine Learning

LASSO is a regression method used to select variables to improve prediction accuracy.[20] The degree of complexity adjustment in LASSO regression is controlled by the parameter λ. The larger the value of λ, the greater the penalty for linear models with more variables, resulting in fewer final variables and more representative hub genes.[21] We used the "glmnet" package (version 4.1–6) in R to perform LASSO regression analysis on DEIRGs. Through 10-fold cross-validation, we determined the optimal value of lambda and identified 15 hub genes using the lambda.1se model. The random forest model calculates the average error rate of candidate hub genes to determine the optimal number of variables, which is utilized to predict continuous variables based on the identified hub genes.[22] We used "randomForest" (version 4.7–1.1) in R for RF analysis. "Mean Decrease Accuracy" represents the degree of reduction in the prediction accuracy of the random forest. A higher value indicates a greater importance of the variable. "Mean Decrease Gini" calculates the impact of each variable on the heterogeneity of the observed values at each node of the classification tree, allowing for a comparison of the importance of different variables. A higher value also indicates a greater importance of the variable. Based on the Mean decrease gini indicator, we selected the top 30 important variables as hub genes. After obtaining the expression levels of differentially expressed genes in the two groups, these two machine learning methods can be used to screen potential key genes. Ultimately, genes that overlapped among the two machine learning algorithms were regarded as key genes.

Analysis of Immune Cells Infiltration in PD and Controls

Cibersortx is an algorithm utilized for deconvolution analysis, which allows for the assessment of cellular abundance and the identification of cell type-specific gene expression patterns based on batch tissue transcriptome profiles.[23] Cibersortx builded a specialized knowledge of gene expression signatures for the deconvolution of interesting cell types to characterize various cell types in complex tissues.[24] The LM22 (leukocyte signature Matrix) is a signature matrix composed of 22 functionally defined human hematopoietic cell subsets. Cibersortx can use the LM22 matrix to calculate the corresponding gene expression profile of these 22 immune cells in patients and estimate their relative abundance.[25] We used this algorithm to calculate immune cells abundance between the two groups in the GSE22491 dataset, and further analyzed the expression differences between them. The analysis using Cibersort package in R. Only Cibersort p value<0.05 was selected as worthy of further analysis. Spearman correlation analysis was used to evaluate the linear correlation between immune cells and gene expression. The ggplot2 package (version 3.3.6) is a powerful graphical visualization R package that we use to display these results.

Sample Collection, RNA Extraction, and Reverse Transcription

The study content and process adhered to international and national ethical requirements for biomedical research. This study was approved by the Ethics Committee of Chongqing General Hospital (Ethics Review number KY S2023-010-01 of 13/03/2023). In our study, we collected blood samples from 28 patients with PD patients and 16 healthy controls (obtained from discarded laboratory testing blood samples). The PD patients were diagnosed by at least two professionally trained neurologists. The severity of the disease was evaluated using the modified Hoehn and Yahr (H&Y) scale. According to the Hoehn-Yahr staging scale for PD patients, PD patients with an H-Y staging of ≤2 were considered as the early stage group, while those with an H-Y staging of >2 were considered as the middle to late stage group. The exclusion criteria were as follows: patients with tumors or a history of anticancer treatment; severe liver and kidney dysfunction; other neurological diseases (such as Alzheimer's disease, Huntington's disease, etc.); stroke and other cerebrovascular diseases. The healthy participants were all from the health examination center. Detailed information about the patients and healthy controls is summarized in Supplementary Table 1.

We collected 2 mL of fasting EDTA anticoagulated blood from each participant, centrifuged the whole blood at 450×g for 10 minutes at room temperature, took the bottom sediment into a free-Nase EP tube, and then added three times the volume of Red Blood Cell Lysis Buffer (Solarbio, Beijing, China). After shaking and mixing, we let it rest at 4 °C for 10 minutes. We then centrifuged it at 450×g for 10 minutes, aspirated and discarded the supernatant, added twice the volume of Red Blood Cell Lysis Buffer, shook and mixed it again, aspirated and discarded the supernatant, and finally centrifuged it at 450×g for 10 minutes to obtain PBMCs.

We used the RNAiso Plus reagent (Takara Bio, Japan) to extract RNA from the PBMCs. The absorption values of the RNA solution at 260nm and 280nm were detected using the Nano-300 micro-spectrophotometer (ALLSHENG, Hangzhou, China) to calculate the extracted RNA's purity and concentration. Subsequently, we used the PrimeScript™ RT reagent Kit and gDNA Eraser (Takara Bio, Japan) to incubate the total RNA at 37°C for 15 minutes and heat it at 85°C for 5 seconds to terminate the reaction. This step eliminated genomic RNA from the total RNA and synthesized cDNA. The resulting cDNA was then stored at −80°C until it was ready for further PCR reactions.

Quantitative Real-Time Polymerase Chain Reaction

We performed qPCR amplification on the Applied Biosystems® QuantStudio®5 real-time PCR system using the TB Green® Premix Ex Taq™ II kit from Takara Bio, Japan. Table 1 shows the specific sequence of primers for amplification. 2-ΔΔCT is a method to obtain the relative expression level of the target gene by using internal parameters to homogenize, and we choose this method to calculate the expression of related genes.

Statistical Analysis

According to whether the data follow the normal distribution, we choose the appropriate statistical analysis method for comparison. If the data follows a normal distribution, pairwise comparisons can be performed using an independent Student's *t*-test. If the data do not follow a normal distribution, we consider using the Mann–Whitney *U*-test for pairwise comparison. Receiver Operating Characteristic (ROC) curve analysis is a graphical tool that represents the performance of a classification model, combining sensitivity and specificity graphically to assess the accuracy of the model's predictions. Area under the curve (AUC) is a measure of diagnostic accuracy. The closer the AUC is to 1, the better the predictive performance of the model. Therefore, it can be used to evaluate the diagnostic significance of key genes in PD patients. When the P-value<0.05, we consider that there is a statistical difference between the two groups. Both R language (version 4.2.1) and GraphPad Prism (version 9.0.0) are powerful statistical analysis software and scientific research drawing tools. We use these to analyze statistical differences and draw results.

Table 1 Primer Sequence for Hub Genes

Gene	Forward Primer (5'->3')	Reverse Primer (5'->3')
CBL	TCTACATGAAGTGCATCCCATCA	AAGAGGACCAGGGCTGAAAGA
GAPDH	ATGGGGAAGGTGAAGGTCGG	CCTGGAAGATGGTGATGGGATT

Results

Identification of Immune-Related Differential Genes

GSE22491 was subjected to principal component analysis, as shown in Figure 1a, the result demonstrated a significant separation between PD samples and controls PC1 (25.5%) and PC2 (16.2%). A total of 1794 DEGs were identified in GSE22491, of which 879 up-regulated genes and 915 down-regulated genes were detected (Figure 1b). The top 20 DEGs with the most significant up-regulation and down-regulation were selected for heat map drawing (Figure 1c). There were 2483 IRGs in the Immport database, and 157 DEIRGs were identified through the intersection of DEGs and IRGs (Figure 1d).

Immune-Related Pathways Were Significantly Enriched in PD

We conducted a GSEA enrichment analysis on PD and HC to explore their biological signaling pathways. We screened the 10 most significantly upregulated and downregulated pathways of DEGs on the Reactome pathway. Pathways such as Developmental Biology, G alpha (s) signaling events, and Keratinization were significantly upregulated in PD patients, while Neutrophil degranulation, Innate Immune System, and Immune System were significantly down-regulated (Figure 2a). Among these, the inflammation-related pathways Neutrophil degranulation (FDR<2.2e-16, ES =−0.63904, NES =−3.7324), Innate Immune System (FDR<2.2e-16, ES=−0.51757, NES =−3.3782), and Immune System

