Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

# Identification and Experimental Validation of Candidate Biomarkers for Parkinson's Disease through Integrated Bioinformatics, Mendelian Randomization, and Machine Learning

**Xiaofei Guan**

University of Shanghai for Science and Technology

**Wenmin Yu**

6140101@jju.edu.cn

Jiujiang University

**Research Article**

**Keywords:** Parkinson's disease, biomarkers, Mendelian randomization, machine learning, drug screening

**Posted Date:** September 9th, 2026

**DOI:** https://doi.org/10.21203/rs.3.rs-10776401/v1

**License:** This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

**Additional Declarations:** No competing interests reported.

# Identification and Experimental Validation of Candidate Biomarkers for Parkinson's Disease through Integrated Bioinformatics, Mendelian Randomization, and Machine Learning

Xiaofei Guan[a,b], Wenmin Yu[b,c,*]

[a]School of Health Science and Engineering, University of Shanghai for Science and Technology, No. 516 Jungong Road, Shanghai 200093, China

[b]The School of Basic Medical Science, Jiujiang University, No. 551 Qianjin East Road, Jiujiang 332000, Jiangxi, China

[c]Jiangxi Provincial Key Laboratory of Cell Precision Therapy, Jiujiang University, No. 551 Qianjin East Road, Jiujiang 332000, Jiangxi, China

*Correspondence: Wenmin Yu, 6140101@jju.edu.cn; Jianjun Xiong, xiongjj1975@163.com

**Abstract**

**Background:** Identifying reliable biomarkers and therapeutic targets is important for improving the early diagnosis and treatment of Parkinson's disease (PD). This study aimed to identify PD-related core genes, validate their expression, and explore potential therapeutic drugs. **Methods:** PD transcriptomic datasets were analyzed using differential expression analysis, Mendelian randomization, and machine learning. Candidate genes were validated by qPCR in an MPTP-induced PD mouse model. Potential drugs were screened using the DSigDB database, followed by molecular docking and 100 ns molecular dynamics simulations to evaluate protein–ligand binding and stability. **Results:** Seven core genes were identified, including AMIGO2, ATG7, CDK2AP2, CYP27A1, MT1X, NIPSNAP1, and TTC19. AMIGO2, ATG7, NIPSNAP1, and TTC19 were downregulated, whereas CDK2AP2, CYP27A1, and MT1X were upregulated. Animal experiments validated the abnormal expression of ATG7, NIPSNAP1, TTC19, CYP27A1, and MT1X. Drug screening identified 239 candidate compounds, of which five met the screening criteria. Molecular docking and molecular dynamics simulations indicated stable interactions between the top three drug–target complexes. **Conclusion:** This study identified and experimentally validated several PD-related core genes and potential therapeutic compounds, providing candidate biomarkers and targets for further investigation of PD mechanisms and treatment.

**Keywords**

Parkinson's disease, biomarkers, Mendelian randomization, machine learning, drug screening

## 1. Introduction

Parkinson's disease (PD) is the second most common neurodegenerative disorder worldwide, and its pathological characteristics are primarily manifested by the progressive loss of dopaminergic neurons in the substantia nigra pars compacta and dopamine depletion in the striatum[1]. According to data from the Global Burden of Disease Study (GBD) 2021, the global prevalence of PD was approximately 11.77 million in 2021 and is projected to rise to approximately 25.20 million by 2050—an increase of about 114% compared to 2021. Among these regions, East Asia is projected to bear the heaviest burden of PD, with the number of people living with the disease expected to reach approximately 10.9 million by 2050, indicating that PD has become an increasingly serious global public health issue amid population aging and growth[2]. The pathogenesis of PD is complex and not yet fully understood, involving multiple interrelated pathological processes such as oxidative stress, neuroinflammation, mitochondrial dysfunction, abnormal protein aggregation, and an imbalance in the autophagy-lysosomal system[3, 4].

These abnormal biological processes may interact to collectively promote the dysfunction and progressive loss of dopaminergic neurons, further leading to impairment of the substantia nigra–striatal dopaminergic pathway[5]. At the same time, oxidative stress and mitochondrial dysfunction can reinforce each other, while abnormal protein aggregation and impaired clearance may further exacerbate neuroinflammation and disrupt cellular homeostasis, ultimately resulting in persistent neurodegenerative damage[6, 7]. The clinical presentation of PD is highly heterogeneous, encompassing not only typical motor symptoms such as resting tremor, rigidity, and bradykinesia, but also a variety of non-motor symptoms, including depression, anxiety, sleep disturbances, and constipation. These symptoms may appear at different stages of the disease and have a significant impact on patients' daily lives, social functioning, and quality of life, while also increasing the burden on family caregivers[1, 8]. Although understanding of the clinical characteristics and pathological mechanisms of PD has deepened in recent years, early identification and accurate diagnosis remain significant challenges. Currently, the clinical diagnosis of PD relies primarily on patient history, neurological examination, and typical clinical manifestations; there is a lack of objective, reliable molecular biomarkers capable of accurately reflecting pathological changes in the early stages of the disease[9]. In addition, the treatment of PD primarily relies on medications such as levodopa and dopamine receptor agonists to alleviate motor symptoms. These treatments are mainly aimed at symptom relief and cannot halt the ongoing damage to dopaminergic neurons or slow the progression of the disease[10, 11], Therefore, screening for biomarkers at the molecular level that are closely associated with the onset and progression of PD and possess high diagnostic utility is of great significance for improving the ability to detect PD at an early stage, gaining a deeper understanding of its underlying molecular mechanisms, and identifying potential therapeutic agents. This study integrates multiple PD transcriptomic datasets, eQTL data, and large-scale PD GWAS data. First, PD-associated differentially expressed genes were identified through differential expression analysis, and candidate genes with potential causal associations with PD were screened using Mendelian randomization analysis. Subsequently, the candidate genes were further screened using various machine learning algorithms, and their diagnostic performance and stability were evaluated using ROC curves, confusion matrices, and an independent dataset[12]. Building on this foundation, we conducted functional enrichment analysis, drug druggability assessment, and experimental validation to evaluate the potential biological significance of the core candidate genes. Through this multi-level integrated analysis, this study aims to identify key PD molecules with stable diagnostic value, thereby providing new evidence for the early diagnosis of PD, research into disease mechanisms, and the discovery of potential therapeutic targets.

## 2. Materials and Methods

### 2.1 Research Design

This study established a comprehensive research framework spanning from omics data mining, causal inference, and machine learning-based screening to animal experiments and validation of molecular mechanisms (Figure 1). First, we integrated the PD transcriptomic datasets from GSE7621 and GSE8397 to perform differential expression analysis, identifying PD-associated differentially expressed genes using the screening criteria of $|\log2FC| > 0.5$ and $p.adjust < 0.05$. Subsequently, eQTL data were integrated with PD GWAS data from the Finnish PDGWAS database to conduct a Mendelian randomization (MR) analysis, screening for PD-associated candidate genes from the perspective of genetic causality, and the results of the MR analysis were intersected with the differentially expressed genes. Building on this, we further screened for core PD genes (hub genes) using various machine learning models and ROC curves, and evaluated their diagnostic performance and generalizability in

independent validation datasets such as GSE20292 and GSE35642. Furthermore, this study established an MPTP-induced PD mouse model. Changes in the animals' motor behavior were evaluated via pole-climbing and open-field tests, and the mRNA expression levels of candidate hub genes were detected using qPCR to validate the bioinformatics screening results at the in vivo experimental level. Finally, molecular docking analyses were conducted on core targets with potential drug development value to preliminarily investigate the binding patterns and key interactions between core target proteins and candidate drug molecules. Building on this, 100 ns molecular dynamics (MD) simulations were conducted to further evaluate the dynamic stability of the protein-ligand complexes. Using metrics such as RMSD, RMSF, radius of gyration (Rg), and solvent-accessible surface area (SASA), a comprehensive analysis of the conformational stability and binding characteristics of the candidate drug-target complexes was performed.

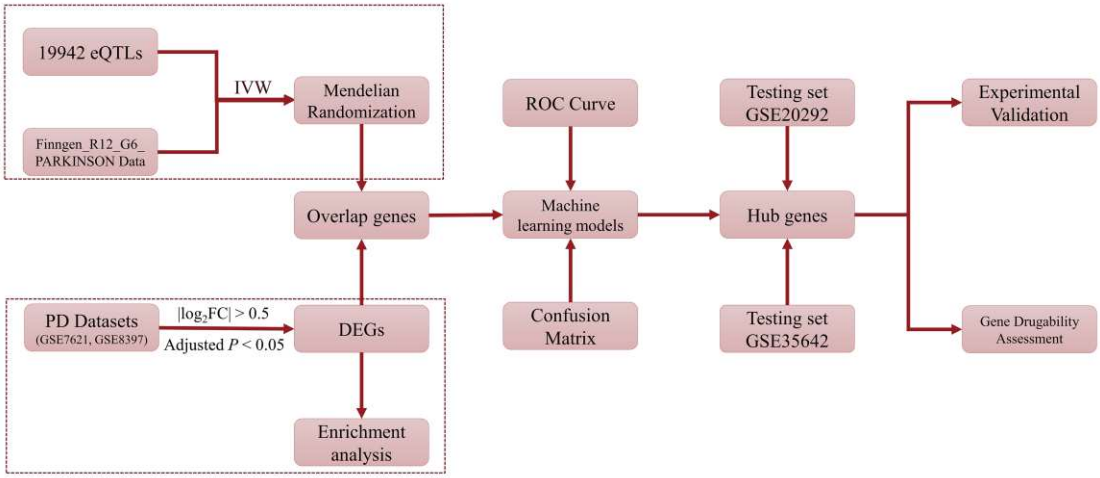



Figure 1. Flowchart of the study.

