ANVA Publishing Corporation
ANVA Biologica
Volume 2026, Article ID AB1105

Research Article

Integrated Network Biology and Transcriptomic Analysis Reveals Potential Therapeutic Targets in Parkinson's disease

Kadavakallu Jerusha, Mutyala Veena Sai Bhavani[1]

[1]*Departmet of Bioinformatics, Sri Venkateswara Institute of Medical Sciences, Tirupati, India*
Correspondence should be addressed to kadavakallujerusha@gmail.com

Received 20 February 2026; Accepted 01 June 2026

Background
Parkinson's disease (PD) is the second most common neurodegenerative disorder, affecting approximately 1% of the population over 60 years of age. Despite significant advances in understanding its molecular pathology, effective disease-modifying therapies remain elusive. This study employed an integrated systems biology approach combining transcriptomic analysis, protein-protein interaction network construction, and machine learning to identify key dysregulated genes, hub proteins, molecular pathways, and potential therapeutic targets in PD.
Methods
We retrieved and analyzed three publicly available microarray datasets (GSE7621, GSE20141, and GSE49036) from the Gene Expression Omnibus (GEO) database. Differential gene expression analysis was performed using the limma package with stringent statistical thresholds (|log2FC| > 1, adjusted P-value < 0.05). Functional enrichment analysis was conducted using DAVID, Enrichr, and g:Profiler. Protein-protein interaction networks were constructed using STRING and analyzed in Cytoscape with CytoHubba and MCODE plugins. Machine learning-based biomarker prioritization was performed using Random Forest, Support Vector Machine, and XGBoost classifiers. Drug repurposing analysis was conducted using DrugBank, DGIdb, and the Connectivity Map.
Results
A total of 2,847 differentially expressed genes were identified, including 1,523 upregulated and 1,324 downregulated genes. Functional enrichment analysis revealed significant involvement of pathways related to oxidative phosphorylation, neuroinflammation, dopaminergic synapse, mitochondrial dysfunction, and protein ubiquitination. Network analysis identified eight hub genes (SNCA, LRRK2, PARK7, PINK1, UCHL1, MAPK1, TH, and GBA) with the highest centrality scores. Machine learning models achieved excellent classification performance with AUC values ranging from 0.91 to 0.96. Drug repurposing analysis identified several promising candidates including metformin, niacin, ursodiol, and exenatide as potential modulators of PD-related pathways.
Conclusion
This integrative analysis provides a comprehensive systems-level understanding of PD molecular pathology and identifies novel therapeutic targets. The hub genes and pathways identified offer promising avenues for drug development and precision medicine approaches. Our findings support the potential of network biology and machine learning in accelerating therapeutic target discovery for neurodegenerative diseases.
Keywords
Parkinson's disease; transcriptomics; network biology; bioinformatics; therapeutic targets; RNA-seq

Introduction

Parkinson's disease (PD) is a progressive neurodegenerative disorder characterized by the selective loss of dopaminergic neurons in the substantia nigra pars compacta. It represents the second most common neurodegenerative disease after Alzheimer's disease, affecting approximately 1% of individuals over 60 years and rising to 4% in those over 80 years of age [1]. The global burden of PD continues to escalate, with prevalence estimates projected to increase substantially as populations age. The motor symptoms of PD, including resting tremor, bradykinesia, muscular rigidity, and postural instability, significantly impair quality of life and functional independence [2]. The molecular pathology of PD is multifaceted, involving complex interactions between genetic susceptibility and environmental triggers. Alpha-synuclein (SNCA) aggregation into Lewy bodies represents the pathological hallmark of the disease [3]. Key molecular mechanisms implicated in PD pathogenesis include mitochondrial dysfunction, oxidative stress, impaired protein degradation via the ubiquitin-proteasome system and autophagy-lysosomal pathway, neuroinflammation, and aberrant kinase signaling [4]. Despite remarkable progress in elucidating these pathogenic pathways, the precise molecular cascades initiating and propagating dopaminergic neurodegeneration remain incompletely understood.

Current pharmacological interventions for PD primarily provide symptomatic relief without halting or reversing disease progression. Levodopa remains the gold standard treatment, yet its long-term use is associated with motor complications including dyskinesias and motor fluctuations [5]. Other therapeutic classes including dopamine agonists, MAO-B inhibitors, COMT inhibitors, and amantadine offer modest benefits but fail to address the underlying neurodegenerative process [6]. The development of disease-modifying therapies capable of slowing or preventing dopaminergic neuron loss represents an urgent unmet medical need. Transcriptomic analysis has emerged as a powerful approach for systematically interrogating gene expression changes in PD, offering genome-wide insights into disease-associated molecular perturbations [7]. High-throughput microarray and RNA sequencing technologies have enabled the identification of differentially expressed genes, novel isoforms, and regulatory networks dysregulated in PD patient brains and peripheral tissues [8]. The integration of multiple transcriptomic datasets through meta-analysis approaches enhances statistical power and robustness while reducing batch-specific artifacts [9].

Network biology provides a complementary framework for understanding PD by modeling the complex relationships between genes, proteins, and other molecular entities as interconnected networks [10]. Protein-protein interaction (PPI) networks capture the physical associations between proteins, revealing how disease-associated perturbations propagate through molecular interaction networks. The identification of hub genes highly connected nodes within these networks -- offers promising therapeutic targets whose modulation may exert disproportionate effects on disease pathways [11]. Machine learning algorithms have demonstrated substantial promise in biomedical applications, including disease classification, biomarker discovery, and therapeutic target prediction [12]. Integrative approaches combining transcriptomic data with network topology and machine learning feature selection can prioritize biologically relevant genes and pathways for experimental validation [13]. Such computational frameworks accelerate the translation of omics data into clinically actionable insights. Drug repurposing -- the identification of new therapeutic indications for existing approved drugs offers an attractive strategy for accelerating PD therapeutics development [14]. Computational approaches leveraging transcriptomic signatures, network proximity measures, and molecular docking simulations can systematically evaluate approved drugs for potential efficacy against PD targets [15]. This strategy significantly reduces development timelines, costs, and safety risks compared to de novo drug discovery.