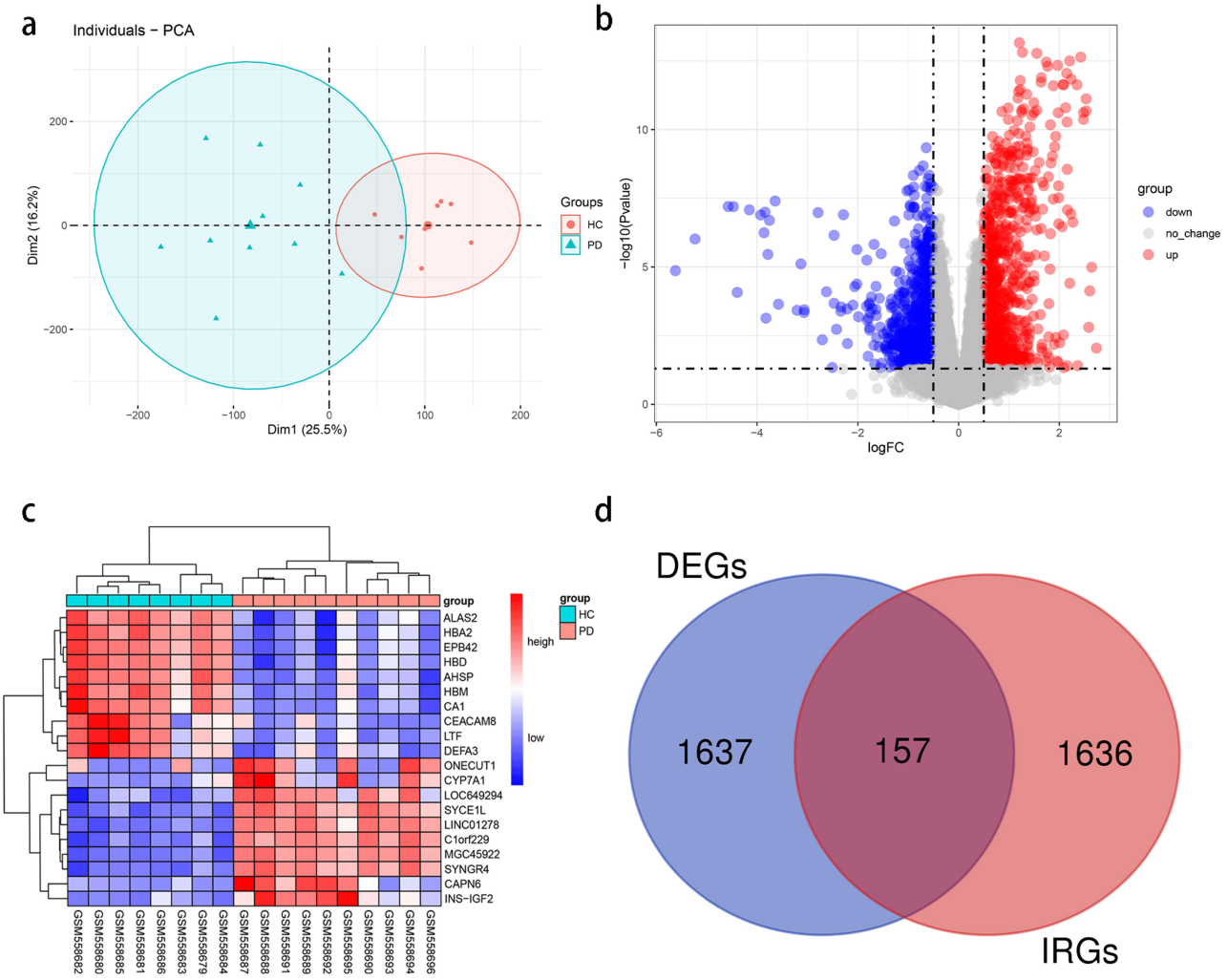

Figure 1 Differential gene analysis of the GSE22491. (**a**) Plot representing principal component analysis of GSE22491. The red dots represent samples from the healthy control group, while the blue triangles represent samples from the Parkinson's disease group. The ellipses indicate the 95% confidence intervals for each group, and the points outside the circles are not statistically significant; (**b**) Volcano plot of DEGs, blue indicates down-regulated genes, red indicates up-regulated genes; (**c**) Heat map of the top 10 most significantly up-regulated genes and the top 10 most significantly down-regulated genes, red represents up-regulation, blue represents down-regulation, "HC" represented control samples, and "PD" represented PD patients; (**d**) DEGs and IRGs of crossover genes, red represents DEGs and purple represents IRGs.
Abbreviations: DEG, differentially expressed genes; IRG, immune-related genes.

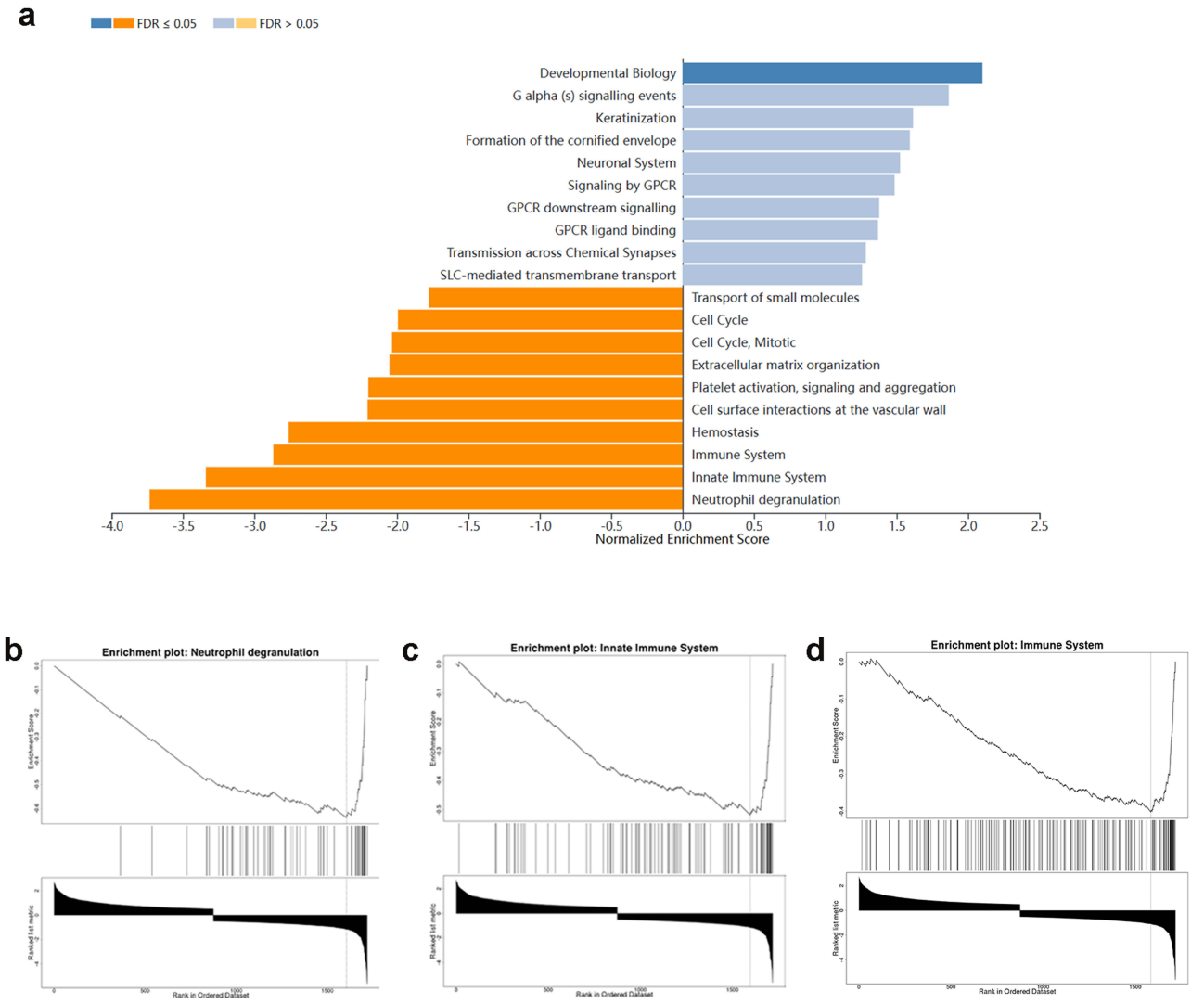

Figure 2 Gene set enrichment analysis of WebGestalt. (**a**) GSEA revealed the following Reactome pathways were enriched in PD group; Orange represents the negative enrichment correlation pathway, blue represents the positive enrichment correlation pathway, and the horizontal coordinate corresponds to the standardized enrichment score of the pathway. FDR, False Discovery Rate; NES, Normalized Enrichment Score, the enrichment score for the gene set after it had been normalized across analyzed gene sets. (**b**) Neutrophil degranulation; (**c**) Innate Immune System; (**d**) Immune System. The topmost section of the figure corresponds to the enrichment score (ES) values of the genes. As each gene is encountered from left to right, an ES value is calculated and plotted as a line. The peak value on the leftmost or rightmost side represents the ES value for the phenotypic profile of the gene set. The middle section of the figure features individual lines representing each gene within the gene set and its corresponding ranking position in the gene list. Genes with positive enrichment scores are those located on the left side of the peak, while genes with negative enrichment scores are those on the right side. The bottom section displays a matrix that illustrates the association between genes and phenotypes.

