Supplementary Material

Transcriptome Analysis Reveals a Two-Gene Signature Links to Motor Progression and Alterations of Immune Cells in Parkinson's Disease

Table of Contents

SUPPLEMENTARY METHODS

Cox proportional hazards analysis

We established a gene-based prognostic signature after confirming the validity of the proportional hazard assumption. The assumption was assessed by Kaplan Meier curves in the discovery cohort and found curves did not cross indicating our modeling satisfies the proportional hazards assumption. The Cox proportional hazards analysis was performed using the survival (version 3.2-13) R package.

The time-dependent receiver operating characteristic (ROC) analysis

We used the time-dependent receiver operating characteristic (ROC) conducted by R package survivalROC (version 1.0.3) based on the nearest neighbor estimation (NNE) method to evaluate the performance of two-gene signature risk score.

Correlation analysis

To estimate the relationship between the risk score and the fraction of immune cells, we firstly used the Kolmogorov–Smirnov test to assess the normal distribution of data. After confirming data normality, Pearson's correlation analysis was performed to evaluate the relation between the risk score and immune cells. we employed the "cor.test" function in R to conduct Pearson's correlation analysis. Results with $p < 0.05$ (adjusted by the Benjamini-Hochberg method) and Pearson coefficient ≥ 0.2 were considered significantly associated.

Linear mixed effects model analysis for predicting longitudinal trajectories MDS-UPDRS part III score

A linear mixed model (LMM) was used for assessing longitudinal trajectories of serial MDS-UPDRS part III scores in patients with high-risk scores (≥ 1.46) versus low-risk scores (< 1.46) in the discovery cohort. If patients were receiving PD medications, the MDS-UPDRS part III was performed while they were off their medication. For patients who were examined only while on

medication, an off-medication score was estimated as the previous method [1]. Age at baseline, sex, duration of PD disease at baseline (years), signature status (high-risk, low-risk), time in study (years), and the last two interactions were modeled as fixed covariates. Subject and time in study intercepts and slopes were included in the model as random terms. LMM analysis was performed using R package lme4 (version 1.1-23) and p value < 0.05 was considered significant.

SUPPLEMENTARY TABLES

Supplementary Table 1. Baseline demographic and clinical characteristics of participants in this study.

	PD patients	Healthy controls	*p*
PPMI cohort			
Total	691	594	/
Caucasian Race (*N*, %)	691 (100%)	594 (100%)	/
Age (mean years, SD)	63.62 (10.57)	57.45 (13.02)	< 0.001
Male (*N*, %)	401 (58.03%)	282 (47.47%)	< 0.001
Years of education (mean years, SD)	15.54 (3.53)	16.84 (3.38)	< 0.001
Disease duration (mean years, SD)	2.99 (4.85)	NA	/
Hoehn and Yahr (mean, SD)	1.80 (0.68)	0.05 (0.26)	< 0.001
MDS-UPDRS Part III (mean, SD)	22.16 (11.45)	2.05 (3.51)	< 0.001
MoCA (mean, SD)	26.53 (3.28)	27.56 (2.05)	< 0.001
GENEPARK cohort (GSE99039 dataset)			
Total	205	233	/
Caucasian Race (*N*, %)	NA	NA	/
Age (mean years, SD)	62.17 (9.52)	61.18 (9.20)	0.30
Male (*N*, %)	101 (49.27%)	70 (30.04%)	< 0.001
Years of education (mean years, SD)	NA	NA	/
Disease duration (mean years, SD)	5.99 (4.59)	NA	/
Hoehn and Yahr (mean, SD)	NA	NA	/
MDS-UPDRS Part III (mean, SD)	20.68 (11.75)	2.10 (1.37)	< 0.001
MoCA (mean, SD)	27.17 (3.14)	27.66 (2.79)	0.11
PDBP cohort			
Total	702	458	/
Caucasian Race (*N*, %)	702 (100%)	458 (100%)	/
Age (mean years, SD)	64.71 (9.05)	62.08 (11.32)	< 0.001
Male (*N*, %)	446 (63.53%)	204 (44.54%)	< 0.001
Years of education (mean years, SD)	15.64 (2.63)	15.69 (2.46)	0.72
Disease duration (mean years, SD)	5.99 (5.36)	NA	/
Hoehn and Yahr stage (mean, SD)	2.00 (0.67)	0.07 (0.38)	< 0.001
MDS-UPDRS Part III (mean, SD)	24.71 (12.92)	1.82 (3.96)	< 0.001
MoCA (mean, SD)	25.23 (3. 59)	26.51 (2.48)	< 0.001

Group-wise comparisons were performed using Chi-Square ($\chi 2$) test for sex, Mann-Whitney for continuous variables. SD, standard deviation; NA, no data available; MDS-UPDRS, Movement Disorder Society Unified Parkinson's Disease Rating Scale; MoCA, Montreal Cognitive Assessment score.

Supplementary Table 2. The most potential genes linked to the motor progression in PD.

Gene	Function description	Risk coefficient$^{\xi}$
LILRB3 (Leukocyte Immunoglobulin Like Receptor B3)	Controls immune responses, inflammatory responses and cytotoxicity [2-5]	0.38
LRRN3 (Leucine Rich Repeat Neuronal 3)	Plays a role in the development and maintenance of the nervous system and may with function to maintain normal T cell function during chronic antigenic stimulation [6, 7]	-0.32

Risk coefficients$^{\xi}$ were derived from multivariate Cox regression analysis, adjusting by covariates of age, sex, years of education, and the duration of disease.