**2.2 Data Sources**

Four gene expression datasets related to PD were obtained from the Gene Expression Omnibus (GEO) database, including GSE7621, GSE8397, GSE20292, and GSE35642. Specifically, GSE7621 included 16 PD patients and 9 healthy controls[13]; GSE8397 included 29 PD patients and 18 healthy controls[14]; GSE20292 included 11 PD patients and 18 healthy controls[15]; and GSE35642 included 12 PD patients and 6 healthy controls[16]. In this study, GSE7621 and GSE8397 were selected as the primary analysis datasets for identifying PD-associated differentially expressed genes; GSE20292 and GSE35642 were used as independent validation datasets to further evaluate the expression profiles and diagnostic performance of candidate core genes. The EQTL data were obtained from the IEU OpenGWAS project (https://gwas.mrcieu.ac.uk/) and included 19,942 genes. The PD dataset was derived from the FinnGen R12 GWAS data (https://www.finngen.fi/fi), which included 5,861 PD patients and 494,487 healthy individuals; all data were from individuals of European descent.

**2.3 Differentially Expressed Genes and Enrichment Analysis**

First, the GSE7621 and GSE8397 datasets were preprocessed, and the "sva" package in R was used to correct for batch effects across the different datasets. The corrected GSE7621 and GSE8397 datasets were merged to construct a PD gene expression training set. Subsequently, the merged gene expression matrix was log-transformed and normalized to reduce technical biases between samples and improve data comparability. The "limma" R package was used to perform differential expression analysis on the merged training set to compare gene expression differences between PD patients and healthy controls. Genes with $|\log_2 FC| > 0.5$ and an adjusted P-value (adj. P) $< 0.05$ were identified as differentially

expressed genes (DEGs), which were then classified as upregulated or downregulated based on the direction of $log_2FC$. To further explore the potential biological functions of the DEGs and the signaling pathways in which they are involved, Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were performed on the identified DEGs. The GO enrichment analysis provided functional annotation across three categories: biological process (BP), cellular component (CC), and molecular function (MF), while the KEGG enrichment analysis was used to identify biological pathways in which the DEGs were significantly involved. An adjusted P-value < 0.05 was used as the significance threshold for the enrichment analysis, and the significantly enriched GO terms and KEGG pathways were visualized.

**2.4 Mendelian Randomization Analysis**

To investigate the potential causal relationship between gene expression and PD, we conducted a two-sample Mendelian randomization (MR) analysis using eQTL data for 19,942 genes from the IEU OpenGWAS database as the exposure and the FinnGen R12 PD GWAS data as the outcome. SNPs significantly associated with gene expression were selected as instrumental variables (IVs), using a selection threshold of $P < 5 \times 10^{-6}$ and performing linkage disequilibrium (LD) clumping ($r^2 < 0.01$, window size 10,000 kb)[17]. We extracted the β values, standard errors (SE), and P-values for the corresponding SNPs from the PD GWAS data, excluded SNPs significantly associated with PD ($P < 5 \times 10^{-6}$), and removed palindromic and directionally inconsistent SNPs through allelic coordination. The F-value was calculated as $F = R^2*(N-2)/(1-R^2)$, where $R^2$ represents the proportion of exposure variance explained by the SNP in the instrumental variable, and N is the sample size of the exposure GWAS; an F-value > 10 was considered to indicate the absence of significant weak instrumental variable bias[18]. MR analysis was performed using the "TwoSampleMR" package in R, with the inverse-variance weighted (IVW) method as the primary analysis approach[19]. The MR-Egger intercept test was used to assess horizontal polygenic effects, and Cochran's Q test was used to assess heterogeneity[20]. The MR-PRESSO method was employed to identify potential outlier SNPs, and causal effects were re-estimated after removing outliers to reduce the impact of horizontal polygenic effects on the MR results[21]. Finally, the influence of individual SNPs on the MR results was assessed using leave-one-out analysis to test the robustness of the results[22].

**2.5 Using Machine Learning to Identify Key Genes**

First, the set of differentially expressed genes was intersected with the set of significant genes identified through Mendelian randomization (MR) analysis to identify candidate genes. To further identify PD-associated genes with high diagnostic value, a machine learning framework was constructed based on the candidate genes, integrating 12 machine learning algorithms, including Elastic Net (Enet), Least Absolute Shrinkage and Selection Operator (Lasso), Ridge regression, stepwise logistic regression (stepglm), support vector machine (SVM), linear discriminant analysis (LDA), glmBoost, plsRglm, random forest (RF), gradient boosting machine (GBM), extreme gradient boosting (XGBoost), and naive Bayes (naive Bayes). Based on these 12 basic learning algorithms, 113 integrated models were further constructed to evaluate the diagnostic performance of different algorithm combinations[23, 24]. The combined dataset of GSE7621 and GSE8397 served as the training set, while GSE20292 and GSE35642 served as the independent test set. Model training employed 10-fold cross-validation for parameter optimization and model performance evaluation, and the area under the receiver operating characteristic (ROC) curve (AUC) was calculated for both the training set and the independent test set. The diagnostic performance of each model was comprehensively compared across the training set and the independent test set to identify the optimal machine learning model and further determine the core PD genes with

high diagnostic value.

**2.6 Development of Animal Models and Behavioral Experiments**

To verify the expression of key genes following the onset of PD, this study utilized an MPTP-induced PD mouse model. After successfully establishing the model, systematic behavioral tests were conducted, including a pole-climbing test and an open-field test[25]. Twelve C57BL/6 mice (weighing 22–24 g) were selected for the experiment, randomly divided into a PD group and a control group (n = 6). Both groups were housed in a standard laboratory animal facility, where the environment was maintained at a constant temperature (22 ± 2°C) and relative humidity (50% ± 10%), with a 12-hour light/12-hour dark cycle. During this period, the mice had free access to water and were fed sterile standard pelleted feed. During the modeling phase, mice in the experimental group received a single intraperitoneal injection of MPTP solution at a dose of 30 mg per mouse per day for 7 consecutive days, while mice in the control group received an equal volume of saline at the same time points[26, 27]. After the final injection and once the mice had stabilized, behavioral tests were conducted sequentially. In the Pole Test, mice were placed head-up at the top of a vertically fixed, rough-surfaced wooden pole, and the total climbing time (T-total) required for the mouse to climb completely from the top of the pole to the bottom was recorded to evaluate the mice's basal ganglia-mediated motor coordination and balance. In the Open Field Test (OFT), mice were placed individually in the center of a custom-built 50 × 50 × 40 cm open field chamber and allowed to explore freely for 5 minutes. Automated software was used to precisely record and extract the mice's total distance traveled (m), time spent in the central area (s), and time spent at rest (s), which were used to assess their spontaneous motor activity and neurobehavioral status[28].

**2.7 qPCR Validation**

After the behavioral tests were completed, the mice were decapitated, and tissue from the substantia nigra compacta (SNc) was rapidly isolated. Total RNA was extracted from the tissue using the RNA Easy Fast Animal Tissue/Cell Total RNA Extraction Kit (Tiangen, Beijing, China; Cat. No.: DP451) in strict accordance with the manufacturer's instructions, and its purity and concentration were verified by measuring the absorbance at 260/280 nm. Subsequently, the extracted total RNA was reverse-transcribed into complementary DNA (cDNA) using the RevertAid First Strand cDNA Synthesis Kit (Thermo Fisher, China, Cat. No.: K1622). On a real-time quantitative PCR platform, quantitative PCR (qPCR) amplification reactions were performed using SYBR® Green Real-Time PCR Master Mix (Toyobo, Shanghai, China, Cat. No.: QPK-201) to detect the transcriptional expression levels of the core causal genes AMIGO2, ATG7, CDK2AP2, CYP27A1, MT1X, NIPSNAP1, TTC19, and the tyrosine hydroxylase TH, which were identified through Mendelian randomization analysis in this study. The specific primer sequences are detailed in Table 1. The transcript expression levels of the target genes were calculated using the relative quantification 2-ΔΔCt method and normalized using β-actin as an internal reference gene. Prior to statistical analysis, the data were tested for normality. Data that followed a normal distribution were compared between groups using an independent samples t-test; data that did not follow a normal distribution were compared using the Mann–Whitney U test. A P-value < 0.05 was considered statistically significant.