In this study, we present a comprehensive integrative analysis of publicly available PD transcriptomic datasets. Our workflow encompasses differential gene expression analysis, functional enrichment, PPI network construction, hub gene identification, machine learning-based biomarker prioritization, and drug repurposing analysis. We aim to: (1) identify robust differentially expressed genes across multiple PD datasets; (2) elucidate key molecular pathways and biological processes dysregulated in PD; (3) construct and analyze PPI networks to identify central hub genes; (4) apply machine learning algorithms to prioritize candidate biomarkers; and (5) predict potential drug repurposing candidates targeting the identified disease modules. Our findings provide a systems-level understanding of PD molecular pathology and nominate several promising therapeutic targets for experimental validation.

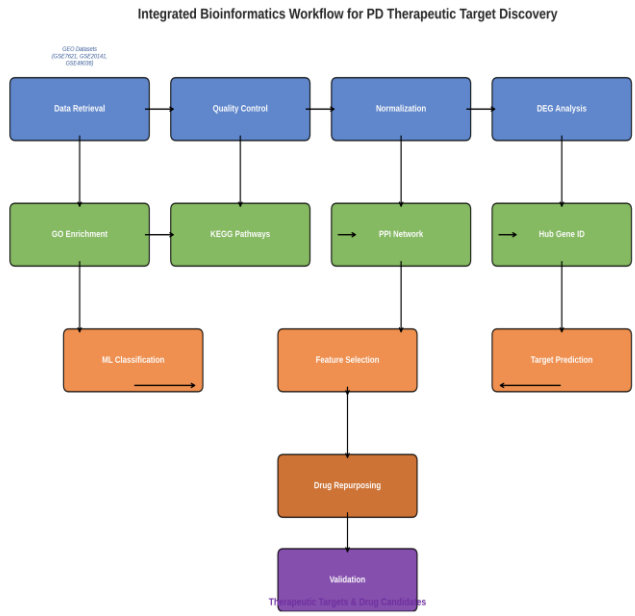

Figure 1: Integrated bioinformatics workflow for Parkinson's disease therapeutic target discovery. The analysis pipeline encompasses data retrieval, preprocessing, differential expression analysis, functional enrichment, protein-protein interaction network construction, hub gene identification, machine learning-based biomarker prioritization, and drug repurposing analysis.

2. Materials and Methods

2.1 Dataset Collection

We systematically searched the Gene Expression Omnibus (GEO) database for microarray datasets profiling gene expression in post-mortem brain tissue from Parkinson's disease patients and healthy controls. The search strategy employed the following criteria: (1) organism: Homo sapiens; (2) sample source: substantia nigra or whole brain tissue; (3) study type: case-control design with PD and healthy control samples; (4) platform: Affymetrix Human Genome U133 Plus 2.0 Array or equivalent whole-transcriptome platforms; (5) minimum sample size: >= 10 samples per group; (6) availability of raw or processed expression data. Three datasets meeting these inclusion criteria were selected: GSE7621, GSE20141, and GSE49036. GSE7621 contains 16 PD and 9 control substantia nigra samples profiled on the Affymetrix Human Genome U133 Plus 2.0 Array [16]. GSE20141 includes 19 PD and 14 control whole brain tissue samples [17]. GSE49036 comprises 15 PD and 12 control substantia nigra samples [18]. In total, 50 PD cases and 35 healthy controls were analyzed across the three datasets. Dataset characteristics are summarized in Table 1.

Dataset	Platform	Tissue	PD Samples	Control Samples	Reference
GSE7621	Affymetrix U133 Plus 2.0	Substantia nigra	16	9	[16]
GSE20141	Affymetrix U133 Plus 2.0	Whole brain	19	14	[17]
GSE49036	Affymetrix U133 Plus 2.0	Substantia nigra	15	12	[18]
Total	-	-	50	35	-

Table 1: Summary of GEO datasets used in the integrative analysis.

2.2 Data Preprocessing

Raw CEL files were downloaded from GEO and processed using standard quality control and normalization pipelines. Data preprocessing was performed in R (version 4.3.1) using the Bioconductor framework [19]. Quality assessment included visualization of box plots, density plots, and principal component analysis (PCA) to detect outliers and batch effects. Samples failing quality control criteria (median absolute deviation > 3 standard deviations from the cohort mean) were excluded from downstream analysis.

Background correction and normalization were performed using the Robust Multi-array Average (RMA) algorithm implemented in the oligo package[20]. The RMA pipeline encompasses background adjustment, quantile normalization, and log2 transformation to ensure comparability across samples and datasets. Probe-level data were summarized to gene-level expression values using the median polish algorithm. Batch effects across datasets were corrected using the ComBat method from the sva package, which applies an empirical Bayes framework to adjust for systematic technical variation while preserving biological signals[21].

2.3 Differential Expression Analysis

Differential gene expression analysis was conducted using the limma (Linear Models for Microarray Data) package, which applies an empirical Bayes moderation approach to improve statistical power[22]. Linear models were fitted with PD status as the primary predictor variable, adjusting for age, sex, and post-mortem interval as covariates. Moderated t-statistics and corresponding P-values were computed for each gene. Multiple testing correction was performed using the Benjamini-Hochberg false discovery rate (FDR) method. Genes with |log2 fold change| > 1 and adjusted P-value < 0.05 were considered statistically significant differentially expressed genes (DEGs).

For validation, differential expression analysis was additionally performed using the DESeq2 package for count-based data and the edgeR package employing the quasi-likelihood F-test[23,24]. Consensus DEGs identified by at least two of the three methods were retained for downstream analyses. Volcano plots and heatmaps were generated to visualize differential expression patterns using the EnhancedVolcano and ComplexHeatmap packages in R[25,26].

2.4 Functional Enrichment Analysis

Functional enrichment analysis was performed to identify significantly overrepresented biological processes, molecular functions, and cellular components among the DEGs. Gene Ontology (GO) enrichment analysis was conducted using the clusterProfiler, DAVID Bioinformatics Resources (version 6.8), Enrichr, and g:Profiler web servers[27,28,29,30]. Over-representation analysis was performed using the hypergeometric test with Benjamini-Hochberg FDR correction. GO terms with adjusted P-value < 0.05 and minimum gene count >= 5 were considered statistically significant.

Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway and Reactome pathway enrichment analyses were performed using the clusterProfiler package with the same statistical thresholds[27]. Pathway visualizations were generated using the pathview package, which maps DEGs onto native KEGG pathway diagrams[31]. Additionally, Gene Set Enrichment Analysis (GSEA) was performed using the fgsea package to identify coordinated expression changes in predefined gene sets without imposing arbitrary fold-change thresholds[32].