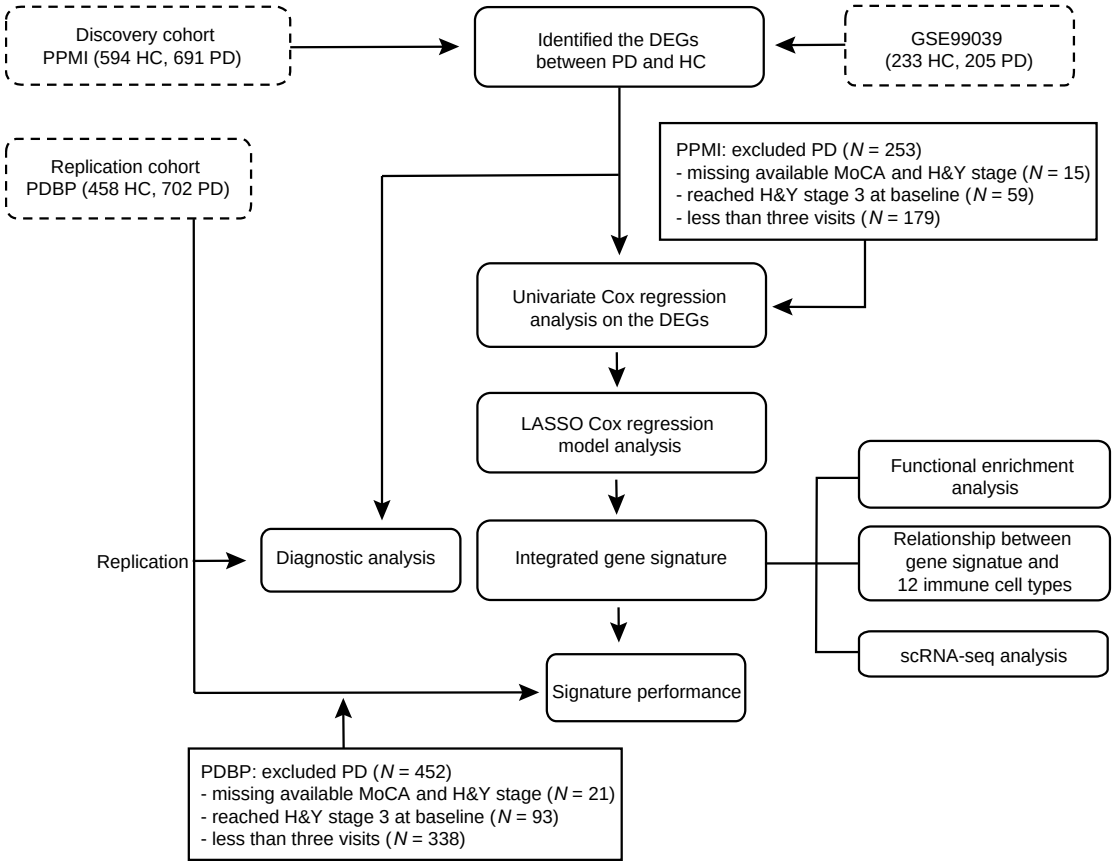
Supplementary Table 3. Baseline demographic and clinical characteristics of participants in single-cell RNA sequencing analysis.

	Patients with high-risk scores	**Patients with low-risk scores**	**Healthy controls**	***p***
Total	6	6	6	/
Caucasian Race (*N*, %)	6 (100%)	6 (100%)	6 (100%)	/
Age (mean years, SD)	67.83	65.0 (8.67)	68.50 (9.46)	0.82[†]
Male (*N*, %)	3 (50%)	3 (50%)	3 (50%)	1[€]
Years of education (mean years, SD)	16.17 (2.23)	16.67 (1.63)	18.0 (3.29)	0.60[†]
Disease duration (mean years, SD)	0.5 (0.55)	0.33 (0.51)	/	0.64[£]
Hoehn and Yahr (mean, SD)	1.67 (0.52)	1.33 (0.52)	0.17 (0.41)	0.31[£]
MDS-UPDRS Part III (mean, SD)	19.83 (8.38)	18.33 (9.81)	3.17 (6.79)	0.75[£]
MoCA (mean, SD)	27.33 (2.58)	27.33 (1.03)	27.33 (1.37)	0.81[£]