(FDR<2.2e-16, ES =−0.40179, NES =−2.9466) all exhibited significant down regulation (Figure 2b–d). The results indicate that immune factors may have a significant impact on PD.

Enrichment Function Analysis of DEIRGs

We performed GO enrichment analysis and KEGG pathway analysis in 157 DEIRGs. According to the GO enrichment analysis results, DEIRGs primarily impact immune-related biological functions such as cytokine-mediated signaling pathway, humoral immune response, cytokine activity and immune receptor activity (Figure 3a). In the KEGG enrichment analysis, DEIRGs are mainly involved in immune-related pathways, including cytokine-cytokine receptor interaction, chemokine signaling pathway, and T cell receptor signaling pathway (Figure 3b).

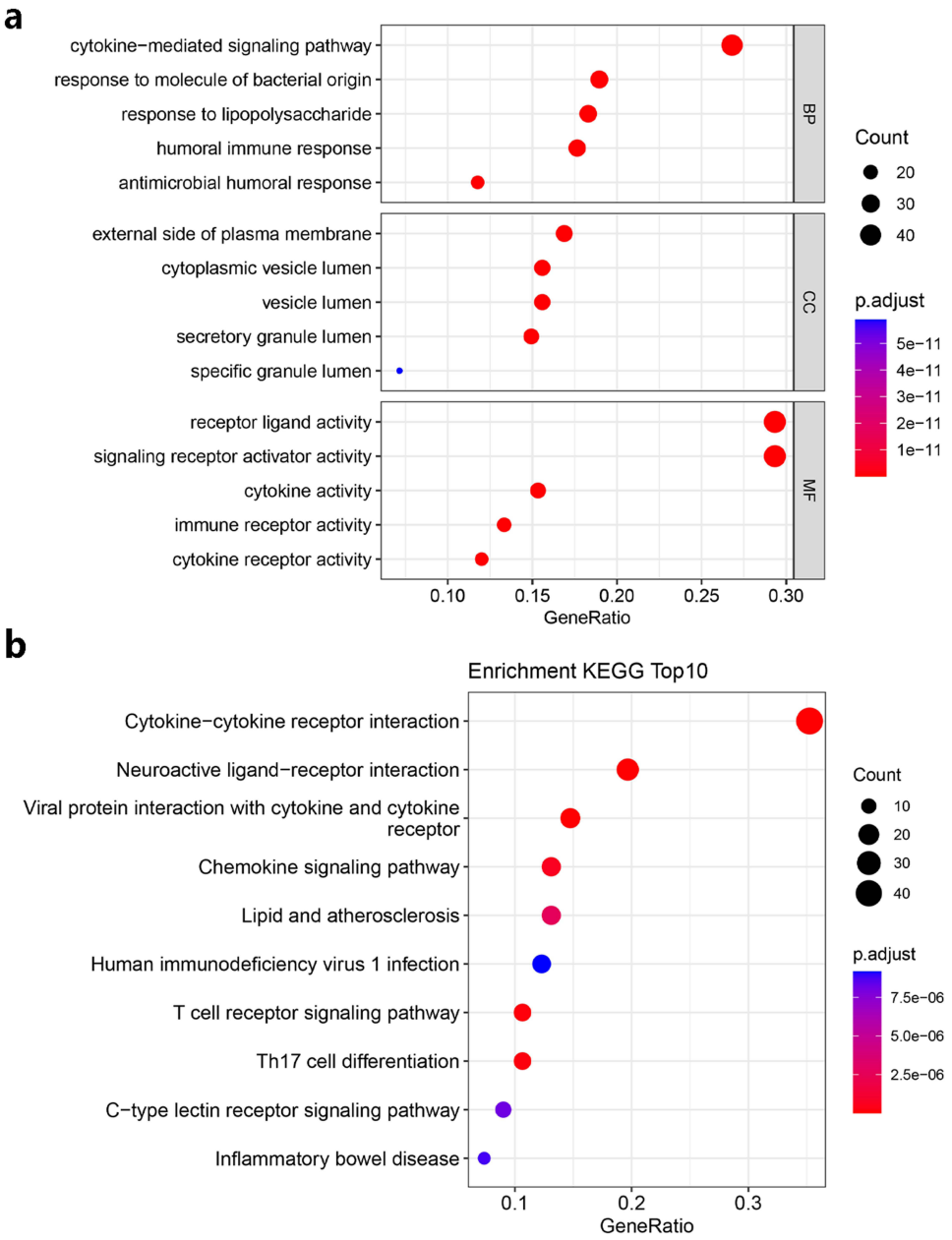

Figure 3 Results of GO and KEGG enrichment analysis of DEIRGs. (**a**) GO bubble plot; (**b**) KEGG pathway bubble plot. The red and blue dots represent corrected p-values, and the radius size of the dots represents gene counts.
Abbreviations: GO, Gene Ontology; BP, Biological Process; CC, Cellular Component; MF, Molecular Function; KEGG, Kyoto Encyclopedia of Genes and Genomes.

8 Key Genes Were Identified by Machine Learning

To study the characteristic genes involved in immune regulation in the development of Parkinson's disease, we utilized the information of 157 DEIRGs analyzed through LASSO and RF techniques to identify key genes of interest. The lambda.1se model in LASSO regression identifies 15 hub genes as the most representative genes related to immunity in PD. The coefficients of the 15 hub genes are presented in Supplementary Table 2. We obtained 15 genes from DEIRGs using LASSO logistic regression algorithm as potential biomarkers (Figure 4a and b). We selected the top 30 genes as candidate genes based on the Mean Decrease Gini index from the random forest algorithm (Figure 4c). By comprehensively considering the overlapping genes obtained from the two algorithms, we have identified the candidate biomarkers. NRAS, GALR3, CD3E, CBL, RELA, AVP, SEMA6B and OGFR were attained as biomarkers (Table 2).