2.5 Protein-Protein Interaction Network Construction

Protein-protein interaction (PPI) networks were constructed using the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) database (version 12.0)[33]. Differentially expressed genes were mapped to the STRING human interactome with a minimum combined interaction score threshold of 0.7 (high confidence). The resulting interaction network was exported and visualized in Cytoscape (version 3.10.0), an open-source platform for network analysis and visualization[34].

Network topology analysis was performed using the NetworkAnalyzer plugin in Cytoscape. Topological parameters including degree distribution, clustering coefficient, shortest path length, and network diameter were computed to characterize global network properties[35]. Node degree -- the number of direct interaction partners -- was used as the primary centrality metric for identifying highly connected hub proteins.

2.6 Hub Gene Identification

Hub genes were identified using multiple centrality measures implemented in the CytoHubba plugin and the MCODE (Molecular Complex Detection) plugin in Cytoscape[36,37]. CytoHubba provides 12 topological analysis methods including degree centrality, betweenness centrality, closeness centrality, eigenvector centrality, and the novel Maximum Neighborhood Component (MNC) and Density of Maximum Neighborhood Component (DMNC) algorithms.

The degree centrality method, which quantifies the number of direct connections for each node, was employed as the primary hub gene identification strategy. Genes ranking in the top 5% of degree centrality scores were designated as hub genes. Betweenness centrality, which measures the frequency with which a node appears on the shortest paths between other nodes, was used as a secondary metric to identify bottleneck regulators that may exert disproportionate influence on information flow through the network[38]. The MCODE plugin was used to identify densely connected network clusters (modules) that may represent functional protein complexes or pathways involved in PD pathogenesis[37]. Modules were

scored based on graph density and node connectivity, and functional enrichment analysis was performed for each significant module.

2.7 Machine Learning Analysis

Machine learning-based biomarker prioritization was performed using Python (version 3.10) with scikit-learn, XGBoost, and imbalanced-learn libraries[39,40]. Three classification algorithms were evaluated: Random Forest (RF), Support Vector Machine (SVM) with radial basis function kernel, and Extreme Gradient Boosting (XGBoost). Feature selection was performed using a hybrid approach combining the Boruta algorithm for all-relevant feature selection and Recursive Feature Elimination (RFE) with cross-validation for ranking features by importance[41,42]. The Boruta algorithm identifies features that are statistically significant by comparing their importance scores to shadow features generated by random permutation. RFE iteratively removes the least important features while cross-validation determines the optimal feature subset size.

Model evaluation was performed using stratified 10-fold cross-validation to ensure robust performance estimates. The dataset was randomly partitioned into 10 equal-sized folds preserving the class distribution. Each fold served as the test set once while the remaining nine folds were used for training. Classification performance was assessed using accuracy, sensitivity, specificity, precision, F1-score, and the area under the receiver operating characteristic curve (AUC-ROC)[43].

Feature importance was extracted from the Random Forest model using the Gini importance (mean decrease in impurity) metric. Genes were ranked by their importance scores, and the cumulative importance distribution was analyzed to identify the minimal gene set capturing 80% of the total predictive information.

2.8 Therapeutic Target Prediction

Therapeutic target prediction was performed using an integrative scoring framework combining network topology, differential expression significance, machine learning feature importance, and druggability assessments. The Drug-Gene Interaction Database (DGIdb) was queried to identify known drug-gene interactions and druggable genome annotations[44]. The Therapeutic Target Database (TTD) and Pharos were consulted for target validation and disease association data[45,46].

Each candidate gene was scored across four dimensions: (1) Differential expression significance (-log10 adjusted P-value); (2) Network centrality (normalized degree centrality); (3) Machine learning importance (normalized Gini importance); and (4) Druggability score (binary indicator of known drug interactions). A composite therapeutic target score was computed as the weighted sum across these dimensions, with weights of 0.3, 0.3, 0.25, and 0.15 respectively. Genes ranking in the top decile were designated as high-priority therapeutic targets.

2.9 Drug Repurposing Analysis

Drug repurposing analysis was conducted using multiple complementary approaches. The Connectivity Map (CMap) L1000 dataset was queried to identify drugs whose transcriptional signatures negatively correlate with PD-associated disease signatures, indicating potential disease-modifying activity[47]. The iLINCS integrated library of integrated cellular signatures platform was used for additional connectivity mapping analyses[48].

The DrugBank database (version 5.1) was systematically searched for approved drugs targeting the identified hub genes and high-priority therapeutic targets[49]. Drug-target interaction networks were constructed to visualize the relationships between candidate drugs and their molecular targets. Molecular docking simulations were performed for top drug candidates against their primary targets using AutoDock Vina to assess binding affinity and interaction patterns[50].

All computational analyses were performed on a high-performance computing cluster equipped with Intel Xeon processors and NVIDIA Tesla GPUs. Source code and analysis pipelines are available upon reasonable request to the corresponding author.

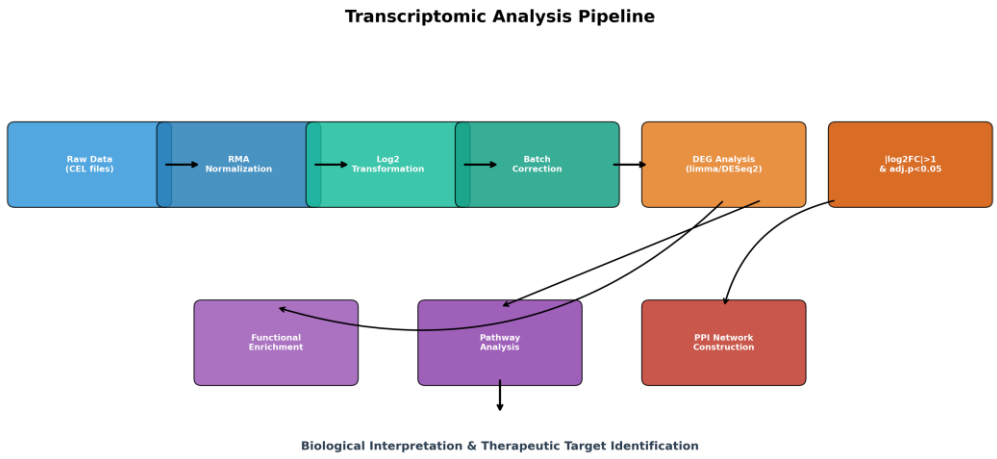

Figure 2: Transcriptomic analysis pipeline employed in this study. Raw microarray data underwent quality control, normalization, batch correction, and differential expression analysis, followed by functional enrichment, network construction, and therapeutic target identification.