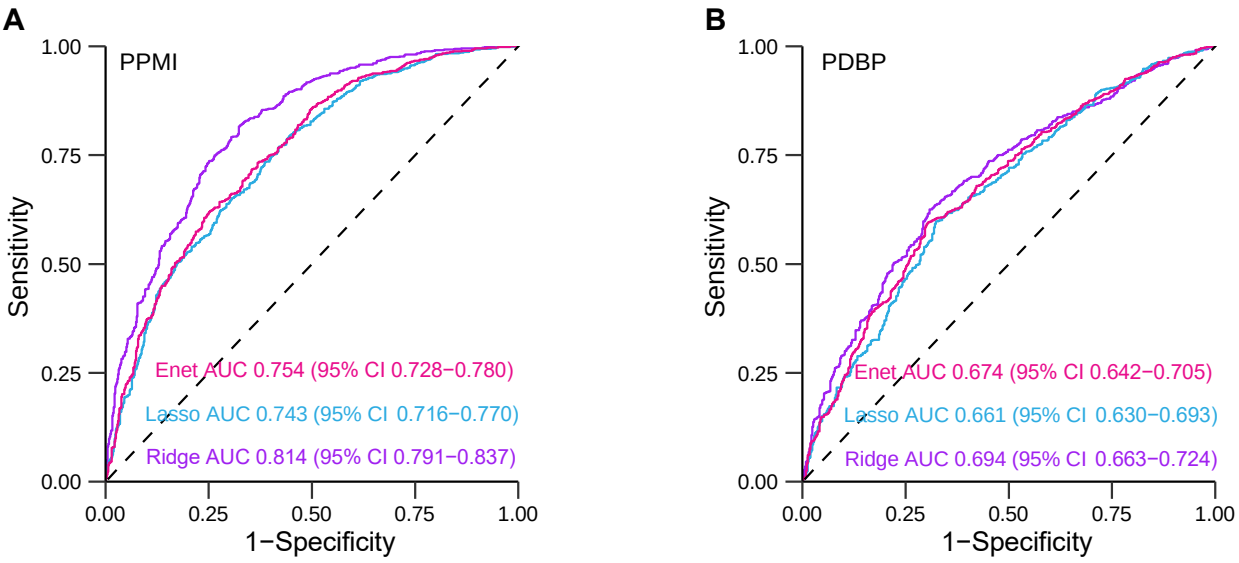
[†]Analysis of variance (ANOVA) was used for age and years of education comparisons among patients with high-risk scores, patients with low-risk scores and healthy controls. [€]The X^2 test was used for sex. [£]To compare clinical characteristics between patients with high-risk scores and patients with low-risk scores, the Mann-Whitney test was used.

SUPPLEMENTARY FIGURES

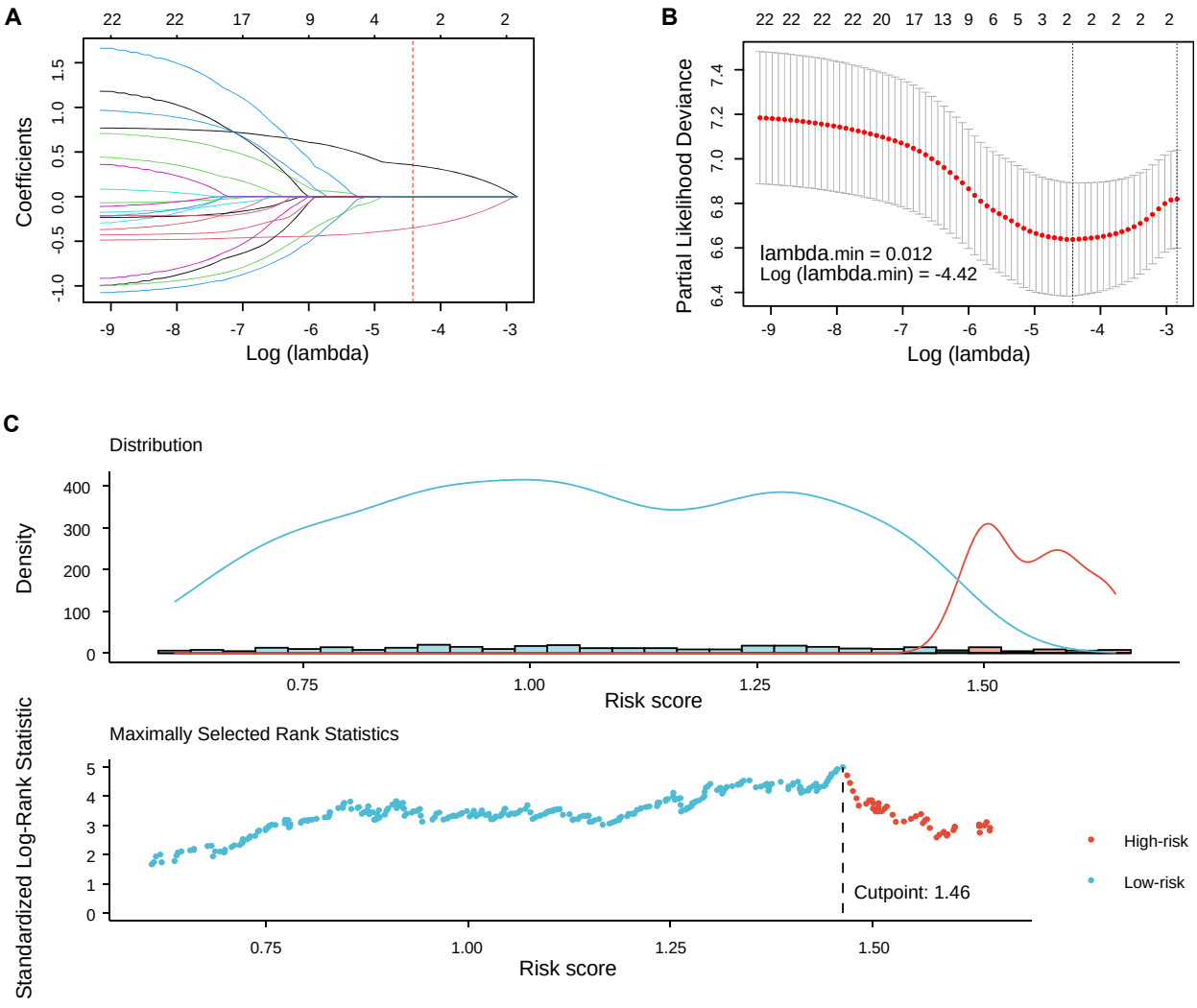
Supplementary Figure 1. The analysis flowchart of this study.
PPMI, The Parkinson's Progression Marker Initiative cohort; PDBP, The Parkinson's disease Biomarkers Program cohort; GSE99039 dataset from GENEPARK cohort, Gene expression data measured by the Affymetrix Human Genome U133 Plus 2.0 Array; HC, Healthy controls; PD, Parkinson's disease; DEGs, Differentially Expressed Genes; MoCA, Montreal Cognitive Assessment score; H&Y, Hoehn and Yahr; scRNA-seq, single-cell RNA sequencing.