Table 1.Sequences of primers used in real-time PCR analysis

| Primer Name | Primer sequences (5' to 3') |
|---|---|
| TH-F | TCTGTGCGTCGGGTGTCTGA |
| TH-R | GAATTGGTTCACCGTGCTTGTA |
| AMIGO2-F | TCCCGGTGTCGTTTGTCAAG |
| AMIGO2-R | ACTCTTTACCGACTTCAGCCTAT |

| | |
|---|---|
| ATG7-F | TGACCTTCGCGGACCTAAAGA |
| ATG7-R | CCCGGATTAGAGGGATGCTC |
| CDK2AP2-F | CTGTTTAACGACTTTGGTCCGC |
| CDK2AP2-R | CAGACAGCAGGTCCGTGTAG |
| CYP27A1-F | AGGGCAAGTACCCAATAAGAGA |
| CYP27A1-R | TCGTTTAAGGCATCCGTGTAGA |
| MT1X-F | AAGAGTGAGTTGGGACACCTT |
| MT1X-R | CGAGACAATACAATGGCCTCC |
| NIPSNAP1-F | ACAACGTGAAGCCCGAATGT |
| NIPSNAP1-R | GCGAGCAGGGATAGTCCTCA |
| TTC19-F | CGCTCTTCGCCTTGCGTAT |
| TTC19-R | AGCTGCCCTCGTATAAATGCTA |
| actb-F | GTGACGTTGACATCCGTAAAGA |
| actb-R | GCCGGACTCATCGTACTCC |

### 2.8 Lead Discovery and Molecular Docking

To further explore potential therapeutic drugs for core PD genes, drug screening was conducted using the DSigDB database in conjunction with the Enrichr platform. Potential drug candidates were identified based on the association between core genes and drug-related gene sets, and molecular docking analysis was subsequently performed to evaluate the binding affinity of the candidate drugs to the core target proteins[29]. The 3D structures of the target proteins were obtained from the PDB database, and the structural information of the candidate drugs was obtained from the PubChem database. Molecular docking was performed directly using the CB-Dock2 online platform[30]. CB-Dock2 automatically identifies potential protein binding pockets and performs receptor-ligand docking calculations based on the AutoDock Vina algorithm. Binding affinity was used to evaluate the binding capacity between candidate drugs and target proteins; generally, a binding affinity below −5.0 kcal/mol indicates good binding capacity, while a value below −7.0 kcal/mol indicates strong binding affinity[31]. Drug candidates with higher binding affinities were screened based on the docking results, and the 3D structures of the protein-ligand complexes were visualized using PyMOL software to further analyze the interactions between the drug candidates and key amino acid residues of the target protein.

### 2.9 Molecular Dynamics Simulation

To further evaluate the structural stability of the protein-ligand complex and validate the molecular docking results, a 100 ns molecular dynamics (MD) simulation was performed using GROMACS software[32]. The protein structures and their corresponding topology files were prepared using GROMACS' built-in commands, while the ligand topology files were generated using the Sobtop 1.0 tool[33]. The protein was modeled using the CHARMM36 force field, the small-molecule ligand using the CGenFF force field, and the TIP3P water model was used to construct the solvated system of the protein-ligand complex. After system neutralization, energy minimization, and NVT (isothermal isochoric) and NPT (isothermal isobaric) pre-equilibration, a 100 ns MD simulation was performed. Simulation trajectories were analyzed using GROMACS' built-in tools to calculate the root mean square deviation (RMSD), root mean square fluctuation (RMSF), radius of gyration (Rg), and solvent-accessible surface area (SASA) were calculated to comprehensively evaluate the complex's overall stability, residue flexibility, structural compactness, and degree of solvent exposure during the simulation.

## 3. Results

### 3.1 Differentially Expressed Genes and Enrichment Analysis

Differential expression analysis of the merged GSE7621 and GSE8397 datasets identified a total of 699 differentially expressed genes, including 236 up-regulated and 463 down-regulated genes (Figures 2A and 2B). Further GO enrichment analysis of the differentially expressed genes revealed that DEGs were primarily enriched in biological processes and cellular components such as synaptic vesicle cycling, neurotransmitter secretion, synaptic signal release, neuronal cell bodies, and synaptic membranes; KEGG analysis showed that DEGs were primarily involved in pathways such as phagocytosis, synaptic vesicle recycling, dopaminergic synapses, neuroactive ligand-receptor interactions, and GABAergic and serotonergic synapses (Figures 2C and D).

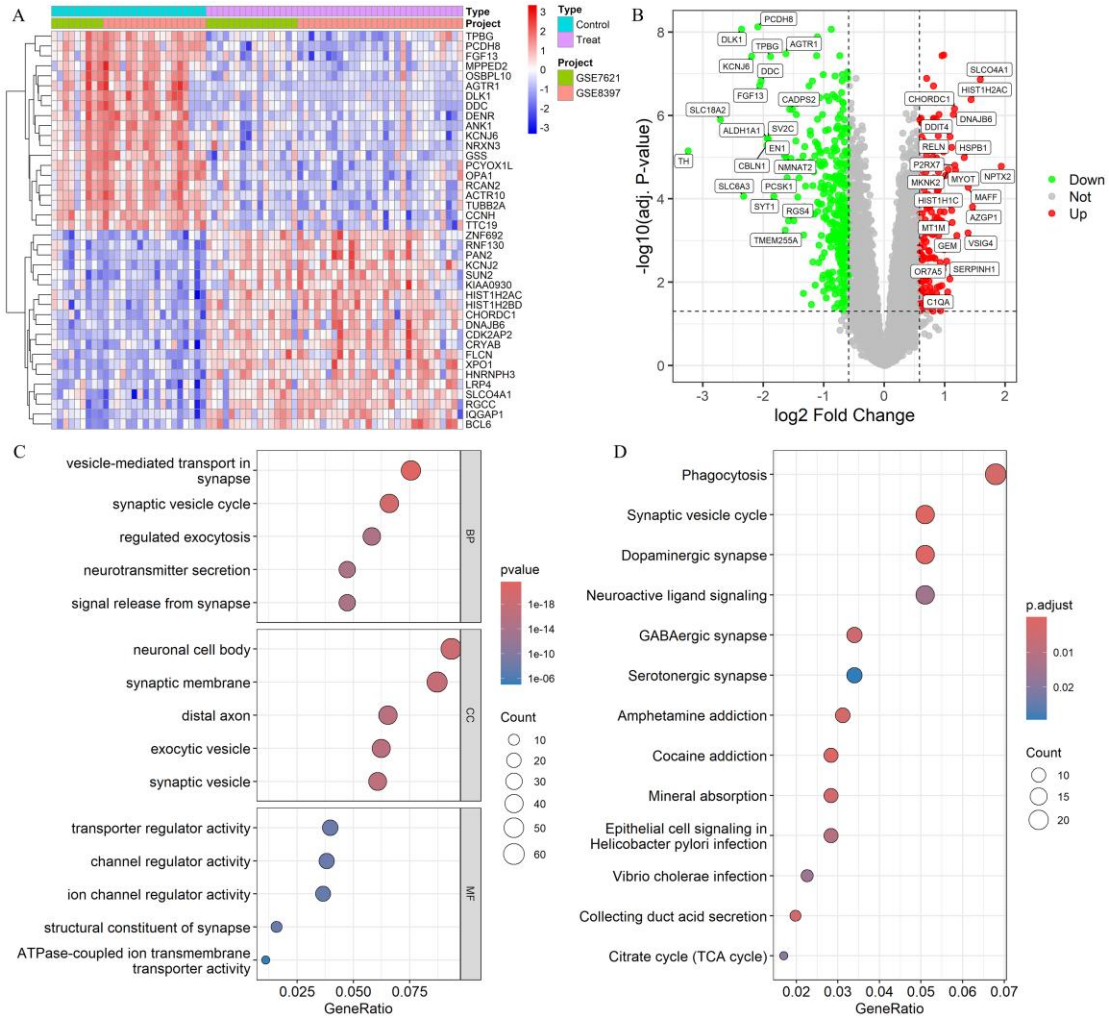



Figure 2. Differentially expressed genes and functional enrichment analysis in Parkinson's disease. (A) Heatmap of differentially expressed genes; (B) volcano plot of differentially expressed genes; (C) GO functional enrichment analysis; (D) KEGG pathway enrichment analysis.

### 3.2 Mendelian Randomization Analysis

A two-sample Mendelian randomization analysis was conducted using eQTL data from 19,942 genes and the FinnGen R12 PD GWAS data, employing the IVW method. A total of 664 genetically significant genes associated with PD were identified. Specifically, increased expression of 328 genes was significantly associated with an increased risk of RA, while increased expression of 336 genes was significantly associated with a reduced risk of RA. (Figure 3A). Further intersection analysis of the MR results with the differential expression analysis revealed that 5 genes simultaneously exhibited an MR

effect size OR > 1 and were upregulated in PD (Figure 3B); another 11 genes simultaneously exhibited an MR effect size OR < 1 and were downregulated in PD (Figure 3C), yielding a total of 16 differentially expressed genes with potential causal associations with PD.