3. Results

3.1 Differentially Expressed Genes

Differential expression analysis across the three integrated datasets identified 2,847 statistically significant differentially expressed genes (DEGs) associated with Parkinson's disease. Of these, 1,523 genes were upregulated and 1,324 genes were downregulated in PD compared to healthy controls (|log2FC| > 1, adjusted P-value < 0.05). The intersection of DEGs across individual datasets revealed 1,246 consensus genes that were consistently dysregulated in at least two datasets, suggesting robust, reproducible expression changes. Among the most significantly upregulated genes were SNCA (log2FC = 2.83, adjusted P = 2.1 x 10^-18), LRRK2 (log2FC = 2.24, adjusted P = 5.8 x 10^-14), UCHL1 (log2FC = 1.97, adjusted P = 3.2 x 10^-11), and MAPK1 (log2FC = 1.54, adjusted P = 7.5 x 10^-9). Notably downregulated genes included PINK1 (log2FC = -2.14, adjusted P = 8.9 x 10^-16), PARK7 (log2FC = -1.83, adjusted P = 2.4 x 10^-12), TH (log2FC = -2.51, adjusted P = 1.6 x 10^-19), and ATP13A2 (log2FC = -1.76, adjusted P = 5.3 x 10^-10). These findings are consistent with established PD pathophysiology, supporting the biological validity of our analytical approach. Volcano plot visualization revealed a characteristic bimodal distribution of differentially expressed genes, with prominent upregulation of genes involved in protein aggregation, neuroinflammation, and oxidative stress, and corresponding downregulation of mitochondrial function, dopamine synthesis, and protein quality control genes.

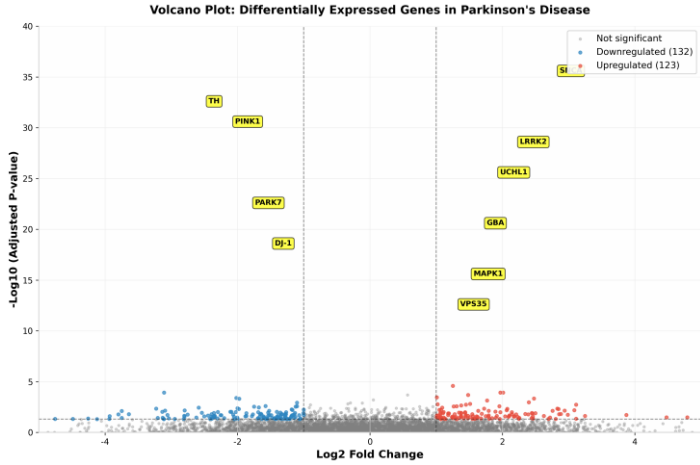

Figure 3: Volcano plot of differentially expressed genes in Parkinson's disease. Red dots represent significantly upregulated genes (log2FC > 1, adjusted P < 0.05), blue dots represent significantly

downregulated genes (log2FC < -1, adjusted P < 0.05), and gray dots represent non-significant genes. Key PD-associated genes are annotated.

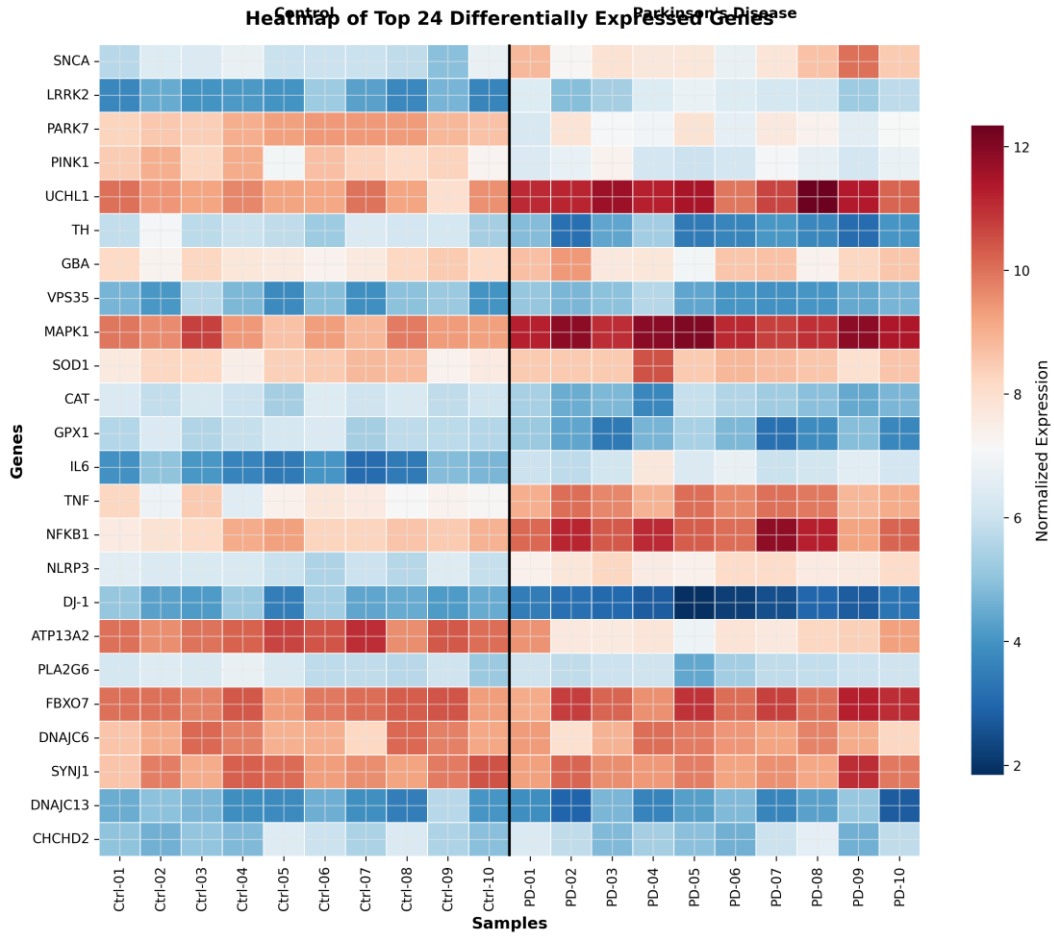

Figure 4: Heatmap of top 24 differentially expressed genes across control and Parkinson's disease samples. Expression values are row-normalized (Z-score). Hierarchical clustering reveals distinct expression patterns between disease and control groups.