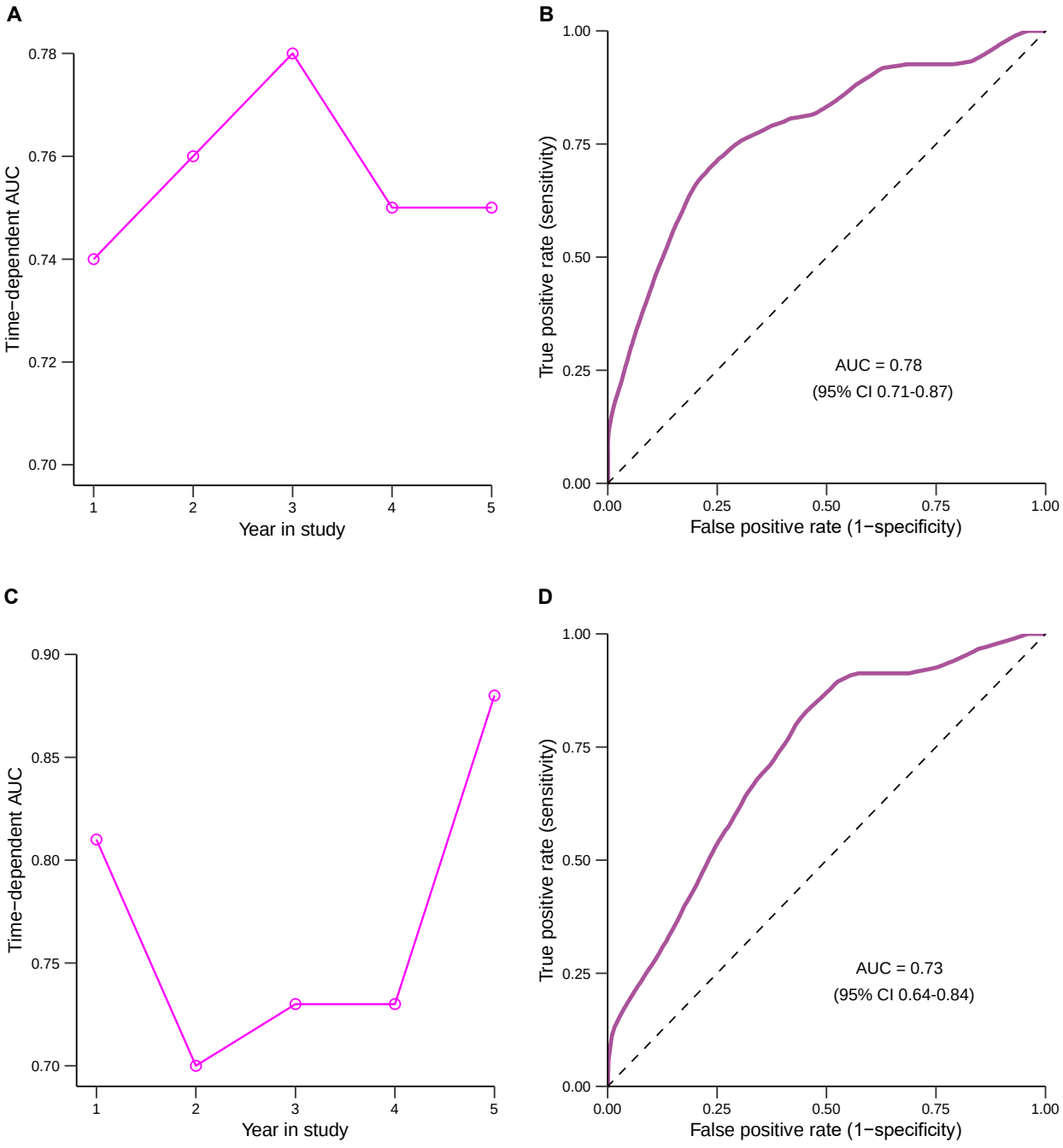
Supplementary Figure 2. Receiver operating characteristic curves in the PPMI cohort and PDBP cohort. The receiving operating characteristic curves (ROC) of three regularized regression models, including ElasticNet (Enet), Lasso and Ridge regression for discriminating patients with PD from healthy controls (HC) in the PPMI cohort with 691 PD and 594 HC (A) and PDBP cohort with 702 PD and 458 HC (B). AUC, area under the curve; 95% CI, 95% confidence interval.