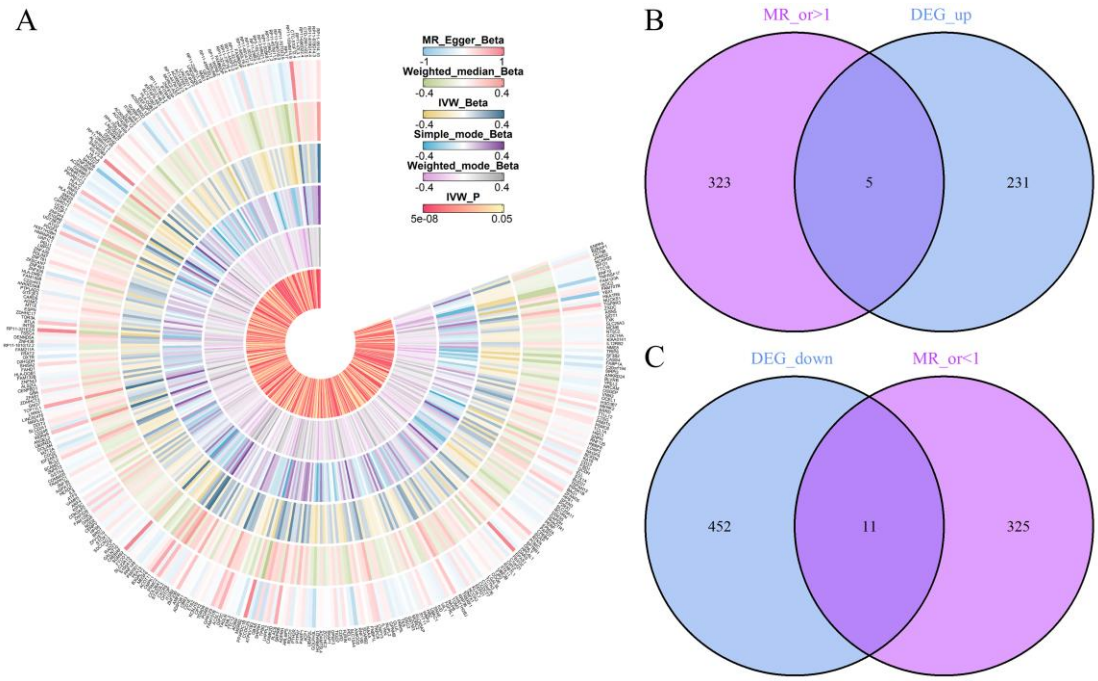



Figure 3. Integration of Mendelian randomization and differential expression analyses. (A) Circular heatmap showing the MR results of PD-related genes; (B) overlap between genes with MR effects (OR > 1) and upregulated DEGs; (C) overlap between genes with MR effects (OR < 1) and downregulated DEGs.

**3.3 Machine Learning**

After identifying differentially expressed genes and candidate genes from the Mendelian randomization analysis, 113 integrated models were constructed using 12 machine learning algorithms. The results showed that the glmBoost+XGBoost model exhibited the best diagnostic performance on the training set, with an AUC of 0.974 (95% CI: 0.937–0.998) (Fig. 4A, B), and achieved AUCs of 0.939 (95% CI: 0.843–1.000) and 0.924 (95% CI: 0.771–1.000) in the independent validation sets GSE20292 and GSE35642, respectively (Fig. 4C, D). Further single-gene ROC analysis of the seven candidate core genes revealed that TTC19 exhibited the highest diagnostic performance (AUC = 0.905), followed by NIPSNAP1 (AUC = 0.848), ATG7 (AUC = 0.836), and CDK2AP2 (AUC = 0.879) (Figure 4E). Furthermore, confusion matrices for the training set and two independent validation sets all demonstrated that the model possesses good classification performance (Fig. 4F–H).

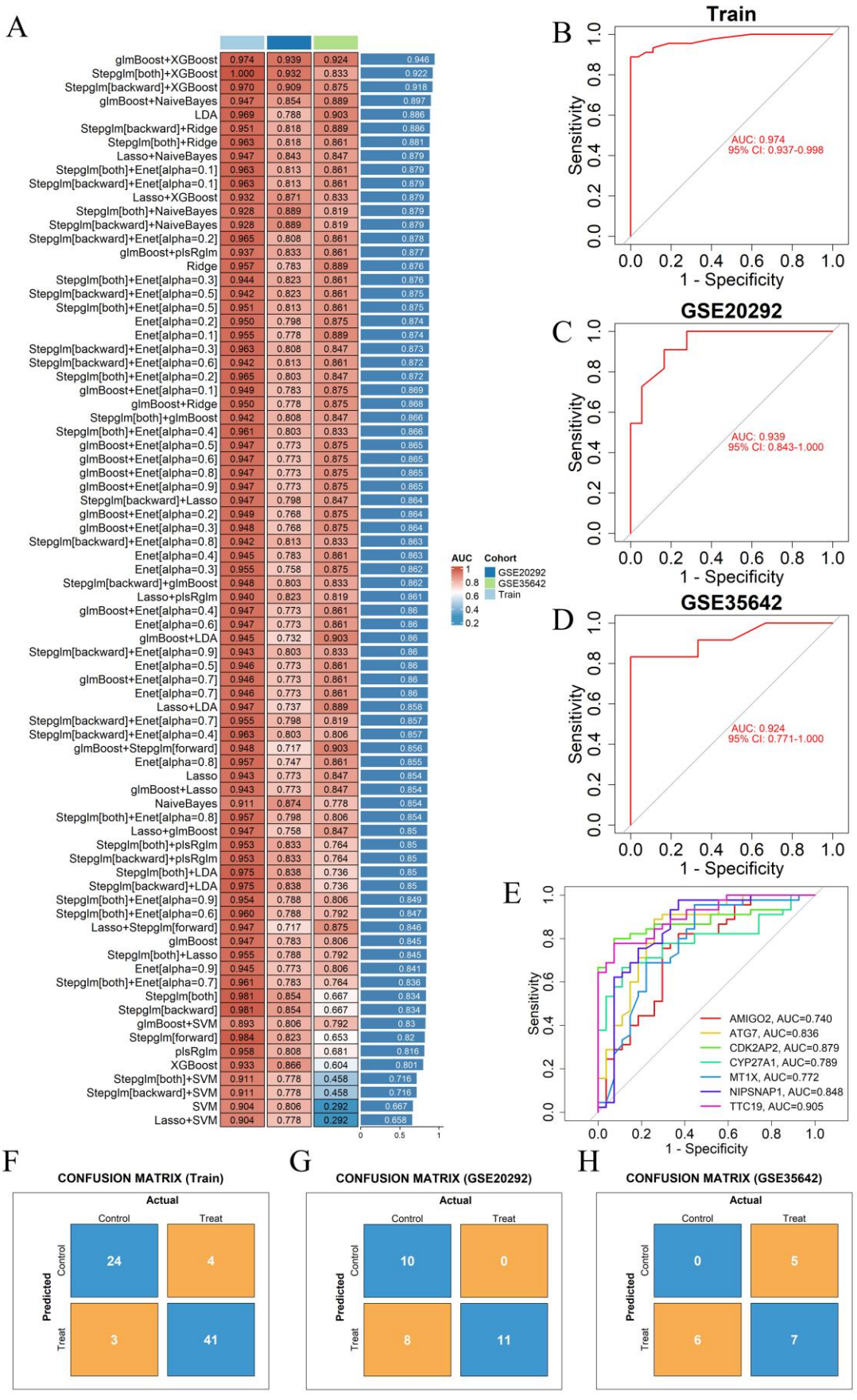



Figure 4. Machine learning-based model selection and diagnostic performance evaluation of the

core genes. (A) Comparison of AUCs among different machine learning algorithm combinations; (B–D) ROC curves of the optimal model in the training set, GSE20292, and GSE35642 test sets; (E) ROC curves and AUC values of the seven core genes; (F–H) Confusion matrices of the optimal model in the training set, GSE20292, and GSE35642 test sets.

**3.4 Mendelian Randomization Analysis of Core Genes**

To further validate the potential causal relationship between the seven core genes identified through screening and PD, a two-sample Mendelian randomization analysis was conducted, and the MR results were visualized. The IVW analysis results showed that AMIGO2 (OR = 0.801, 95% CI: 0.714–0.898, P < 0.001), ATG7 (OR = 0.885, 95% CI: 0.813–0.964, P = 0.005), NIPSNAP1 (OR = 0.854, 95% CI: 0.729–0.999, P = 0.049), and TTC19 (OR = 0.924, 95% CI: 0.867–0.985, P = 0.015) were significantly associated with a reduced risk of PD; whereas CDK2AP2 (OR = 1.170, 95% CI: 1.010–1.356, P = 0.036), CYP27A1 (OR = 1.054, 95% CI: 1.020–1.090, P = 0.002), and MT1X (OR = 1.114, 95% CI: 1.042–1.190, P = 0.001) were significantly associated with an increased risk of PD (Figure 5). The genetic instrumental variables used in the MR analyses for each core gene and their associated information are presented in Supplementary Table 2. The estimated effect directions were generally consistent across different MR methods; detailed analysis results for the MR-Egger, simple mode, and weighted mode methods are presented in Supplementary Table 3. Additionally, sensitivity analyses were conducted to assess the robustness of the MR results. Cochran's Q test was used to assess heterogeneity among instrumental variables, the MR-Egger intercept test was used to detect potential horizontal polyvalence, and MR-PRESSO was used to identify potential outliers and polyvalence. The results showed no obvious heterogeneity or horizontal polyvalence, and the MR-PRESSO analysis also did not detect any significant outliers. The leave-one-out analysis showed that the direction of the MR effects for each core gene did not change significantly after sequentially excluding individual SNPs, suggesting that the aforementioned causal associations are robust. The corresponding MR scatter plots, forest plots, leave-one-out analysis plots, and funnel plots are shown in Supplementary Figures 1–4, respectively.