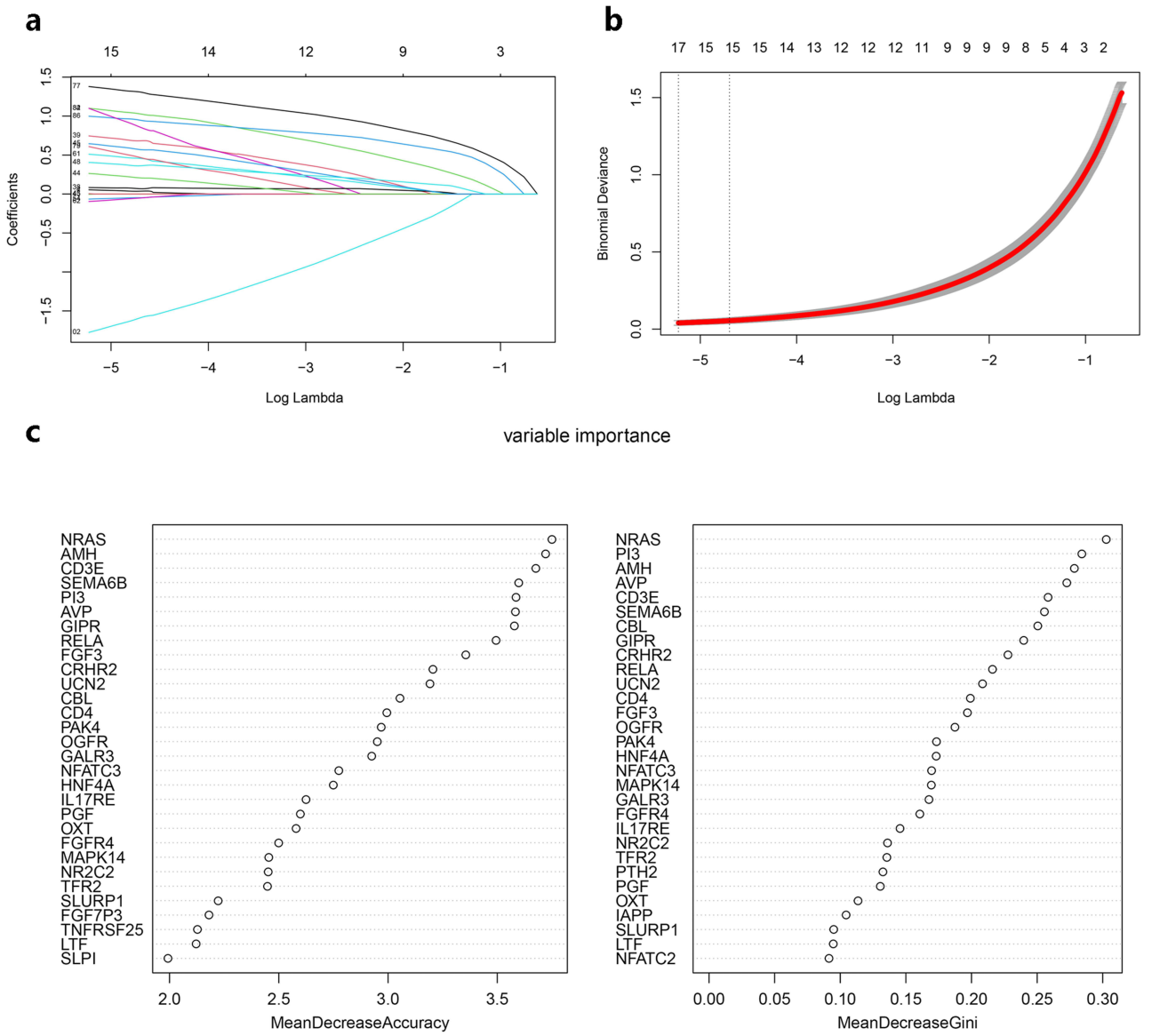

Figure 4 Machine learning screening for DEIRGs. (**a**) Path diagram of the LASSO coefficients for the 15 hub genes in PD. Each curve represents the trajectory of each hub gene, with the vertical axis indicating the coefficient of the gene. The lower horizontal coordinate is log(lambda), and the upper horizontal coordinate is the number of non-zero hub genes in the model at this time. As lambda varies, variables with coefficients that are compressed to 0 later are considered more important. (**b**) LASSO regression cross-validation curve. Optimal lambda values were determined using 10-fold cross-validation. There are two dashed lines in the figure, corresponding to lambda.min and lambda.1se from left to right. (**c**) RF algorithm.

Abbreviations: LASSO, least absolute shrinkage and selection operator; RF, Random Forest.

Abnormal Infiltration of Immune Cells in PD

We used the Cibersort algorithm to calculate the percentage of 22 different types of immune infiltrating cells in patients with PD and HC. The state of monocytes, M0 macrophages and resting mast cells in PD group was significantly lower than that in HC group, while the memory B cells was significantly higher ($p<0.05$) (Figure 5a). This suggests that there may be abnormal infiltration and heterogeneity of immune cells in the occurrence of PD, which has important clinical significance. Then, we studied the correlation between the expression of eight hub genes and the ratio of immune cells. Spearman correlation analysis showed that OGFR was negatively correlated with neutrophils, M0 macrophages and monocytes, and positively correlated with memory B cells. SEMA6B was significantly negatively correlated with M0 macrophages, monocytes, activated memory CD4T cells and naive B cells, and significantly positively correlated with activated NK cells and memory B cells. AVP was significantly negatively correlated with M0 macrophages and

Table 2 Hub Genes Selected by of RF and LASSO Analysis

Analysis	Gene
LASSO	IGKC, RBP5, CYSLTR2, EBI3, FABP5, IL4, CCR4, **NRAS, GALR3, CD3E, CBL, RELA, AVP, SEMA6B, OGFR**
RF	LTF, FGFR4, PI3, **CD3E**, SLURP1, NFATC2, GIPR, TFR2, FGF3, CD4, HNF4A, IL17RE, PGF, NFATC3, **NRAS**, MAPK14, IAPP, AMH, **RELA**, UCN2, PAK4, **AVP**, CRHR2, PTH2, OXT, NR2C2, **GALR3, CBL, SEMA6B, OGFR**

Note: *The genes in bold font are the same genes selected by the two algorithms.

neutrophils, and positively correlated with memory B cells. RELA was negatively correlated with M0 macrophages, resting dendritic cells and monocytes, and positively correlated with memory B cells and activated mast cells. CBL was negatively correlated with M0 macrophages, resting mast cells and monocytes, and positively correlated with γδT cells, memory B cells and activated NK cells. CD3E was negatively correlated with M0 macrophages, monocytes and resting mast cells, and positively correlated with activated NK cells, CD8 T cells and memory B cells. GALR3 was negatively correlated with neutrophils, resting mast cells and M0 macrophages, and positively correlated with γδT cells. NRAS was negatively correlated with memory B cells and positively correlated with resting mast cells, neutrophils, M0 macrophages and monocytes (Figure 5b).

The Diagnostic Efficacy of the Eight Hub Genes for PD

To further validate the accuracy of the biomarkers selected through machine learning methods, we used both training and validation datasets to assess their expression patterns and diagnostic values. In the training dataset, compared to the HC, the expression levels of AVP, CBL, CD3E, GALR3, RELA, OGFR and SEMA6B were significantly higher in PBMC and only NRAS were significantly lower ($p < 0.05$) in PD patients (Figure 6a–h). The expression levels of these eight hub genes effectively distinguished PD from HC. Subsequently, we employed ROC curves to evaluate the diagnostic performance of eight hub genes in the validation dataset for PD. The results showed that the AUC of AVP was 0.658 (95% CI: 0.505–0.811), CBL was 0.722 (95% CI: 0.580–0.864), CD3E was 0.542 (95% CI: 0.377–0.706), GALR3 was 0.522 (95% CI: 0.343–0.700), NRAS was 0.567 (95% CI: 0.381–0.752), OGFR was 0.537 (95% CI: 0.360–0.713), RELA was 0.562 (95% CI: 0.402–0.721), and SEMA6B was 0.616 (95% CI: 0.451–0.782) for PD diagnosis in the validation set GSE49126 respectively (Figure 7a–h). Among them, CBL exhibited an AUC >0.7, indicating its effective ability to distinguish PD from HC. AVP and SEMA6B also showed AUC values above 0.6, suggesting that they possess certain diagnostic values. In our subsequent research, we focused on the diagnostic efficacy of CBL in PD.

CBL Exhibits Good Diagnostic Value as a Potential Biomarker

To validate CBL's expression levels, QPCR analysis was performed on 28 samples from individuals with PD and 16 samples from HC. The results indicated a notable increase in the expression levels of CBL in PBMC from individuals with PD in comparison to those from HC ($p<0.0001$) (Figure 8a). The diagnostic value of CBL was analyzed by ROC curve, the AUC of CBL was 0.808 (95% CI: 0.628–0.989, $p < 0.001$) (Figure 8b). It suggested CBL had good diagnostic values in PD. Consistent with previous studies, CBL is expressed in peripheral immune cells such as T cells and B cells, potentially playing a crucial role in inflammatory processes and influencing disease progression.[26,27]

Discussion

The study found 1794 DEGs from the original dataset of GSE22491 in the GEO database and 2483 IRGs from the Immport Database and obtained 157 DEIRGs by intersecting the two. In GO and KEGG enrichment analysis, these DEIRGs are involved in different immune response-related biological processes, such as cytokine-mediated signaling pathways, cytokine-cytokine receptor interactions and neuroactive ligand-receptor interactions. After that, we selected eight hub genes through LASSO and RF analysis, including AVP, CBL, CD3E, GALR3, NRAS, OGFR, RELA and SEMA6B. Our immune infiltration analysis revealed abnormal infiltration of four types of immune cells in PBMC samples from PD patients compared to healthy controls: monocytes, M0 macrophages, memory B cells, and resting mast cells. Additionally, we evaluated the correlation between eight genes and these four immune cell types. The systemic