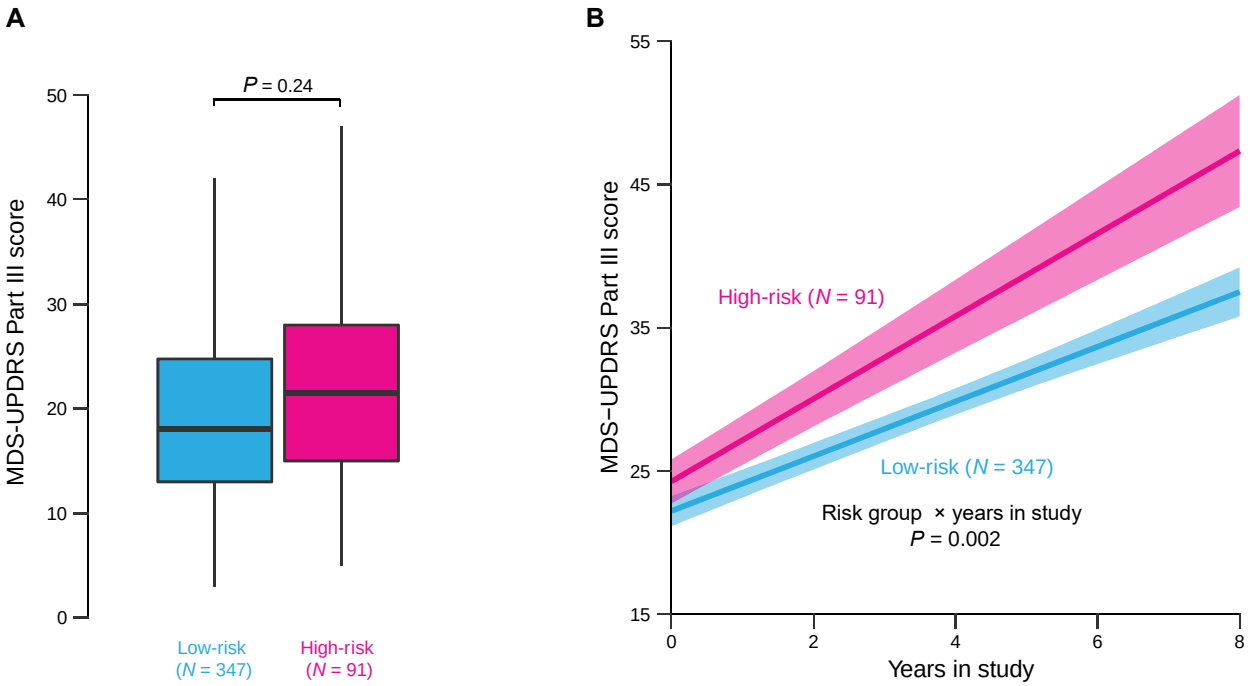
Supplementary Figure 3. The LASSO regression analysis for the construction of a prognostic gene signature. (A) Tenfold cross-validation for tuning parameter screening upon LASSO regression analysis of the 22 genes. The red dotted vertical line was drawn at the optimal parameter (lambda.min) by the minimum criteria. (B) Two genes selected using LASSO regression analysis. The left dotted vertical line was drawn at the optimal parameter (lambda.min) and the right dotted vertical line was drawn at the lambda.1se. LASSO, the least absolute shrinkage and selection operator Cox regression model; lambda.min, lambda of minimum mean cross-validated error; lambda.1se, largest value of lambda such that error is within 1 standard error of the cross-validated errors for lambda.min. (C) The optimal cutoff value of risk scores in the PPMI cohort.



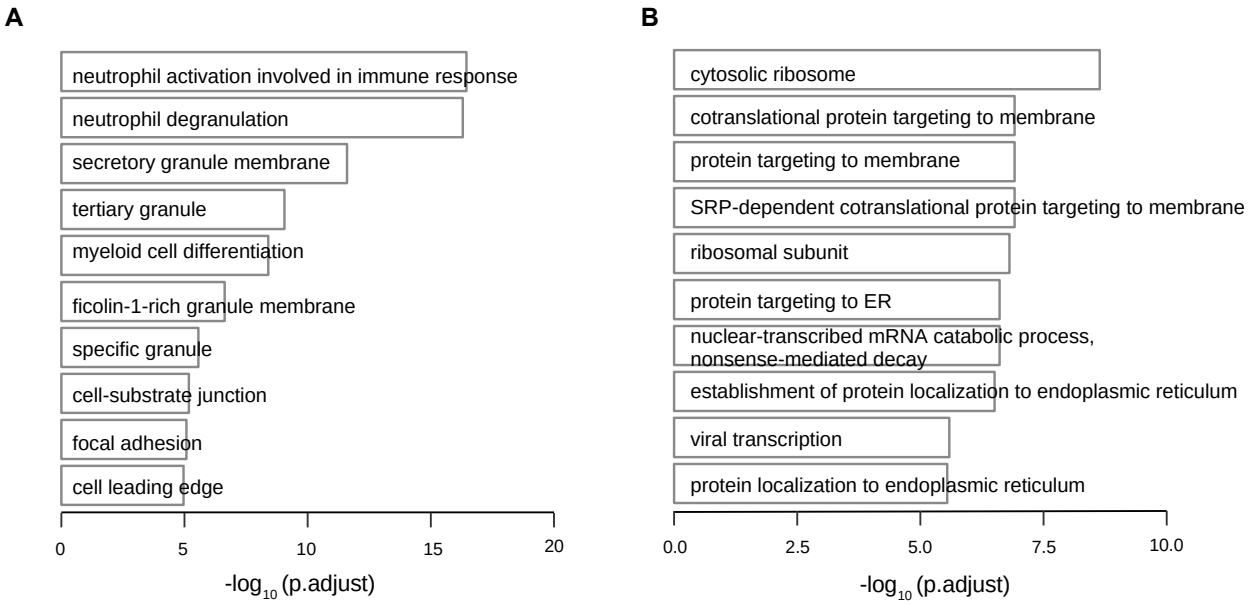
Supplementary Figure 4. Time-dependent AUCs of the two-gene signature risk score in the discovery and replication cohorts. Based on the time-dependent AUCs, the prediction performance of the two-gene signature risk score at various time points (interval of one year, up to five years during study) in the PPMI (A) and PDBP cohorts (C). The ROC curves presented prognostic capacities with the two-gene signature prior to year three since the study in the PPMI (B) and PDBP cohorts (D). ROC, receiver operating characteristic curve; AUC, area under the ROC curve; 95% CI, 95% confidence interval.