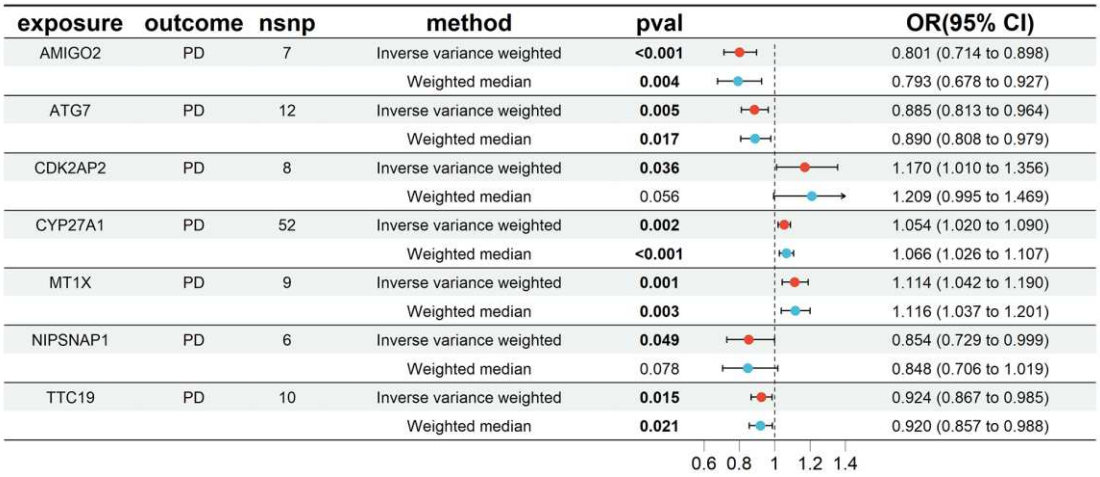



Figure 5. Forest plot of Mendelian randomization analysis of PD-related core genes.

Table 2. Information on forward MR sensitivity analysis.

| Exposure | Outcome | Heterogeneity | | Pleiotropy | | MR-PRESSO | | |
|---|---|---|---|---|---|---|---|---|
| | | Q statistic (IVW) | P value | MR-Egger Intercept | P value | Global P | Number of outliers | Distortion P |
| AMIGO2 | PD | 5.256 | 0.511 | 0.027 | 0.235 | 0.426 | NA | NA |
| ATG7 | PD | 6.430 | 0.843 | 0.002 | 0.872 | 0.911 | NA | NA |
| CDK2AP2 | PD | 6.962 | 0.433 | -0.028 | 0.161 | 0.516 | NA | NA |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| CYP27A1 | PD | 59.177 | 0.202 | 0.000 | 0.947 | 0.252 | NA | NA |
| MT1X | PD | 4.807 | 0.778 | 0.004 | 0.770 | 0.856 | NA | NA |
| NIPSNAP1 | PD | 3.913 | 0.562 | -0.043 | 0.550 | 0.704 | NA | NA |
| TTC19 | PD | 8.979 | 0.439 | -0.003 | 0.866 | 0.571 | NA | NA |

### 3.5 Results of Behavioral Experiments and qPCR Validation

To validate the reliability of the core genes identified through bioinformatics analysis and the PD model, a C57BL/6 mouse PD model was established using MPTP, followed by behavioral and qPCR assays. Results from the pole-climbing test showed that, compared with the control group, mice in the MPTP group exhibited a significantly prolonged descent time, suggesting impaired motor coordination and balance (Figure 6A). In the open-field test, MPTP-treated mice exhibited significantly reduced total distance traveled and time spent in the central area, while time spent stationary was significantly increased, indicating that MPTP treatment led to decreased spontaneous motor activity and abnormal motor behavior (Fig. 6B–D). Further qPCR analysis of SNc tissue revealed that TH mRNA expression levels were significantly reduced in MPTP-treated mice, further validating the successful establishment of the PD model (Fig. 6E). Regarding core gene expression, compared with the control group, mRNA expression levels of ATG7, NIPSNAP1, and TTC19 were significantly reduced in the MPTP group, while mRNA expression levels of CYP27A1 and MT1X were significantly elevated (Fig. 6F–J). These results indicate that MPTP-induced PD mice exhibit marked motor dysfunction and abnormal core gene expression, which is generally consistent with the results of previous bioinformatics analyses.

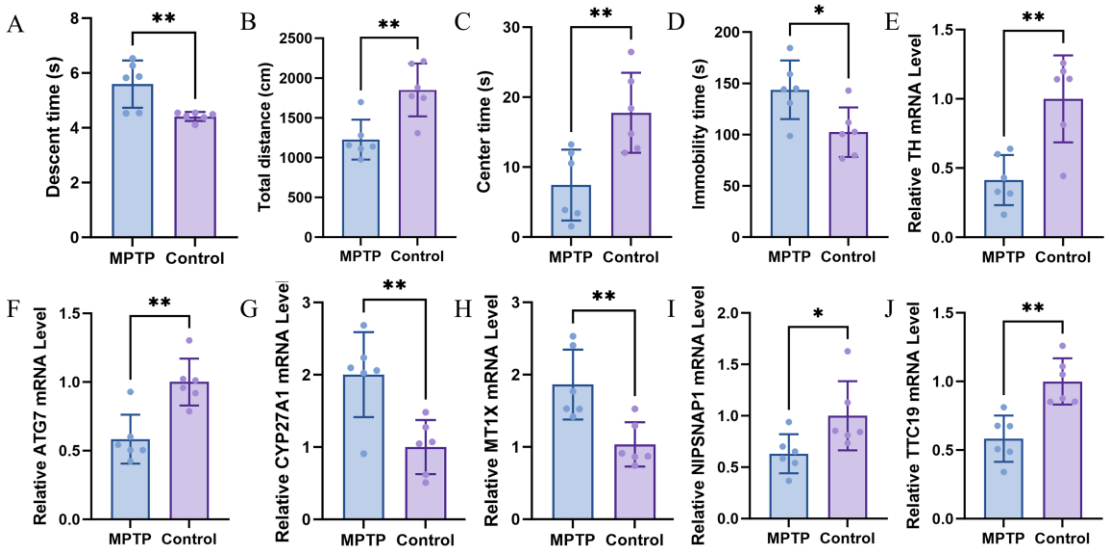



Figure 6. Efficacy evaluation of the MPTP-induced PD mouse model and validation of core genes. (A) Descent time (s) in the pole test to assess motor coordination. (B–D) Open field test (OFT) metrics: total distance (cm) (B), center time (s) (C), and immobility time (s) (D). (E) Relative mRNA level of TH in the substantia nigra pars compacta (SNc). (F–J) Relative mRNA levels of core genes ATG7 (F), CYP27A1 (G), MT1X (H), NIPSNAP1 (I), and TTC19 (J) in the SNc quantified via qPCR ($n = 6$). Data are expressed as mean ± SD. Statistical comparisons were analyzed using the Student's t-test or the Mann–Whitney U test. *$P < 0.05$, **$P < 0.01$ versus the Control group.

### 3.6 Prediction of Potential Drugs and Molecular Docking

To further explore potential therapeutic drugs for the core genes, a drug enrichment analysis was performed on the five common core genes using the DSigDB database, resulting in the identification of 239 potential candidate compounds (Supplementary Table 4). Among these, further screening of drug

candidates with P < 0.05 identified five potential drugs: periodate-oxidized adenosine, diltiazem, ambroxol, epivincamine, and eugenol (Table 3). Subsequently, molecular docking was performed between these five candidate drugs and their corresponding target genes to evaluate their potential binding affinity. The results showed that the binding energies of most drug-target protein complexes were lower than −5.0 kcal/mol, suggesting that the candidate drugs exhibit a strong binding affinity for the core target proteins (Table 4). Among them, epivincamine had the lowest binding energy with TTC19, at −7.1 kcal/mol; diltiazem had a binding energy of −6.7 kcal/mol with NIPSNAP1, the binding energy of periodate-oxidized adenosine with NIPSNAP1 was −6.6 kcal/mol, that of diltiazem with TTC19 was −6.4 kcal/mol, and that of epivincamine with ATG7 was −6.3 kcal/mol, all of which demonstrated strong predicted binding affinities. We further used PyMOL to visualize the binding patterns of the candidate drugs with the core target proteins to examine their potential binding sites and key amino acid residue interactions (Figure 7).