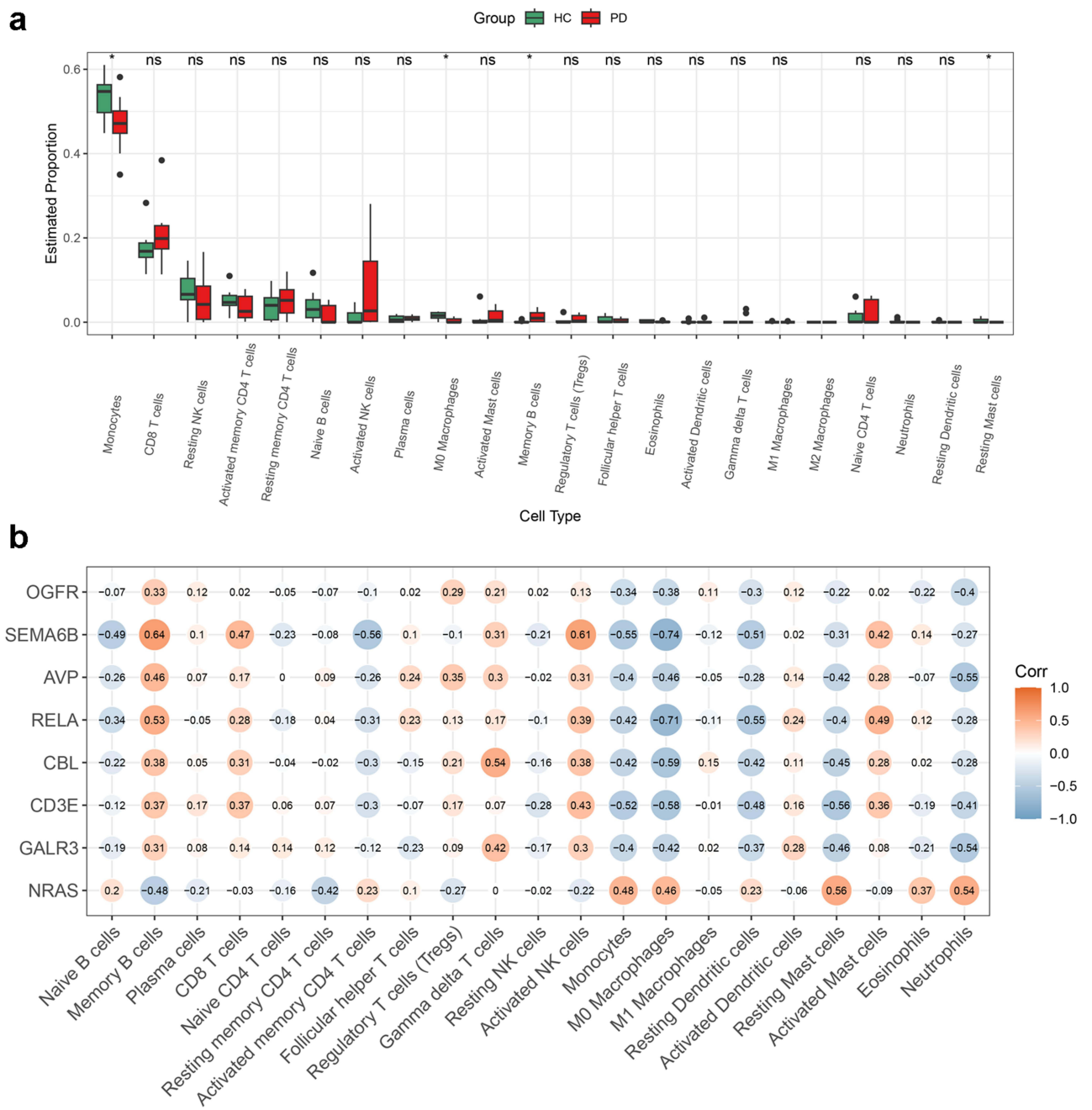

Figure 5 Analysis of immune cell infiltration in PD patients. (**a**) Boxplot showing the proportions of 22 immune cell types in PBMCs from PD and healthy control. Red represents the PD group, green represents the healthy control group; (**b**) Heat map of the correlation between eight hub genes and immune cells infiltration. The size of the dots represents the magnitude of the correlation coefficient, while the darkness of the color indicates the strength of the correlation. Blue represents a negative correlation, while Orange represents a positive correlation. *p<0.05.

inflammation that occurs in PD may lead to changes in the proportion of immune cells, and the expression of hub genes may be associated with this inflammation.

AVP is a posterior pituitary hormone that is considered a neurotransmitter and/or neuromodulator in the central nervous system. The function or level of AVP hormone may decline with age, leading to neurodegenerative diseases, especially the occurrence of PD.[28,29] The Cbl family comprises Cbl-b, c-Cbl, and Cbl3. Cbl-b is expressed in all leukocyte subsets and regulates diverse signaling pathways in T cells, NK cells, B cells, and various types of myeloid cells.[26,27] One of the subunits of CD3 is encoded by the CD3ε (CD3E) gene. The CD3 molecule can bind to the T-cell

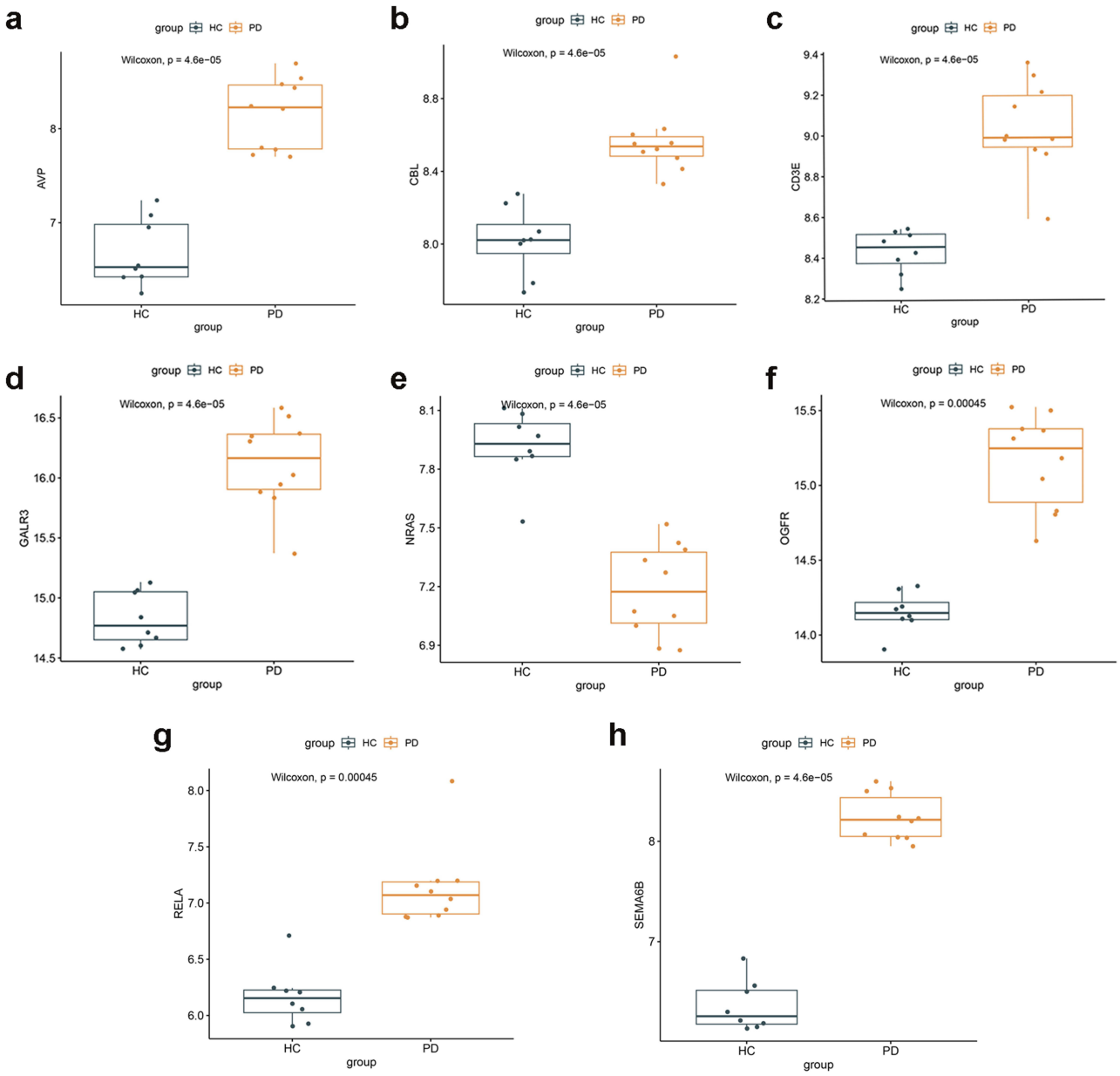

Figure 6 Boxplot of the expression levels of eight hub genes between Parkinson's disease group and healthy control group in GSE22491. (**a**) AVP; (**b**) CBL; (**c**) CD3E; (**d**) GALR3; (**e**) NRAS; (**f**) OGFR; (**g**) RELA; (h) SEMA6B. p<0.05 was considered to be statistically different.
Abbreviations: HC, healthy controls; PD, Parkinson's disease.