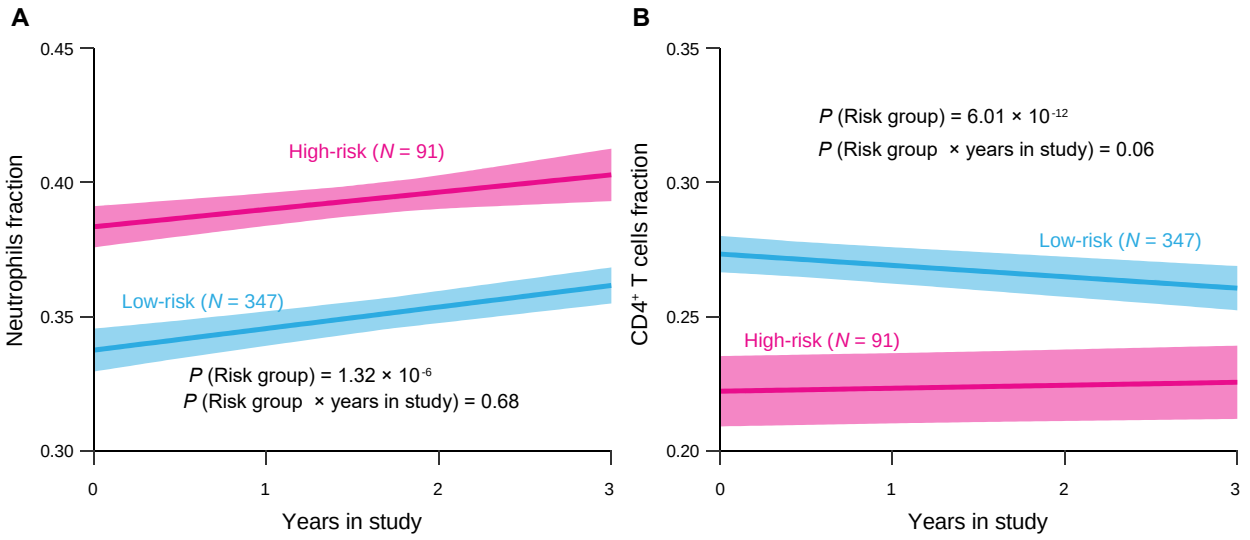
Supplementary Figure 5. High-risk PD patients with a rapid progression of motor disability. (A) The MDS-UPDRS part III score was not different between the high-risk (magenta box) and low-risk patients (cyan box) at baseline. The p value was obtained from the linear regression model analysis adjusting for covariates. (B) Adjusted mean MDS-UPDRS III scores across time predicted from the estimated fixed-effect parameters in the linear mixed effects model analysis are shown for patients with high-risk scores (magenta line) and patients with low-risk scores (cyan line). Age at baseline, sex, duration of PD disease at baseline (years), signature status (high-risk, low-risk), time in study (years), and the last two interactions were modeled as fixed covariates. Subject and time in study intercepts and slopes were included in the model as random terms. the s.e.m are displayed as light-colored bands around predicated MDS-UPDRS III scores across time. The p value from two-sided tests. MDS-UPDRS III, Movement Disorder Society-Unified Parkinson Disease Rating Scale part III.