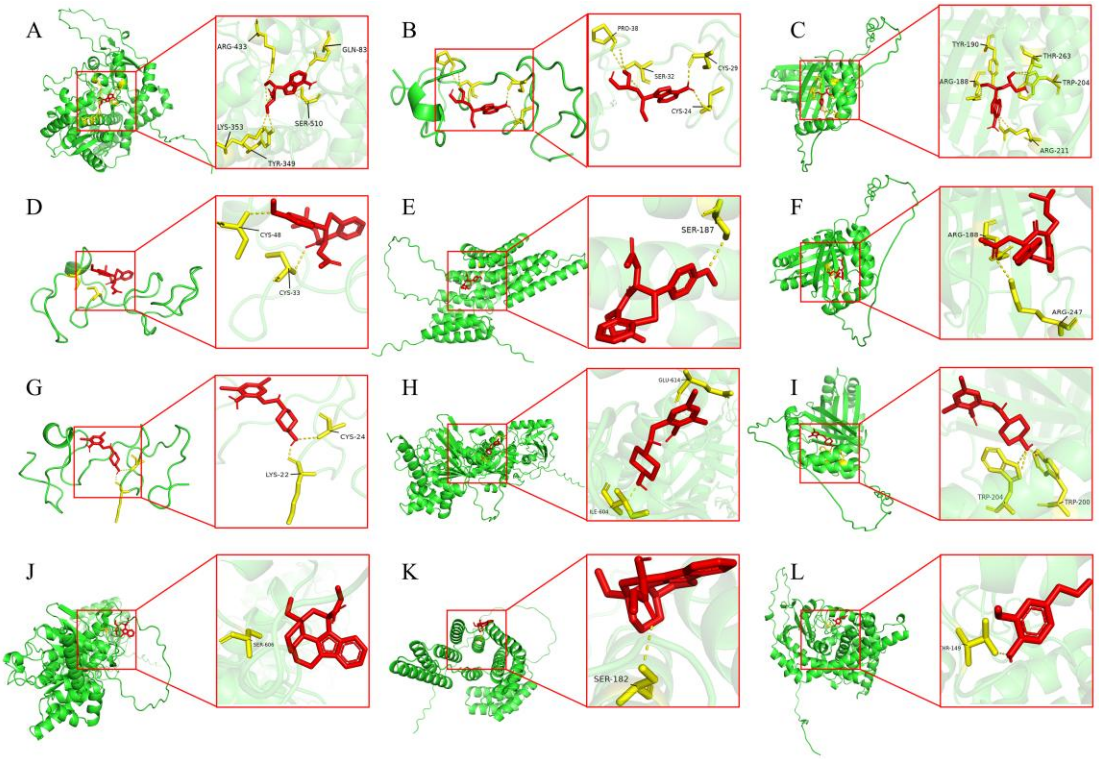



Figure 7. Three-dimensional binding modes and key amino acid residue interactions of candidate drugs with core target proteins based on molecular docking. (A) Periodate-oxidized adenosine – CYP27A1; (B) Periodate-oxidized adenosine – MT1X; (C) Periodate-oxidized adenosine – NIPSNAP1; (D) Diltiazem – MT1X; (E) Diltiazem – TTC19; (F) Diltiazem – NIPSNAP1; (G) Ambroxol – MT1X; (H) Ambroxol – ATG7; (I) Ambroxol – NIPSNAP1; (J) Epivincamine – ATG7; (K) Epivincamine – TTC19; (L) Eugenol – CYP27A1. The green structures represent target proteins, the red structures represent ligand molecules, and the yellow dashed lines indicate the key amino acid residues involved in the interaction.

Table 3. Top 5 andidate drugs predicted by DSigDB

| Drugs | Genes | *P*-value |
|---|---|---|
| Periodate-oxidized adenosine CTD 00001296 | CYP27A1;MT1X;NIPSNAP1 | 0.002922 |
| diltiazem PC3 DOWN | MT1X;TTC19;NIPSNAP1 | 0.006954 |
| ambroxol PC3 DOWN | MT1X;ATG7;NIPSNAP1 | 0.022757 |
| epivincamine HL60 DOWN | ATG7;TTC19 | 0.030297 |

| | | |
|---|---|---|
| eugenol CTD 00005949 | CYP27A1;MT1X | 0.030297 |

Table 4. Molecular docking results between genes and top 5 andidate drugs

| Drugs | Target genes | Binding energy (kcal/mol) |
|---|---|---|
| Periodate-oxidized adenosine CTD 00001296 | CYP27A1 | -5.9 |
| Periodate-oxidized adenosine CTD 00001296 | MT1X | -5.4 |
| Periodate-oxidized adenosine CTD 00001296 | NIPSNAP1 | -6.6 |
| diltiazem PC3 DOWN | MT1X | -5.9 |
| diltiazem PC3 DOWN | TTC19 | -6.4 |
| diltiazem PC3 DOWN | NIPSNAP1 | -6.7 |
| ambroxol PC3 DOWN | MT1X | -5.4 |
| ambroxol PC3 DOWN | ATG7 | -5.5 |
| ambroxol PC3 DOWN | NIPSNAP1 | -5.9 |
| epivincamine HL60 DOWN | ATG7 | -6.3 |
| epivincamine HL60 DOWN | TTC19 | -7.1 |
| eugenol CTD 00005949 | CYP27A1 | -5.3 |
| eugenol CTD 00005949 | MT1X | -3.9 |

**3.7 Molecular Dynamics Simulations**

To further validate the molecular docking results and evaluate the dynamic stability of the core target protein–candidate drug complexes under physiological conditions, this study performed 100 ns full-atom molecular dynamics simulations on three complex systems: Epivincamine-TTC19, Diltiazem-NIPSNAP1, and Periodate-oxidized adenosine-NIPSNAP1. Root-mean-square deviation (RMSD) analysis showed that after initial fluctuations, all systems stabilized around 100 ns, with RMSD values remaining within a relatively stable range, indicating that the protein-ligand complexes maintained good structural stability throughout the simulation (Figure 8A); Root-mean-square fluctuations (RMSF) results showed that residues in the protein core exhibited minimal fluctuations, indicating a robust backbone structure, with only the expected localized flexibility changes occurring at the protein termini (Figure 8B); the radius of gyration (Rg) curves remained stable throughout the simulation, showing no significant contraction or expansion, confirming the compactness of the complex's spatial structure (Figure 8C); At the same time, analysis of the solvent-accessible surface area (SASA/Area) showed that the surface exposure area of each complex gradually converged over time and maintained a dynamic equilibrium, suggesting that no significant denaturation occurred in the protein in the aqueous environment (Fig. 8D). In summary, the molecular dynamics simulation results confirmed the dynamic stability of the binding between the aforementioned drug candidates and the target protein, further supporting the reliability of molecular docking predictions and their potential practical biological applications.

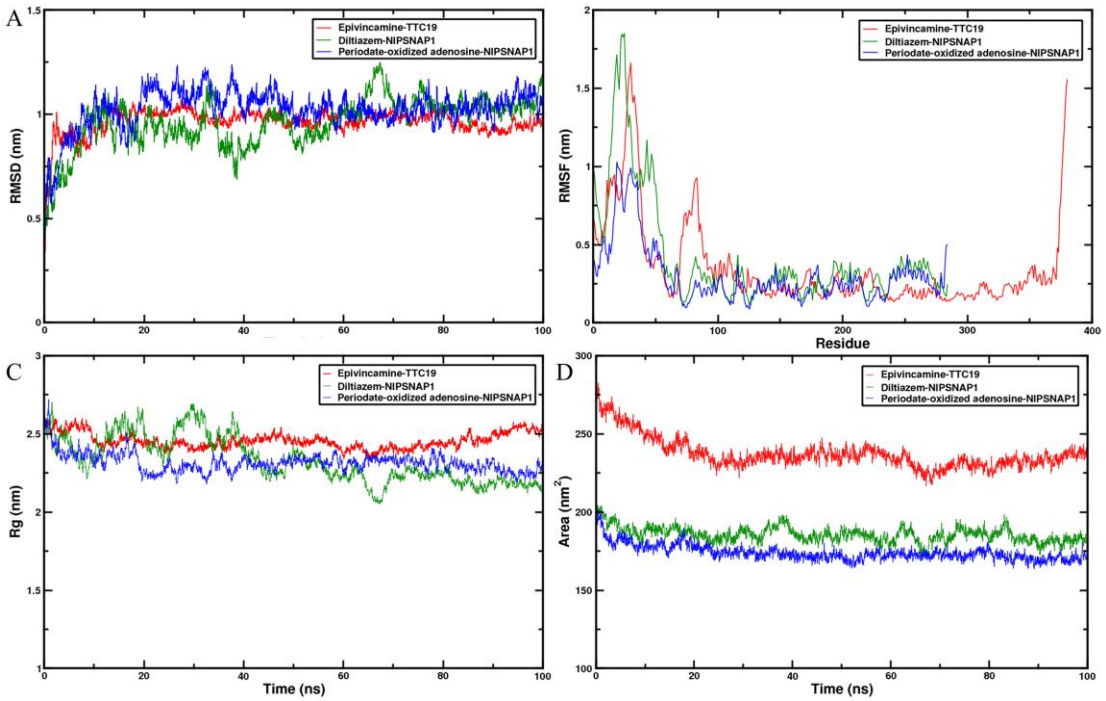



Figure 8. Molecular dynamics (MD) simulation analysis of core protein-ligand complexes over 100 ns. (A) Root mean square deviation (RMSD) profiles. (B) Root mean square fluctuation (RMSF) values per residue. (C) Radius of gyration (Rg) trajectories. (D) Solvent accessible surface area (SASA / Area) values. The analyzed systems include Epivincamine-TTC19 (red), Diltiazem-NIPSNAP1 (green), and Periodate-oxidized adenosine-NIPSNAP1 (blue).