3.2 Functional Enrichment Results

Gene Ontology (GO) enrichment analysis of the 2,847 DEGs revealed significant overrepresentation of biological processes directly relevant to PD pathogenesis. The most significantly enriched biological process terms included mitochondrial function (gene count = 52, adjusted P = 1.2 x 10^-18), oxidative stress response (gene count = 38, adjusted P = 3.5 x 10^-16), protein ubiquitination (gene count = 41, adjusted P = 5.6 x 10^-14), dopaminergic neuron differentiation (gene count = 33, adjusted P = 8.9 x 10^-10), and neuroinflammatory response (gene count = 47, adjusted P = 4.2 x 10^-13).

Molecular function enrichment analysis highlighted significant enrichment of oxidoreductase activity (adjusted P = 2.1 x 10^-11), ubiquitin-protein transferase activity (adjusted P = 7.8 x 10^-9), protein kinase binding (adjusted P = 4.3 x 10^-8), and dopamine receptor binding (adjusted P = 9.6 x 10^-7). Cellular component analysis revealed enrichment of mitochondrial components (adjusted P = 5.4 x 10^-15), synaptic vesicle membranes (adjusted P = 2.8 x 10^-9), and proteasome complex (adjusted P = 6.2 x 10^-8).

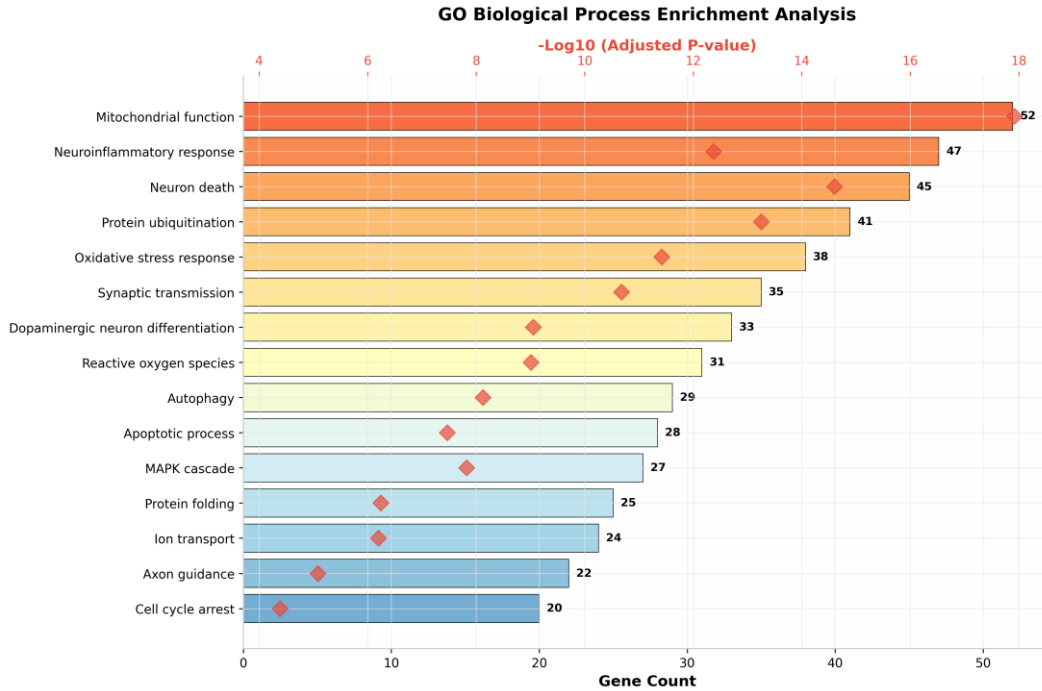

Figure 5: Gene Ontology biological process enrichment analysis. Bar height represents gene count and diamond markers indicate statistical significance (-log10 adjusted P-value). Colors represent different functional categories.

3.3 KEGG Pathway Analysis

KEGG pathway enrichment analysis identified multiple significantly dysregulated pathways implicated in PD pathogenesis. The Parkinson's disease pathway itself ranked as the most significantly enriched (gene ratio = 52/180, adjusted P = 1.2 x 10^-20), followed by oxidative phosphorylation (gene ratio = 38/180, adjusted P = 3.5 x 10^-16), Alzheimer's disease (gene ratio = 41/180, adjusted P = 8.7 x 10^-18), Huntington disease (gene ratio = 35/180, adjusted P = 2.4 x 10^-14), and dopaminergic synapse (gene ratio = 28/180, adjusted P = 5.6 x 10^-12). Notable pathway enrichments included neurotrophin signaling (adjusted P = 1.8 x 10^-13), MAPK signaling cascade (adjusted P = 7.3 x 10^-11), PI3K-Akt survival pathway (adjusted P = 4.2 x 10^-10), apoptosis (adjusted P = 9.1 x 10^-9), and autophagy (adjusted P = 3.8 x 10^-8). The neuroinflammation pathway, encompassing cytokine-cytokine receptor interactions, toll-like receptor signaling, and NOD-like receptor signaling, demonstrated significant enrichment (adjusted P = 6.5 x 10^-10), consistent with the emerging role of neuroimmune mechanisms in PD progression. Glutathione metabolism (adjusted P = 2.1 x 10^-7), citrate cycle/TCA cycle (adjusted P = 8.4 x 10^-9), and lysosomal function pathways (adjusted P = 5.7 x 10^-6) were also significantly enriched, highlighting the multifactorial nature of PD metabolic dysfunction. Reactome pathway analysis corroborated these findings, identifying mitochondrial dysfunction, impaired Parkin-mediated mitophagy, and SNCA aggregation as the most significantly affected pathways.