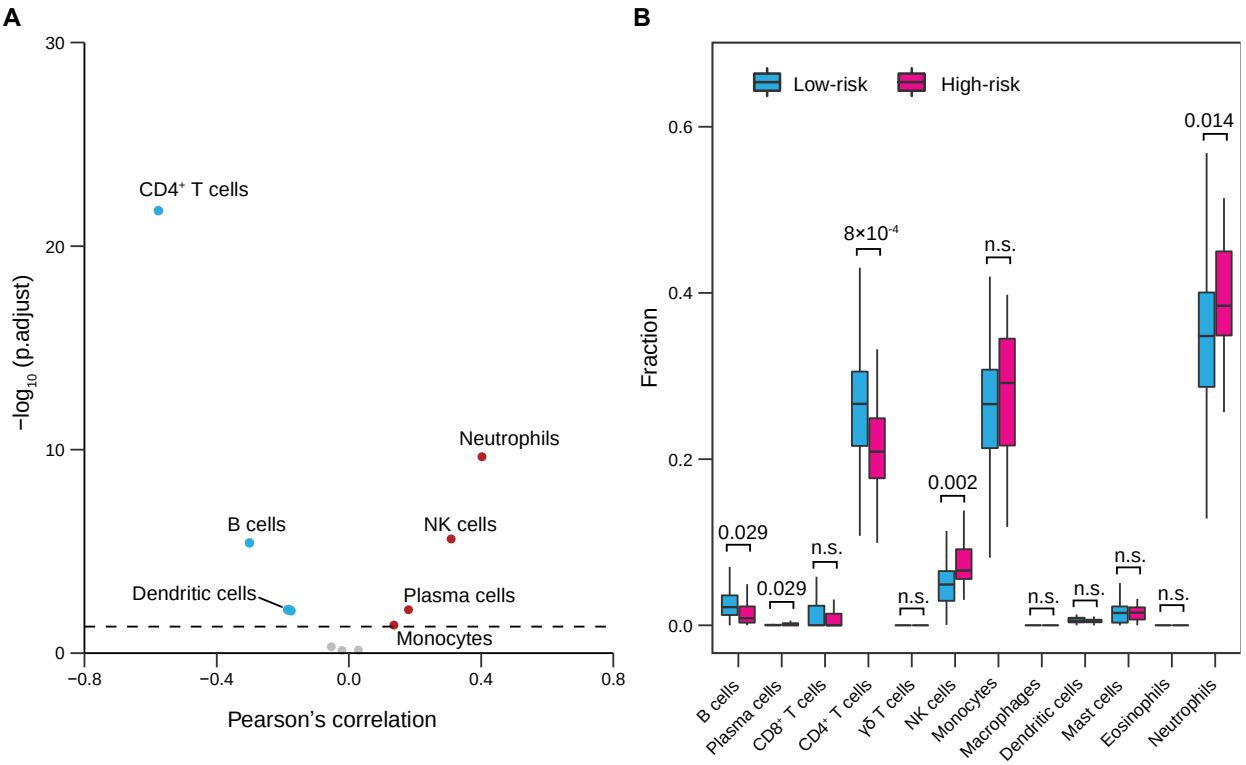
Supplementary Figure 6. The gene ontology enrichment analysis of differential expression genes between patients with PD in the high- and low-risk score groups. (A) Top 10 enriched GO terms for upregulated genes and (B) top 10 enriched GO terms for downregulated genes were shown in bar plots. Adjusted *p* values were obtained by the Benjamini-Hochberg procedure.



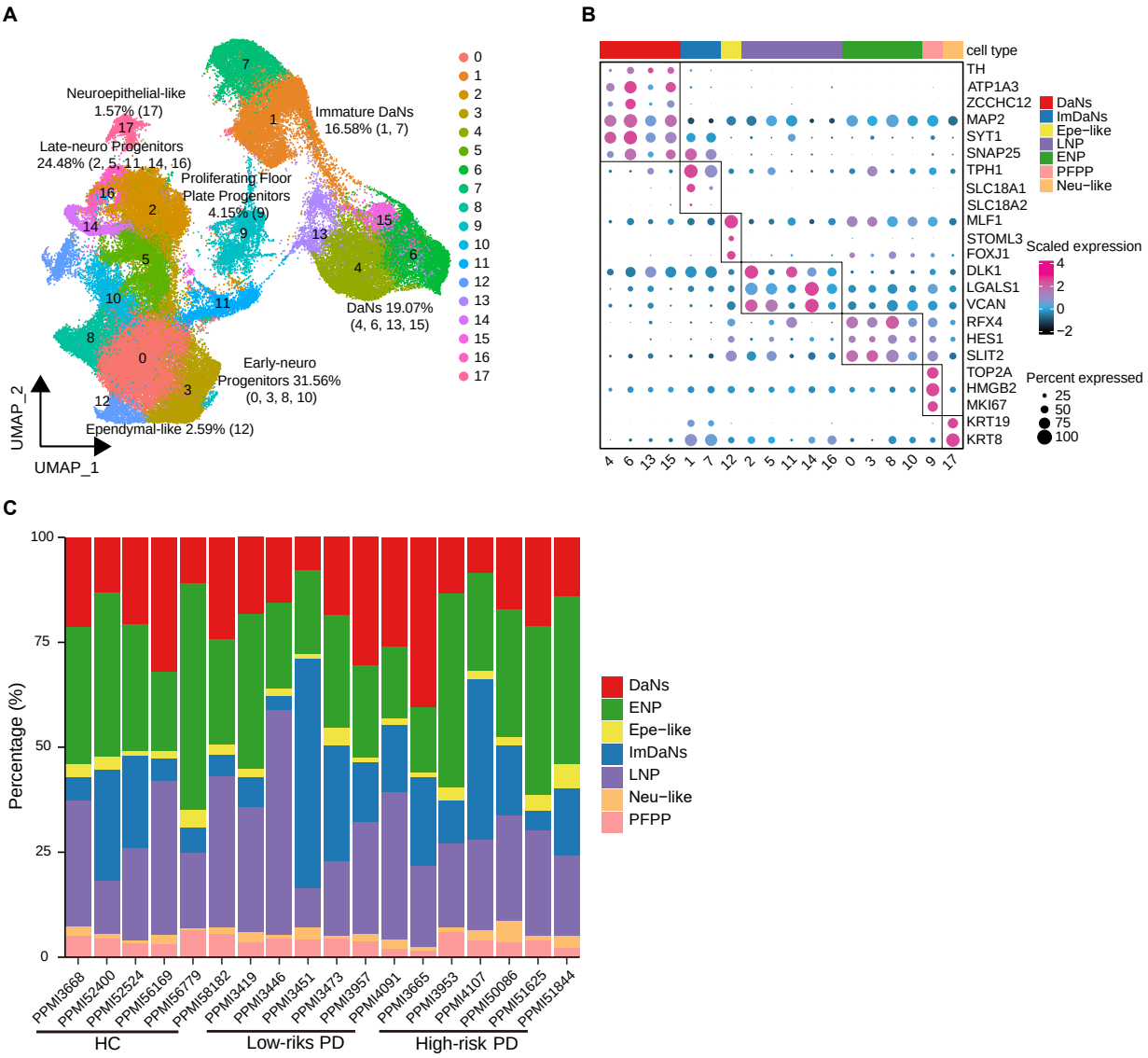
Supplementary Figure 7. The trends of neutrophils and CD4$^+$ T cells fractions in peripheral blood during the progression of PD from the PPMI cohort. Adjusted mean fractions of neutrophils (A) and CD4$^+$ T cells (B) in peripheral blood across all visits predicted from the estimated fixed-effect parameters in the linear mixed effects model analysis are shown for patients with high-risk scores (magenta line) and patients with low-risk scores (cyan line). Age at baseline, sex, duration of PD disease at baseline (years), signature status (high-risk, low-risk), time in study (years), and the last two interactions were modeled as fixed covariates. Subject and time in study intercepts and slopes were included in the model as random terms. the s.e.m are displayed as light-colored bands around predicated fractions of cell types across time. The p values from two-sided tests.