## 4. Discussion

Based on multiple independent transcriptomic datasets from PD, this study combined differential expression analysis, machine learning-based screening, and validation with external data to ultimately identify seven potential core genes: AMIGO2, ATG7, CDK2AP2, CYP27A1, MT1X, NIPSNAP1, and TTC19. Among these, ATG7, NIPSNAP1, and TTC19 were downregulated, while CYP27A1 and MT1X were upregulated. Five of these core genes were further experimentally validated using a PD animal model; the results showed that changes in the expression of ATG7, NIPSNAP1, TTC19, CYP27A1, and MT1X were generally consistent with the direction predicted by bioinformatics analysis, thereby providing experimental support for the association between these core genes and PD. Compared to computational analyses relying solely on public databases, the combination of bioinformatics screening and animal model validation can, to some extent, reduce biases arising from single datasets, sample heterogeneity, and analytical methods, thereby enhancing the reliability of candidate biomarkers. In recent years, multi-transcriptome integration studies have also demonstrated that cross-validation using different datasets and multiple computational methods aids in identifying relatively stable disease-associated molecular features in PD[34]. In this study, ATG7, NIPSNAP1, and TTC19 were all downregulated; this result suggests that systemic changes in organelle quality control and mitochondrial homeostasis may occur in PD [35]. ATG7 is a key regulator of the autophagy process, participating in autophagosome formation and the clearance of damaged proteins and organelles within cells. When ATG7 expression is reduced, the cell's autophagy flux may be impaired, leading to the gradual accumulation of abnormal proteins and damaged organelles. This change is of particular concern in PD, as dopaminergic neurons have high metabolic demands and long axonal structures, making them highly dependent on intracellular protein and organelle quality control. Previous studies have shown that

abnormal autophagy is closely associated with PD-related neuronal damage, while restoring proper autophagy function may exert a neuroprotective effect[36, 37]. Therefore, the downregulation of ATG7 observed in this study may reflect an impairment in cellular autophagy and quality control in the substantia nigra of PD patients and may contribute to subsequent neuronal damage. The downregulation of NIPSNAP1 further supports the notion of abnormal mitochondrial quality control from another perspective. NIPSNAP1 belongs to the NIPSNAP family of proteins and acts as an “eat-me” signal on the surface of damaged mitochondria, facilitating mitochondrial autophagy so that damaged mitochondria can be recognized by the autophagy system and subsequently cleared[38]. Therefore, reduced NIPSNAP1 expression may lead to impaired mitochondrial recognition efficiency, making it easier for abnormal mitochondria to accumulate persistently within cells. Given that mitochondrial dysfunction is one of the most significant cytopathological changes in PD, the decrease in NIPSNAP1 observed in this study may be associated with the susceptibility of dopaminergic neurons to mitochondrial damage. TTC19 is involved in maintaining the function of mitochondrial respiratory chain complex III; its dysfunction can lead to complex III dysfunction and abnormalities in oxidative phosphorylation[39]. Therefore, the reduced expression of TTC19 observed in this study may further impair the efficiency of the electron transport chain and ATP production, thereby increasing the cellular oxidative stress burden. It is worth noting that ATG7, NIPSNAP1, and TTC19 point to abnormalities in mitochondrial homeostasis at three distinct levels—autophagy, recognition of damaged mitochondria, and respiratory chain function, respectively—and these findings exhibit a certain degree of intrinsic consistency. CYP27A1 was upregulated in this study. CYP27A1 is a key enzyme in cholesterol metabolism that promotes the conversion of cholesterol to 27-hydroxycholesterol (27-OHC)[40]. In recent years, a growing body of research has suggested that abnormalities in brain cholesterol metabolism may be involved in PD-related pathological processes[41]. 27-OHC can increase α-synuclein protein levels in dopaminergic neurons and may promote α-synuclein accumulation by affecting proteasome function[42]. Further studies have shown that 27-OHC promotes the pathological spread of α-synuclein and induces mitochondrial dysfunction and damage to dopaminergic neurons, whereas reducing CYP27A1 activity can alleviate these pathological changes[43]. MT1X was also upregulated. As a member of the metallothionein family, MT1X is involved in metal ion binding, redox regulation, and cellular stress responses; therefore, changes in its expression may reflect alterations in the intracellular environment in PD[44]. PD-associated substantia nigra tissue may exhibit redox imbalances and abnormalities in metal ion homeostasis; under these conditions, elevated MT1X levels may represent a compensatory response by cells to the abnormal environment[45]. In summary, downregulation of ATG7 may lead to reduced autophagy and organelle clearance capacity; downregulation of NIPSNAP1 may further impair the recognition and clearance of damaged mitochondria; and downregulation of TTC19 may affect mitochondrial respiratory chain function; At the same time, the upregulation of CYP27A1 suggests a remodeling of cholesterol metabolism, while the upregulation of MT1X may reflect the cell’s response to oxidative stress and changes in metal homeostasis. Compared to treating these genes solely as independent diagnostic markers, this interpretation of synergistic multi-process interactions may better align with the complex molecular pathophysiology of PD. Recent multi-omics studies have similarly found that PD-related molecular alterations often involve multiple interconnected cellular processes rather than being limited to a single signaling pathway[46, 47]. Building on the identification and experimental validation of core genes, DSigDB analysis identified 239 candidate compounds, of which 5 met the screening criteria. Molecular docking showed that most candidate drugs exhibited favorable binding affinities with the corresponding target proteins through hydrogen bonds and hydrophobic

interactions. To further assess the stability of the top three drug–target complexes, 100 ns molecular dynamics simulations were performed. All three complexes maintained relatively stable RMSD, RMSF, Rg, and SASA profiles after equilibration, indicating stable overall conformations, limited residue fluctuations, and preserved structural compactness and solvent exposure. These findings further support the stability of the predicted drug–target interactions.

## 5. Conclusions

This study identified several potential core genes associated with PD, including ATG7, NIPSNAP1, TTC19, CYP27A1, and MT1X, and validated their abnormal expression through animal experiments. Further screening identified several potential drug candidates; molecular docking and molecular dynamics simulation results support that some of these drugs exhibit good binding stability with their target proteins.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

The authors sincerely thank related investigators for sharing the GWAS summary statistics included in this study.

## Author Contributions

Xiaofei Guan: Conceptualization, methodology, data curation, formal analysis, investigation, visualization, and writing—original draft. Wenmin Yu: Conceptualization, supervision, project administration, and writing—review and editing. Both authors contributed to the interpretation of the results, critically revised the manuscript, and approved the final version of the manuscript.

## Ethics Approval

All animal experiments were conducted in accordance with the institutional guidelines for the care and use of laboratory animals and were approved by the Animal Welfare and Ethics Committee of Jiujiang University (Ethics No. JJU20260512).

[1] POEWE W, SEPPI K, TANNER C M, et al. Parkinson disease [J]. Nature reviews Disease primers, 2017, 3: 17013.

[2] SU D, CUI Y, HE C, et al. Projections for prevalence of Parkinson's disease and its driving factors in 195 countries and territories to 2050: modelling study of Global Burden of Disease Study 2021 [J]. BMJ (Clinical research ed), 2025, 388: e080952.

[3] MORRIS H R, SPILLANTINI M G, SUE C M, et al. The pathogenesis of Parkinson's disease [J]. Lancet (London, England), 2024, 403(10423): 293-304.

[4] TANSEY M G, WALLINGS R L, HOUSER M C, et al. Inflammation and immune dysfunction in Parkinson disease [J]. Nature reviews Immunology, 2022, 22(11): 657-73.

[5] COUKOS R, KRAINC D. Key genes and convergent pathogenic mechanisms in Parkinson disease [J]. Nature reviews Neuroscience, 2024, 25(6): 393-413.

[6] HENCHCLIFFE C, BEAL M F. Mitochondrial biology and oxidative stress in Parkinson disease pathogenesis [J]. Nature clinical practice Neurology, 2008, 4(11): 600-9.

[7] NEGI S, KHURANA N, DUGGAL N. The misfolding mystery: α-synuclein and the pathogenesis of Parkinson's disease [J]. Neurochemistry international, 2024, 177: 105760.

[8] TAN Q Y, COX N J, LIM S E R, et al. The Experiences of Treatment Burden in People with

Parkinson's Disease and Their Caregivers: A Systematic Review of Qualitative Studies [J]. Journal of Parkinson's disease, 2021, 11(4): 1597-617.

[9] TOLOSA E, POEWE W, NOYCE A J, et al. Towards biomarker-based diagnosis of Parkinson disease [J]. Nature reviews Neurology, 2026, 22(3): 139-51.