receptor (TCR) to assemble the CD3/TCR complex, which mediates TCR signal transduction and T-cell differentiation. This complex plays an important role in the coupling of antigen recognition and several intracellular signal transduction pathways, and may play a role in PD through inflammatory pathways.[30] Galanin exerts its biological activities through three different G protein-coupled receptors, GalR1, GalR2, and GalR3. Many studies on inflammatory animal models and immune cells have revealed the proinflammatory and anti-inflammatory functions of galanin, especially its role as an effective regulator of cytokine and chemokine expression in human macrophages.[31] NRAS, as a member of the RAS family, plays a role in the MAPK signaling pathway, and its dysregulation can lead to tumorigenesis.[32] Some studies have pointed out that abnormal activation of B cells in the body may affect the expression of NRAS, therefore, NRAS may play a role in PD by participating in changes in immune cells.[33] The protein encoded by OGFr is the receptor for opioid growth factor (OGF). When OGFr binds to OGF, it initiates signal cascades related to DNA synthesis and growth,

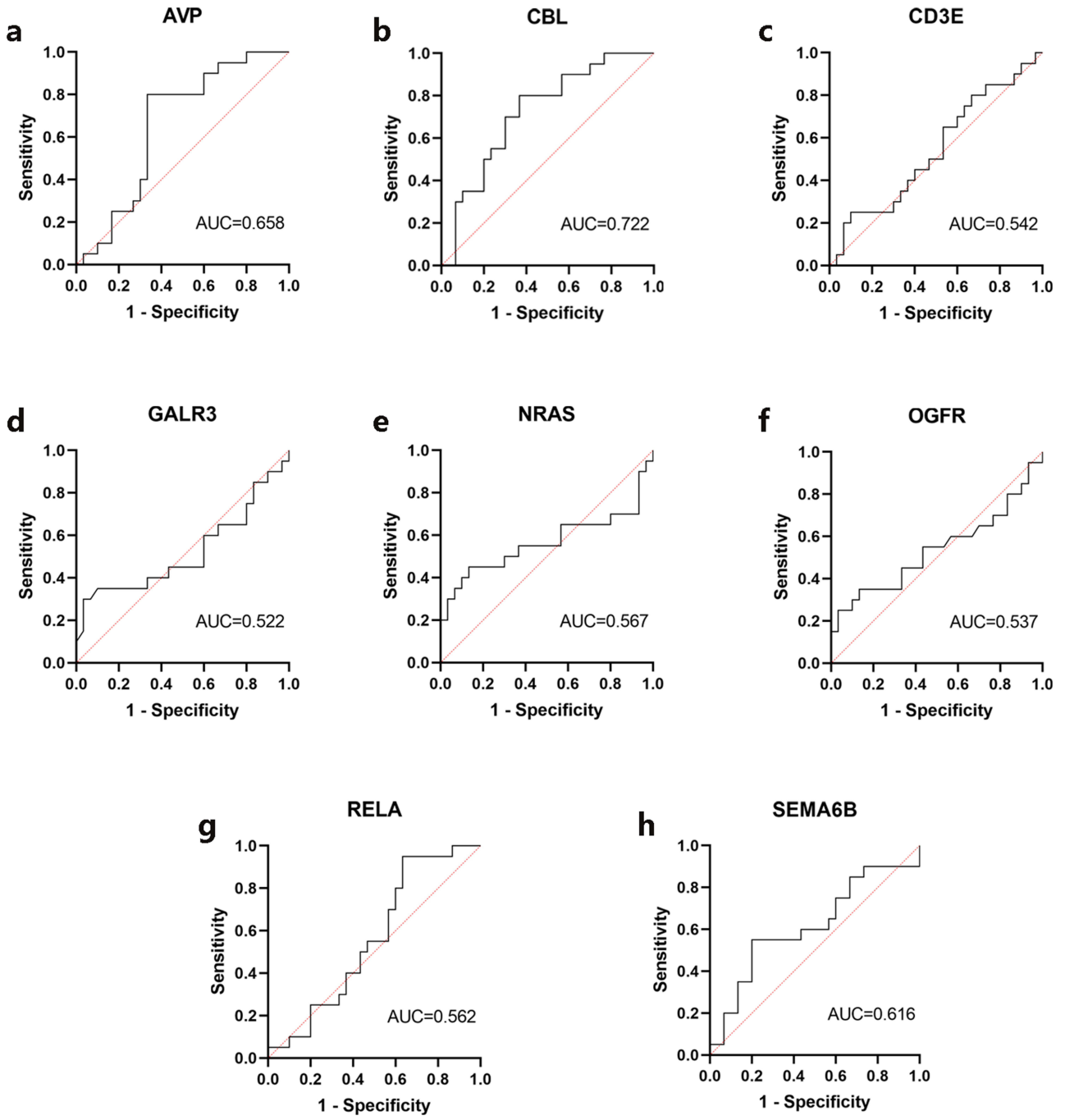

Figure 7 The ROC curve of eight hub genes based on the data set of the validation cohort GSE49126. (a) AVP; (b) CBL; (c) CD3E; (d) GALR3; (e) NRAS; (f) OGFR; (g) RELA; (h) SEMA6B.

Abbreviations: ROC, Receiver operating characteristic, AUC, area under curve.

regulates the immune and neuroendocrine systems after injury, and participates in the pathogenesis of diseases.[34] RELA is one of the factors in the NF-κB family, and the NF-κB transcription factor is a key regulator of inflammation and apoptosis. It participates in multiple immune effector cells (such as Th1, Th2, Th17, and regulatory T cells) and the inflammatory system through the secretion of proinflammatory cytokines.[35] SEMA6B is a member of the semaphorin axonal guidance family and is closely related to immunosuppressive cells such as Tregs, macrophages, and MDSCs.[36] SEMA6B is highly expressed in the human brain and may be involved in the development of the peripheral and central nervous systems.[37]

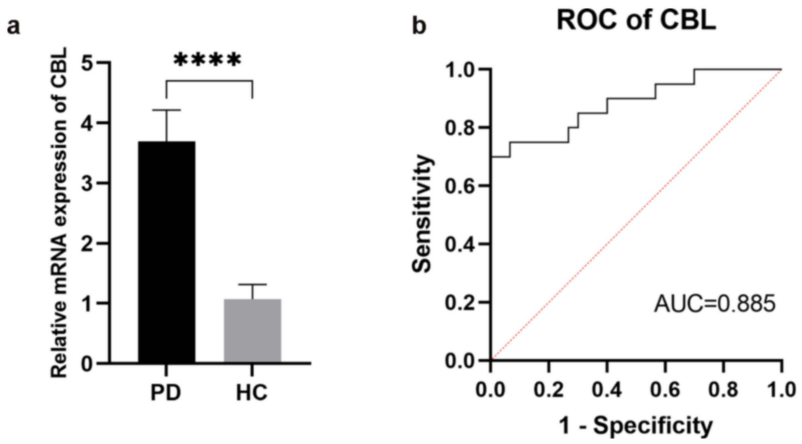

Figure 8 The expression and diagnostic value of CBL in clinical samples. (**a**) The relative mRNA expression levels of CBL expression in the PBMC of PD and HC; (**b**) The ROC curves of CBL in validation set. Data are shown as mean ±SEM, ****p<0.0001.
Abbreviations: HC, healthy controls; PD, Parkinson's disease. ROC, Receiver operating characteristic; AUC, area under curve.