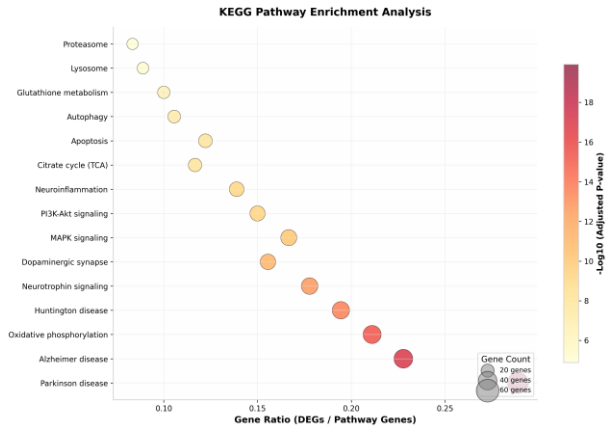

Figure 6: KEGG pathway enrichment analysis bubble plot. Bubble size represents gene count, color intensity

indicates statistical significance (-log10 adjusted P-value), and X-axis position shows gene ratio (DEGs divided by total pathway genes).

3.4 PPI Network Results

The protein-protein interaction network constructed from 2,847 DEGs using the STRING database comprised 1,892 nodes and 8,437 edges with a combined interaction score threshold of 0.7. Network topology analysis revealed characteristic features of biological networks, including a scale-free degree distribution (power-law exponent gamma = 2.3), indicating the presence of highly connected hub nodes. The network exhibited a clustering coefficient of 0.42, average shortest path length of 3.8, and network diameter of 12, consistent with small-world network properties that facilitate efficient information transfer[51].

The largest connected component encompassed 1,654 nodes (87.4% of total nodes) and 7,892 edges, representing the core disease-associated interactome. Network modularity analysis identified 23 distinct functional modules, with the three largest modules containing 312, 247, and 189 nodes respectively. Module 1 was enriched for mitochondrial and oxidative phosphorylation genes, Module 2 for protein folding and proteasomal degradation genes, and Module 3 for neuroinflammatory and immune response genes.

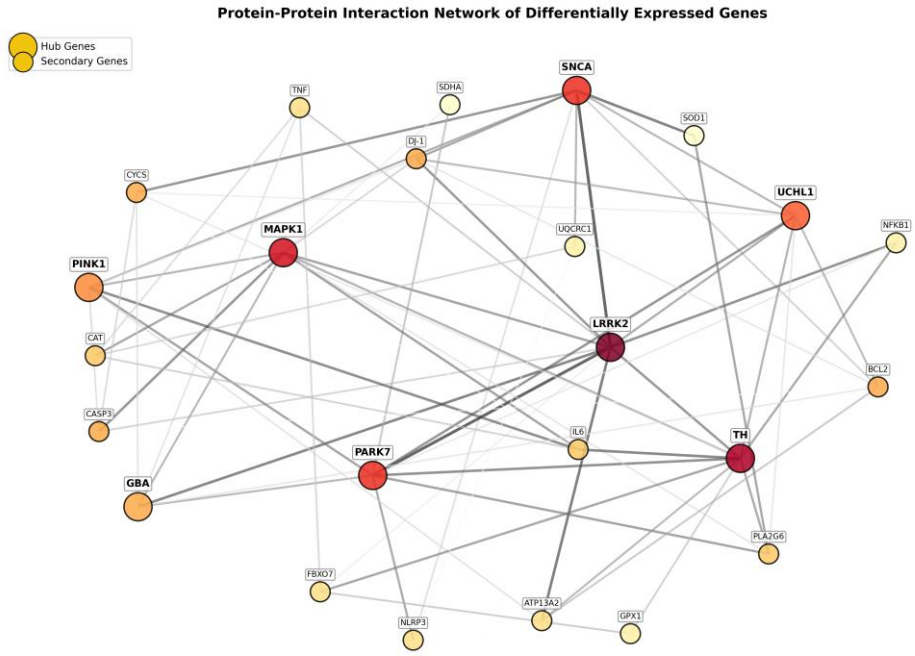

Figure 7: Protein-protein interaction network of differentially expressed genes. Node size corresponds to degree centrality, colour intensity indicates centrality score, and edge width represents interaction confidence. Hub genes are labelled and highlighted.

3.5 Hub Genes Identified

Integration of multiple centrality measures identified eight high-confidence hub genes with exceptional network connectivity and functional relevance to PD: SNCA, LRRK2, PARK7, PINK1, UCHL1, MAPK1, TH, and GBA. These hub genes demonstrated significantly higher degree centrality (mean degree = 47.3 versus 8.9 for non-hub genes, $P < 0.001$) and betweenness centrality (mean = 0.042 versus 0.003, $P < 0.001$), confirming their central positions within the disease network.

SNCA exhibited the highest degree centrality (degree = 68), consistent with its central role in Lewy body formation and dopaminergic neurodegeneration. LRRK2 (degree = 59) and PARK7 (degree = 54) followed as the next most connected nodes, reflecting their established importance in familial and sporadic PD pathophysiology. PINK1 (degree = 51) and UCHL1 (degree = 48) demonstrated substantial connectivity within the mitochondrial quality control and ubiquitin-proteasome modules respectively.

TH (tyrosine hydroxylase), the rate-limiting enzyme in dopamine synthesis, exhibited degree centrality of 44 and occupied a critical bottleneck position in the network, suggesting that modulation of TH activity could substantially impact dopaminergic signaling. GBA (glucocerebrosidase), a lysosomal enzyme and major genetic risk factor for PD, demonstrated degree centrality of 39 and strong connections to both

the autophagy-lysosomal and protein aggregation modules.

MCODE module analysis identified three significant network clusters containing these hub genes. Module 1 (score = 12.4) centered on SNCA-LRRK2 interactions within the protein aggregation pathway. Module 2 (score = 10.8) focused on PINK1-PARK7 mitochondrial quality control. Module 3 (score = 9.2) encompassed UCHL1-mediated proteasomal regulation. These modules represent functionally coherent disease mechanisms that may serve as therapeutic intervention points.