[10] KWON D K, KWATRA M, WANG J, et al. Levodopa-Induced Dyskinesia in Parkinson's Disease: Pathogenesis and Emerging Treatment Strategies [J]. Cells, 2022, 11(23).

[11] PIRKER W, KATZENSCHLAGER R, HALLETT M, et al. Pharmacological Treatment of Tremor in Parkinson's Disease Revisited [J]. Journal of Parkinson's disease, 2023, 13(2): 127-44.

[12] LANDOLFI A, RICCIARDI C, DONISI L, et al. Machine Learning Approaches in Parkinson's Disease [J]. Current medicinal chemistry, 2021, 28(32): 6548-68.

[13] LESNICK T G, PAPAPETROPOULOS S, MASH D C, et al. A genomic pathway approach to a complex disease: axon guidance and Parkinson disease [J]. PLoS genetics, 2007, 3(6): e98.

[14] MORAN L B, DUKE D C, DEPREZ M, et al. Whole genome expression profiling of the medial and lateral substantia nigra in Parkinson's disease [J]. Neurogenetics, 2006, 7(1): 1-11.

[15] ZHANG Y, JAMES M, MIDDLETON F A, et al. Transcriptional analysis of multiple brain regions in Parkinson's disease supports the involvement of specific protein processing, energy metabolism, and signaling pathways, and suggests novel disease mechanisms [J]. American journal of medical genetics Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics, 2005, 137b(1): 5-16.

[16] CABEZA-ARVELAIZ Y, SCHIESTL R H. Transcriptome analysis of a rotenone model of parkinsonism reveals complex I-tied and -untied toxicity mechanisms common to neurodegenerative diseases [J]. PloS one, 2012, 7(9): e44700.

[17] ABECASIS G R, ALTSHULER D, AUTON A, et al. A map of human genome variation from population-scale sequencing [J]. Nature, 2010, 467(7319): 1061-73.

[18] BURGESS S, THOMPSON S G. Avoiding bias from weak instruments in Mendelian randomization studies [J]. International journal of epidemiology, 2011, 40(3): 755-64.

[19] BURGESS S, DAVEY SMITH G, DAVIES N M, et al. Guidelines for performing Mendelian randomization investigations: update for summer 2023 [J]. Wellcome open research, 2019, 4: 186.

[20] BURGESS S, THOMPSON S G. Interpreting findings from Mendelian randomization using the MR-Egger method [J]. European journal of epidemiology, 2017, 32(5): 377-89.

[21] VERBANCK M, CHEN C Y, NEALE B, et al. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases [J]. Nature genetics, 2018, 50(5): 693-8.

[22] NOLTE I M. Metasubtract: an R-package to analytically produce leave-one-out meta-analysis GWAS summary statistics [J]. Bioinformatics (Oxford, England), 2020, 36(16): 4521-2.

[23] YING H, ZHENG J, JIANG M, et al. Selective cytotoxicity of anhydroicaritin in ER-positive breast cancer via ESR1-mediated MAPK and apoptotic signaling [J]. Chemico-biological interactions, 2026, 423: 111825.

[24] LI T, JING H, GAO X, et al. Identification of key genes as diagnostic biomarkers for IBD using bioinformatics and machine learning [J]. Journal of translational medicine, 2025, 23(1): 738.

[25] ZENG X S, GENG W S, JIA J J. Neurotoxin-Induced Animal Models of Parkinson Disease: Pathogenic Mechanism and Assessment [J]. ASN neuro, 2018, 10: 1759091418777438.

[26] KOLACHEVA A, BANNIKOVA A, PAVLOVA E, et al. Modeling of the Progressive Degradation of the Nigrostriatal Dopaminergic System in Mice to Study the Mechanisms of Neurodegeneration and Neuroplasticity in Parkinson's Disease [J]. International journal of molecular sciences, 2022, 24(1).

[27] JACKSON-LEWIS V, PRZEDBORSKI S. Protocol for the MPTP mouse model of Parkinson's disease [J]. Nature protocols, 2007, 2(1): 141-51.

[28] KUŞçU G C, TUT E, GüREL Ç, et al. Mitochondrial Dynamics-Related Gene Regulation by Epigenetic Suppression of GCN5 Exerts Neuroprotective Effects in Rotenone-Induced Parkinson's Disease Model [J]. Molecular neurobiology, 2026, 63(1): 383.

[29] YOO M, SHIN J, KIM J, et al. DSigDB: drug signatures database for gene set analysis [J]. Bioinformatics (Oxford, England), 2015, 31(18): 3069-71.

[30] LIU Y, YANG X, GAN J, et al. CB-Dock2: improved protein-ligand blind docking by integrating cavity detection, docking and homologous template fitting [J]. Nucleic acids research, 2022, 50(W1): W159-w64.

[31] TONG X, YANG Y, GONG S, et al. Data Mining and Network Pharmacology Analysis of Kidney-Tonifying Herbs on the Treatment of Renal Osteodystrophy Based on the Theory of "Kidney Governing Bones" in Traditional Chinese Medicine [J]. Evidence-based complementary and alternative medicine : eCAM, 2022, 2022: 1116923.

[32] KUTZNER C, KNIEP C, CHERIAN A, et al. GROMACS in the Cloud: A Global Supercomputer to Speed Up Alchemical Drug Design [J]. Journal of chemical information and modeling, 2022, 62(7): 1691-711.

[33] HATHOUT R M, ABDELHAMID S G, EL-HOUSSEINY G S, et al. Comparing cefotaxime and ceftriaxone in combating meningitis through nose-to-brain delivery using bio/chemoinformatics tools [J]. Scientific reports, 2020, 10(1): 21250.

[34] PAPPU M A A, ALAMIN M, NOMAN M A, et al. Multi-Transcriptome-Informed Network Pharmacology Reveals Novel Biomarkers and Therapeutic Candidates for Parkinson's Disease [J]. Genes, 2025, 16(12).

[35] ELDEEB M A, THOMAS R A, RAGHEB M A, et al. Mitochondrial quality control in health and in Parkinson's disease [J]. Physiological reviews, 2022, 102(4): 1721-55.

[36] FRUDD K, BURGOYNE T, BURGOYNE J R. Oxidation of Atg3 and Atg7 mediates inhibition of autophagy [J]. Nature communications, 2018, 9(1): 95.

[37] NODA S, HATTORI N. Unraveling the mystery: How autophagy deficiency in dopaminergic neurons drives human Parkinson's disease [J]. Molecular brain, 2025, 18(1): 66.

[38] ABUDU Y P, PANKIV S, MATHAI B J, et al. NIPSNAP1 and NIPSNAP2 act as "eat me" signals to allow sustained recruitment of autophagy receptors during mitophagy [J]. Autophagy, 2019, 15(10): 1845-7.

[39] GHEZZI D, ARZUFFI P, ZORDAN M, et al. Mutations in TTC19 cause mitochondrial complex III deficiency and neurological impairment in humans and flies [J]. Nature genetics, 2011, 43(3): 259-63.

[40] PIKULEVA I A. Cholesterol-metabolizing cytochromes P450 [J]. Drug metabolism and disposition: the biological fate of chemicals, 2006, 34(4): 513-20.

[41] HUANG X, STERLING N W, DU G, et al. Brain cholesterol metabolism and Parkinson's disease [J]. Movement disorders : official journal of the Movement Disorder Society, 2019, 34(3): 386-95.

[42] SCHOMMER J, MARWARHA G, SCHOMMER T, et al. Retraction Note: 27-Hydroxycholesterol increases α-synuclein protein levels through proteasomal inhibition in human dopaminergic neurons [J]. BMC neuroscience, 2024, 25(1): 70.

[43] DAI L, WANG J, ZHANG X, et al. 27-Hydroxycholesterol Drives the Spread of α-Synuclein Pathology in Parkinson's Disease [J]. Movement disorders : official journal of the Movement Disorder Society, 2023, 38(11): 2005-18.

[44] BJøRKLUND G, BUTNARIU M, CAUNII A, et al. Metallothioneins in Neurodegenerative Diseases: Metal Homeostasis, Autoimmunity, and Therapeutic Potential [J]. Molecular neurobiology, 2026, 63(1): 373.

[45] CARMONA A, CARBONI E, GOMES L C, et al. Metal dyshomeostasis in the substantia nigra of patients with Parkinson's disease or multiple sclerosis [J]. Journal of neurochemistry, 2024, 168(2): 128-41.

[46] YANG W, XU S, ZHOU M, et al. Aging-related biomarkers for the diagnosis of Parkinson's disease based on bioinformatics analysis and machine learning [J]. Aging, 2024, 16(17): 12191-208.

[47] WU J, WU W, JIANG P, et al. Identification of SV2C and DENR as Key Biomarkers for Parkinson's Disease Based on Bioinformatics, Machine Learning, and Experimental Verification [J]. Journal of molecular neuroscience : MN, 2024, 74(1): 6.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryFigure4.png
- SupplementaryFigure1.png
- SupplementaryFigure2.png
- SupplementaryFigure3.png
- SupplementaryTable3.xlsx
- SupplementaryTable4.xlsx
- SupplementaryTable2.xlsx
- SupplementaryTable1.xlsx