Due to the impairment of the blood-brain barrier in PD patients, peripheral immune cells can migrate to the brain and infiltrate the central nervous system through this barrier. They may also return to the peripheral circulation, establishing a cycle between the brain and the periphery. In addition to the activation of central nervous system-resident cells and the infiltration of peripheral immune cells into the central nervous system observed in PD, there are also changes in peripheral immune cells in the blood of PD patients. This phenomenon indicates that PD is a systemic disease whose pathogenesis and progression are influenced by the immune system.[38] The correlation between infiltrating peripheral monocytes and degenerative events in PD has been confirmed in rodent PD models of dopaminergic degeneration.[39] Su et al also pointed out that the proportions of intermediate and non-classical monocytes in PD patients were lower than those in the control group, which is consistent with our research results and suggests that monocytes play a role in the pathogenesis and progression of PD.[40] When the chemokine receptors of classical monocytes are activated, they can differentiate into monocyte-derived macrophages and dendritic cells (DCs), playing an indispensable role in shaping inflammation and its resolution in tissues. They connect the innate and adaptive systems, affecting the pathogenesis of PD.[41] Some studies have pointed out that mast cells can interact with neurons and glial cells, leading to the release of proinflammatory cytokines/chemokines and neuroactive mediators. Excessively high levels of these mediators may increase the permeability of the BBB and enhance inflammatory infiltration.[42,43] A published study observed the peripheral B cell phenotype in PD through single-cell sequencing, finding a decrease in naive cells and an increase in memory B cells, as well as an increase in IgG and IgA B cells and clonal expansion of memory B cells participating in antigen presentation.[44,45] This is consistent with our research results, indicating that B cells play a role in activating T cells and participate in the pathogenesis of PD.

In this study, CBL demonstrated good diagnostic value in both the validation set and clinical samples, suggesting that CBL may be a marker gene for PD. The CBL gene is a proto-oncogene that encodes the RING finger E3 ubiquitin ligase.[46] Cbl-b and c-Cbl are widely expressed in T cells and play unique roles in T cell development and tolerance induction.[26] Cbl-b is involved in the control of the immune system, cell proliferation, differentiation, and the regulation of cell morphology, serving as a core factor in maintaining the tolerance of peripheral immune cells.[27] They regulate the development, tolerance, and function of T cells and B cells by modulating signals transmitted by T cell and B cell antigen receptors and co-receptors.[47] Studies have shown that c-Cbl inhibits microglia-mediated neuroinflammation by negatively regulating the PI3K/Akt/NF-κB pathway.[48] Microglia, as active contributors to neuronal damage in PD, play an important role in the inflammatory process caused by PD pathogenesis, and inhibiting microglial activation is considered a promising therapeutic strategy for the treatment of PD.

In 2022, Lin Chen's team revealed several potential PD biomarkers through bioinformatics analysis.[49] Compared to this study, our research has made the following advancements and extensions: Firstly, we selected 2483 immune-related genes from the Immport database to identify immune-related differentially expressed genes in PD, which covers a larger

number of immune-related genes compared to the 782 genes used by Lin Chen's team, ensuring a larger-scale screening of immune-related genes. Secondly, we utilized two machine learning algorithms, LASSO and RF, to screen for diagnostic biomarkers, while Lin Chen's study only used the LASSO regression method to screen for DEGs. Compared to a single machine learning algorithm, the combination of LASSO and RF algorithms has stronger and more accurate identification capabilities in identifying biomarkers. Finally, compared to Lin Chen's work, we not only validated the diagnostic value of the biomarkers in public datasets but also further confirmed them in clinical specimens, demonstrating the potential application value of these biomarkers.

There are still some shortcomings and limitations in this study. First, the sample size of PD subjects recruited is small, the expression level of key genes may lead to different test results depending on the different sample collection method, disease severity, and test method. Therefore, it is necessary to expand the detection amount of samples, strictly control PD subjects according to the inclusion standard, and exclude other confounding factors that may affect the results. Secondly, although RT-PCR has been used to verify the expression of hub genes in clinical specimens, its regulatory mechanism in PD is still unclear, and further relevant mechanism studies are needed to verify the diagnostic and prognostic value of key genes in PD.

Conclusion

Through bioinformatics analysis, we have identified eight genes that can be used as immune-related hub genes in peripheral mononuclear cells of PD patients, including RELA, OGFR, AVP, CD3E, CBL, SEMA6B, NRAS, and GALR3. Among them, CBL exhibits good diagnostic value in the validation cohort and can be used as a biomarker for PD. At the same time, we discovered four types of immune cell infiltration abnormalities in PD patients, as well as the connection between the expression of these eight hub genes and the proportion of different immune cells. Our findings provide potential biomarkers for PD patients and insights into the relationship between gene expression and immune cell proportions, offering a new direction for future exploration of immunotherapy for PD.

Data Sharing Statement

The data that support the findings of this study are openly available in the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/) and Immport database (https://immport.niaid.nih.gov) for the data where available.

Ethics Approval

All procedures involving human participants followed the principles of the Declaration of Helsinki and were approved by the Ethics Committee of Chongqing General Hospital (KY S2023-010-01). The clinical samples involved in this study were all discarded samples from laboratory testing samples. All data have been anonymized to protect the privacy and confidentiality of the participants, informed consent was not required. Our institutional ethics committee approved the waiver of informed consent.

Acknowledgments

We gratefully acknowledge the contributions from North Sichuan Medical College, Department of Clinical Laboratory of Chongqing General Hospital, Immport database and GEO databases.

Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Disclosure

The authors report no conflicts of interest in this work.