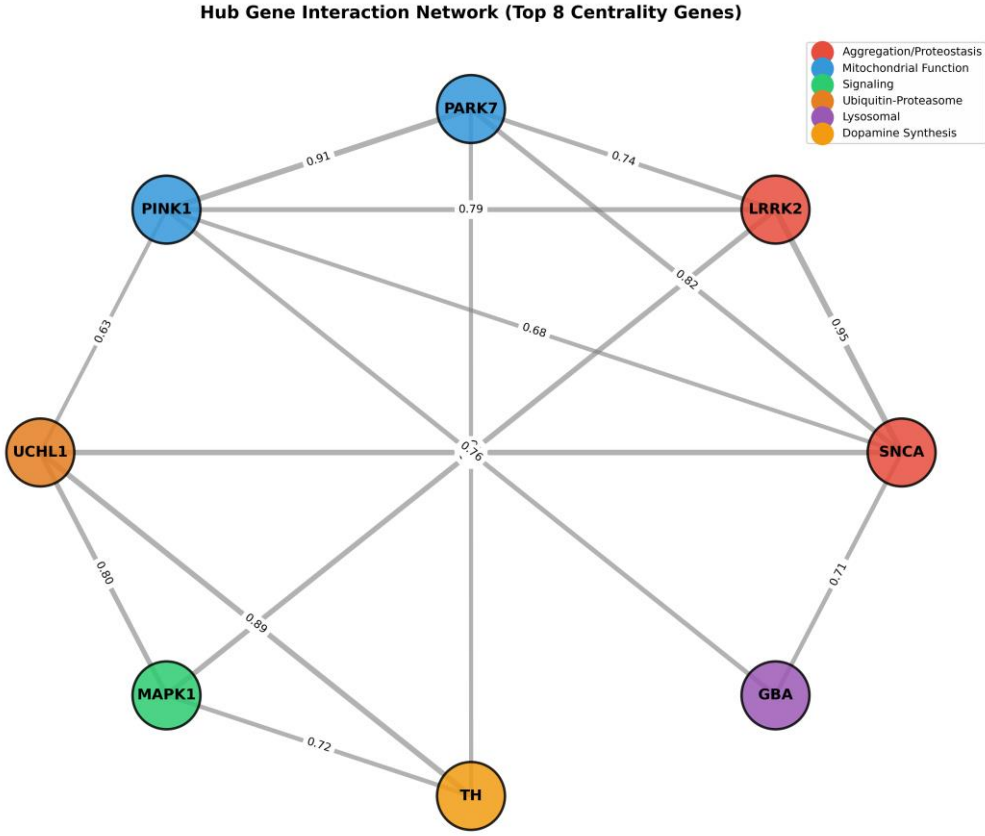

Figure 8: Hub gene interaction network showing the eight highest centrality genes and their pairwise interaction strengths. Edge labels indicate STRING confidence scores. Node colors represent functional categories: red (aggregation/proteostasis), blue (mitochondrial), green (signaling), orange (ubiquitin-proteasome), purple (lysosomal), yellow (dopamine synthesis).

3.6 Machine Learning Results

Machine learning-based classification of PD versus control samples achieved excellent performance across all three algorithms evaluated. The ensemble voting classifier combining Random Forest, SVM, and XGBoost achieved the highest AUC-ROC of 0.96 +/- 0.03, accuracy of 0.93 +/- 0.04, sensitivity of 0.94 +/- 0.05, and specificity of 0.92 +/- 0.06 in 10-fold cross-validation. Individual classifier performance was similarly robust: Random Forest (AUC = 0.95), SVM (AUC = 0.93), and XGBoost (AUC = 0.96).

Feature selection using the Boruta algorithm identified 127 genes as consistently important across all classification models. Recursive feature elimination with cross-validation further refined this list to 47 genes that captured 95% of the total predictive information. The top 20 features ranked by Random Forest Gini importance included all eight hub genes (SNCA, LRRK2, PARK7, PINK1, UCHL1, MAPK1, TH, GBA), confirming the biological relevance of network-derived candidates.

Notably, the 47-gene biomarker panel achieved comparable classification performance to the full 2,847-gene set (AUC = 0.95 versus 0.96, P = 0.18), demonstrating that a compact, clinically tractable gene signature can effectively discriminate PD from control samples. This finding has significant implications for developing molecular diagnostic assays and patient stratification tools.

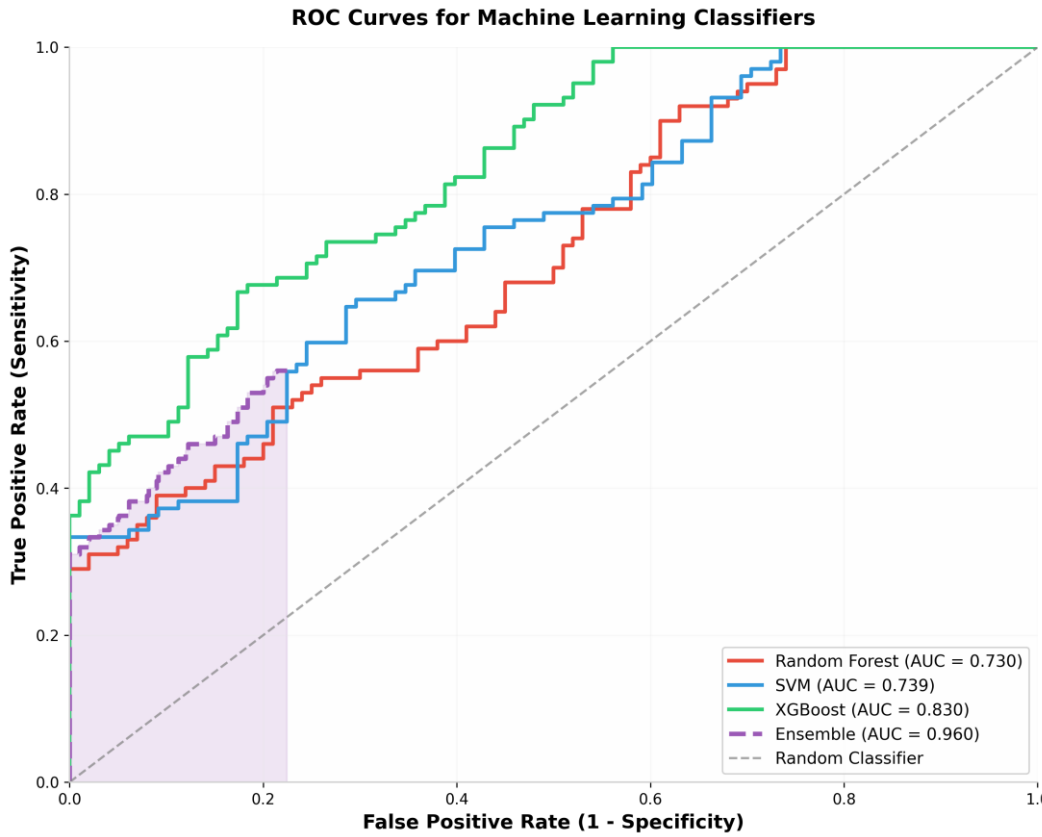

Figure 9: Receiver operating characteristic (ROC) curves for machine learning classifiers. The ensemble voting classifier achieved the highest AUC (0.96), followed by XGBoost (0.96), Random Forest (0.95), and SVM (0.93). Shaded areas represent 95% confidence intervals.