Supplementary Figure 8. The relationship between the risk scores and distinct types of immune cells in the PDBP cohort. (A) Volcano plot showed the correlations between the risk scores and 12 major immune cell fractions in 250 patients with PD using Pearson's correlation analysis. Negative correlation was marked with cyan and positive correlation with dark red. (B) Differences of the 12 major types of immune cells between patients in low-risk (*N* = 222) and high-risk (*N* = 28) score groups. The upper and lower ends of the boxes represented interquartile range of values. The lines in the boxes represented median value. n.s., not significant.



Supplementary Figure 9. Cell type composition of human iPSC-derived dopaminergic neurons. (A) UMAP visualized major cell types with their respective percentages and associated clusters in iPSC-derived DaNs from six healthy controls, six PD patients with high-risk scores and six PD patients with low-risk scores. (B) Bubble plot showed the expression levels of marker genes among all clusters. The color of bubble represented the scaled level of gene expression, and the size of bubble represented the percentage of cells expressing a specific gene. (C) The proportions of cell types in each sample. UMAP, the uniform manifold approximation and projection [8]; DaNs, dopaminergic neurons; ImDaNs, immature dopaminergic neurons; Epe-like, ependymal-like cells; LNP, late-neuro progenitors; ENP, early-neuron progenitors; PFPP, proliferating floor plate progenitors; Neu-like, neuroepithelia-like cells.



REFERENCES

[1] Paul KC, Schulz J, Bronstein JM, Lill CM, Ritz BR (2018) Association of polygenic risk score with cognitive decline and motor progression in Parkinson disease. *JAMA Neurol* **75**, 360-366.

[2] Truong AD, Hong Y, Lee J, Lee K, Tran HTT, Dang HV, Nguyen VK, Lillehoj HS, Hong YH (2019) Chicken novel leukocyte immunoglobulin-like receptor subfamilies B1 and B3 are transcriptional regulators of major histocompatibility complex class I genes and signaling pathways. *Asian-Australas J Anim Sci* **32**, 614-628.

[3] Ming S, Li M, Wu M, Zhang J, Zhong H, Chen J, Huang Y, Bai J, Huang L, Chen J, Lin Q, Liu J, Tao J, He D, Huang X (2019) Immunoglobulin-like transcript 5 inhibits macrophage-mediated bacterial killing and antigen presentation during sepsis. *J Infect Dis* **220**, 1688-1699.

[4] Huang J, Burke PS, Cung TD, Pereyra F, Toth I, Walker BD, Borges L, Lichterfeld M, Yu XG (2010) Leukocyte immunoglobulin-like receptors maintain unique antigen-presenting properties of circulating myeloid dendritic cells in HIV-1-infected elite controllers. *J Virol* **84**, 9463-9471.

[5] Yeboah M, Papagregoriou C, Jones DC, Chan HTC, Hu G, McPartlan JS, Schiott T, Mattson U, Mockridge CI, Tornberg UC, Hambe B, Ljungars A, Mattsson M, Tews I, Glennie MJ, Thirdborough SM, Trowsdale J, Frendeus B, Chen J, Cragg MS, Roghanian A (2020) LILRB3 (ILT5) is a myeloid cell checkpoint that elicits profound immunomodulation. *JCI Insight* **5**, e141593.

[6] Haines BP, Gupta R, Jones CM, Summerbell D, Rigby PW (2005) The NLRR gene family and mouse development: Modified differential display PCR identifies NLRR-1 as a gene expressed in early somitic myoblasts. *Dev Biol* **281**, 145-159.

[7] Chou JP, Ramirez CM, Wu JE, Effros RB (2013) Accelerated aging in HIV/AIDS: novel biomarkers of senescent human CD8+ T cells. *PLoS One* **8**, e64702.

[8] Becht E, McInnes L, Healy J, Dutertre CA, Kwok IWH, Ng LG, Ginhoux F, Newell EW (2019) Dimensionality reduction for visualizing single-cell data using UMAP. *Nat Biotechnol* **37**, 38–44.