References

1. Tansey MG, Wallings RL, Houser MC, Herrick MK, Keating CE, Joers V. Inflammation and immune dysfunction in Parkinson disease. *Nat Rev Immunol*. 2022;22(11):657–673. doi:10.1038/s41577-022-00684-6
2. Tan E-K, Chao Y-X, West A, Chan -L-L, Poewe W, Jankovic J. Parkinson disease and the immune system - associations, mechanisms and therapeutics. *Nat Rev Neurol*. 2020;16(6):303–318.
3. Rizzo G, Copetti M, Arcuti S, Martino D, Fontana A, Logroscino G. Accuracy of clinical diagnosis of Parkinson disease: a systematic review and meta-analysis. *Neurology*. 2016;86(6):566–576. doi:10.1212/WNL.0000000000002350
4. Postuma RB, Poewe W, Litvan I, et al. Validation of the MDS clinical diagnostic criteria for Parkinson's disease. *Mov Disord*. 2018;33 (10):1601–1608. doi:10.1002/mds.27362
5. Zhao S, Zhang L, Ji W, et al. Machine learning-based characterization of cuprotosis-related biomarkers and immune infiltration in Parkinson's disease. *Front Genet*. 2022;13:1010361. doi:10.3389/fgene.2022.1010361
6. Singh G, Khatri DK. MicroRNA-gene regulatory network of TLR signaling in neuroinflammation-induced Parkinson's disease: a bioinformatics approach. *Netw Model Anal Health Inform Bioinforma*. 2024;13(1):7. doi:10.1007/s13721-024-00445-6
7. Samim Khan S, Janrao S, Srivastava S, Bala Singh S, Vora L, Kumar Khatri D. GSK-3β: an exuberating neuroinflammatory mediator in Parkinson's disease. *Biochem Pharmacol*. 2023;210:115496. doi:10.1016/j.bcp.2023.115496
8. Terkelsen MH, Klaestrup IH, Hvingelby V, Lauritsen J, Pavese N, Romero-Ramos M. Neuroinflammation and immune changes in prodromal parkinson's disease and other synucleinopathies. *J Parkinsons Dis*. 2022;12(s1):S149–S163. doi:10.3233/JPD-223245
9. Stephenson J, Nutma E, Valk P, Amor S. Inflammation in CNS neurodegenerative diseases. *Immunology*. 2018;154(2):204–219. doi:10.1111/imm.12922
10. Sweeney MD, Sagare AP, Zlokovic BV. Blood–brain barrier breakdown in Alzheimer's disease and other neurodegenerative disorders. *Nat Rev Neurol*. 2018;14(3):133. doi:10.1038/nrneurol.2017.188
11. Zhang X, Shao Z, Xu S, et al. Immune profiling of parkinson's disease revealed its association with a subset of infiltrating cells and signature genes. *Front Aging Neurosci*. 2021;13:605970.
12. Rajan S, Tryphena KP, Khan S, et al. Understanding the involvement of innate immunity and the Nrf2-NLRP3 axis on mitochondrial health in Parkinson's disease. *Ageing Res Rev*. 2023;87:101915. doi:10.1016/j.arr.2023.101915
13. Abdi IY, Ghanem SS, El-Agnaf OM. Immune-related biomarkers for Parkinson's disease. *Neurobiol Dis*. 2022;170:105771.
14. Farmen K, Nissen SK, Stokholm MG, et al. Monocyte markers correlate with immune and neuronal brain changes in REM sleep behavior disorder. *Proc Natl Acad Sci U S A*. 2021;118(10):e2020858118. doi:10.1073/pnas.2020858118
15. Williams GP, Schonhoff AM, Sette A, Lindestam Arlehamn CS. Central and peripheral inflammation: connecting the immune responses of parkinson's disease. *J Parkinsons Dis*. 2022;12(s1):S129–S136. doi:10.3233/JPD-223241
16. Zhang W-Y, Chen Z-H, An -X-X, et al. Analysis and validation of diagnostic biomarkers and immune cell infiltration characteristics in pediatric sepsis by integrating bioinformatics and machine learning. *World J Pediatr*. 2023;19(11):1094–1103. doi:10.1007/s12519-023-00717-7
17. Yang Y, Hua Y, Zheng H, et al. Biomarkers prediction and immune landscape in ulcerative colitis: findings based on bioinformatics and machine learning. *Comput Biol Med*. 2024;168:107778. doi:10.1016/j.compbiomed.2023.107778
18. Duan Y, Ni S, Zhao K, Qian J, Hu X. Immune cell infiltration and the genes associated with ligamentum flavum hypertrophy: identification and validation. *Front Cell Dev Biol*. 2022;10:914781. doi:10.3389/fcell.2022.914781
19. Wang J, Vasaikar S, Shi Z, Greer M, Zhang B. WebGestalt 2017: a more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit. *Nucleic Acids Res*. 2017;45(W1):W130–W137. doi:10.1093/nar/gkx356
20. Zhou Y, Shi W, Zhao D, Xiao S, Wang K, Wang J. Identification of immune-associated genes in diagnosing aortic valve calcification with metabolic syndrome by integrated bioinformatics analysis and machine learning. *Front Immunol*. 2022;13:1.
21. Wang Z, Liu J, Wang Y, et al. Identification of key biomarkers associated with immunogenic cell death and their regulatory mechanisms in severe acute pancreatitis based on WGCNA and machine learning. *Int J Mol Sci*. 2023;24(3):3033. doi:10.3390/ijms24033033
22. Fan J, Shi S, Qiu Y, Liu M, Shu Q. Analysis of signature genes and association with immune cells infiltration in pediatric septic shock. *Front Immunol*. 2022;13:1056750. doi:10.3389/fimmu.2022.1056750
23. Steen CB, Liu CL, Alizadeh AA, Newman AM. Profiling cell type abundance and expression in bulk tissues with CIBERSORTx. *Methods Mol Biol*. 2020;2117:135–157.
24. Chen B, Khodadoust MS, Liu CL, Newman AM, Alizadeh AA. Profiling tumor infiltrating immune cells with CIBERSORT. *Methods Mol Biol*. 2018;1711:243–259.
25. Newman AM, Liu CL, Green MR, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods*. 2015;12(5):453–457. doi:10.1038/nmeth.3337
26. Liu Q, Zhou H, Langdon WY, Zhang J. E3 ubiquitin ligase Cbl-b in innate and adaptive immunity. *Cell Cycle*. 2014;13(12):1875–1884. doi:10.4161/cc.29213
27. Lutz-Nicoladoni C, Wolf D, Sopper S. Modulation of immune cell functions by the E3 Ligase Cbl-b. *Front Oncol*. 2015;5:58.
28. Marcinkowska AB, Biancardi VC, Winklewski PJ. Arginine vasopressin, synaptic plasticity, and brain networks. *Curr Neuropharmacol*. 2022;20 (12):2292–2302. doi:10.2174/1570159X20666220222143532
29. Zhang X, Zhao F, Wang C, et al. AVP(4–8) improves cognitive behaviors and hippocampal synaptic plasticity in the APP/PS1 mouse model of alzheimer's disease. *Neurosci Bull*. 2019;36(3):254–262. doi:10.1007/s12264-019-00434-0
30. Liu Y, Wu Y, Zhang P, et al. CXCL12 and CD3E as indicators for tumor microenvironment modulation in bladder cancer and their correlations with immune infiltration and molecular subtypes. *Front Oncol*. 2021;11:2.
31. Koller A, Brunner SM, Bianchini R, et al. Galanin is a potent modulator of cytokine and chemokine expression in human macrophages. *Sci Rep*. 2019;9:7237.
32. Nicoś M, Krawczyk P, Jarosz B, Sawicki M, Trojanowski T, Milanowski J. Prevalence of NRAS, PTEN and AKT1 gene mutations in the central nervous system metastases of non-small cell lung cancer. *Brain Tumor Pathol*. 2017;34(1):36–41. doi:10.1007/s10014-016-0276-2
33. Klinman DM, Mushinski JF, Honda M, et al. Oncogene expression in autoimmune and normal peripheral blood mononuclear cells. *J Exp Med*. 1986;163(5):1292–1307. doi:10.1084/jem.163.5.1292

34. Qu N, Wang X, Meng Y, Shan F. Prospective oncotarget for gynecological cancer: opioid growth factor (OGF) - opioid growth factor receptor (OGFr) axis. *Int Immunopharmacol*. 2019;75:105723. doi:10.1016/j.intimp.2019.105723
35. Barnabei L, Laplantine E, Mbongo W, Rieux-Laucat F, Weil R. NF-κB: at the borders of autoimmunity and inflammation. *Front Immunol*. 2021;12:716469. doi:10.3389/fimmu.2021.716469
36. Li T, Yan Z, Wang W, et al. SEMA6B overexpression predicts poor prognosis and correlates with the tumor immunosuppressive microenvironment in colorectal cancer. *Front Mol Biosci*. 2021;8:687319. doi:10.3389/fmolb.2021.687319
37. Correa RG, Sasahara RM, Bengtson MH, et al. Human semaphorin 6B [(HSA)SEMA6B], a novel human class 6 semaphorin gene: alternative splicing and all-trans-retinoic acid-dependent downregulation in glioblastoma cell lines. *Genomics*. 2001;73(3):343–348. doi:10.1006/geno.2001.6525
38. Huang B, Zhenxin Y, Chen S, et al. The innate and adaptive immune cells in alzheimer's and parkinson's diseases. *Oxid Med Cell Longev*. 2022;2022:1315248. doi:10.1155/2022/1315248
39. Tansey MG, Romero-Ramos M. Immune system responses in Parkinson's Disease: early and dynamic. *Eur J Neurosci*. 2019;49(3):364–383. doi:10.1111/ejn.14290
40. Su Y, Shi C, Wang T, et al. Dysregulation of peripheral monocytes and pro-inflammation of alpha-synuclein in Parkinson's disease. *J Neurol*. 2022;269(12):6386–6394. doi:10.1007/s00415-022-11258-w
41. Harms AS, Ferreira SA, Romero-Ramos M. Periphery and brain, innate and adaptive immunity in Parkinson's disease. *Acta Neuropathol*. 2021;141(4):527–545. doi:10.1007/s00401-021-02268-5
42. Jones MK, Nair A, Gupta M. Mast Cells in Neurodegenerative Disease. *Front Cell Neurosci*. 2019;13:1. doi:10.3389/fncel.2019.00171
43. Hendriksen E, Bergeijk D, Oosting RS, Redegeld FA. Mast cells in neuroinflammation and brain disorders. *Neurosci Biobehav Rev*. 2017;79:119–133. doi:10.1016/j.neubiorev.2017.05.001
44. Wang P, Luo M, Zhou W, et al. Global characterization of peripheral B cells in parkinson's disease by single-cell RNA and BCR sequencing. *Front Immunol*. 2022;13:814239. doi:10.3389/fimmu.2022.814239
45. Scott KM, Bloem BR, Brundin P. B lymphocytes in parkinson's disease. *J Parkinsons Dis*. 2022;12(s1):S75–S81. doi:10.3233/JPD-223418
46. Qiao G, Zhao Y, Li Z, et al. T cell activation threshold regulated by E3 ubiquitin ligase Cbl-b determines fate of inducible regulatory T cells. *J Iimmunol*. 2013;191(2):632. doi:10.4049/jimmunol.1202068
47. Tong H, Li X, Zhang J, et al. Ubiquitin ligases CBL and CBL-B maintain the homeostasis and immune quiescence of dendritic cells. *Front Immunol*. 2021;12:757231. doi:10.3389/fimmu.2021.757231
48. Dong L, Li Y, An H, et al. The E3 ubiquitin ligase c-Cbl inhibits microglia-mediated CNS inflammation by regulating PI3K/Akt/NF-κB Pathway. *CNS Neurosci Ther*. 2016;22(8):661–669. doi:10.1111/cns.12557
49. Chen L, Wang Y, Huang J, Hu B, Huang W. Identification of immune-related hub genes in parkinson's disease. *Front Genet*. 2022;13:914645. doi:10.3389/fgene.2022.914645