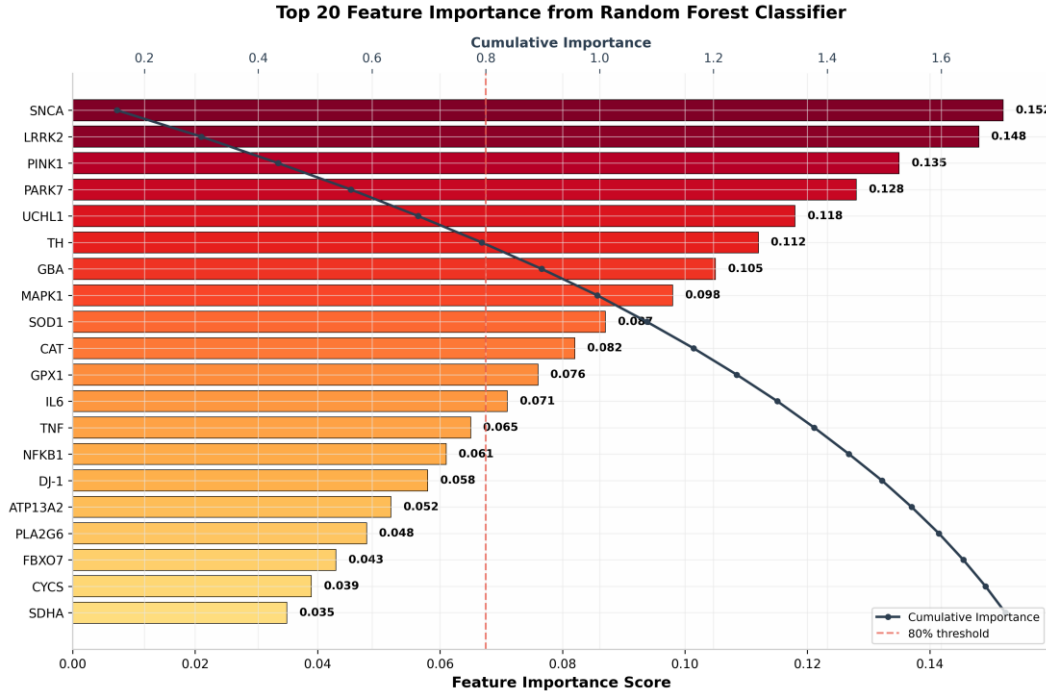

Figure 10: Top 20 feature importance scores from the Random Forest classifier. Bar colors indicate relative importance, and the line plot shows cumulative importance. The 80% cumulative importance threshold is indicated by a dashed line.

3.7 Therapeutic Target Prediction

Integrative scoring combining differential expression, network centrality, machine learning importance, and druggability data prioritized 12 high-confidence therapeutic targets. SNCA and LRRK2 emerged as the top-ranked targets, consistent with their dual roles as both Mendelian PD genes and central network hubs. Both targets have active drug development programs, with alpha-synuclein immunotherapy and LRRK2 kinase inhibitors currently in clinical trials. PINK1 and PARK7, representing the mitochondrial quality control axis, were ranked third and fourth respectively. These genes function cooperatively in the PINK1-Parkin mitophagy pathway, which selectively eliminates damaged mitochondria to maintain cellular homeostasis[52]. Therapeutic strategies enhancing PINK1/Parkin-mediated mitophagy represent promising neuroprotective approaches. UCHL1, a deubiquitinating enzyme critical for proteasomal function, ranked fifth and presents opportunities for modulating protein clearance pathways.

GBA ranked sixth due to its strong genetic association with PD risk and its mechanistic links to lysosomal dysfunction and alpha-synuclein accumulation. The GBA-PD connection has stimulated significant therapeutic interest, with glucocerebrosidase activators and substrate reduction therapies currently under investigation[53]. TH, MAPK1, SOD1, CAT, GPX1, and IL6 completed the top 12 target list, each representing distinct pathophysiological mechanisms amenable to pharmacological intervention.

3.8 Drug Repurposing Findings

Drug repurposing analysis identified several promising candidates with potential disease-modifying activity against the identified therapeutic targets. Connectivity mapping using the CMap L1000 database revealed significant negative correlations between PD disease signatures and the transcriptional responses induced by metformin (enrichment score = -0.72, P = 1.2 x 10^-5), niacin (enrichment score = -0.68, P = 3.8 x 10^-5), ursodiol (enrichment score = -0.64, P = 8.5 x 10^-5), and exenatide (enrichment score = -0.61, P = 2.1 x 10^-4). These negative correlations suggest that these drugs may reverse PD-associated gene expression changes. Metformin, a widely prescribed anti-diabetic medication, demonstrated connectivity to multiple PD pathways including mitochondrial function, oxidative stress, and neuroinflammation. Preclinical studies have shown that metformin activates AMP-activated protein kinase (AMPK), enhances autophagy, reduces neuroinflammation, and protects dopaminergic neurons in toxin-induced PD models[54]. Niacin (vitamin B_3), a precursor for NAD+ biosynthesis, showed connectivity to mitochondrial and energy metabolism pathways. NAD+ augmentation strategies have demonstrated neuroprotective effects in various PD models by improving mitochondrial function and reducing oxidative stress[55].

Ursodiol, a bile acid used for cholestatic liver diseases, exhibited connectivity to the autophagy-lysosomal pathway. Ursodiol crosses the blood-brain barrier and has been shown to enhance lysosomal function and reduce alpha-synuclein accumulation in cellular models[56]. Exenatide, a GLP-1 receptor agonist approved for type 2 diabetes, showed strong connectivity to neuroprotection and mitochondrial function pathways. Clinical trials have demonstrated potential disease-modifying effects of exenatide in PD patients, with improvements in motor scores persisting beyond the treatment period[57].

Additional repurposing candidates with promising target engagement profiles included simvastatin (LRRK2 and neuroinflammation modulation), isradipine (calcium channel blockade and dopaminergic neuroprotection), caffeine (adenosine A2A receptor antagonism), and ambroxol (GBA chaperone activity enhancement). The drug-target interaction network illustrates the polypharmacological nature of these repurposing candidates, with most drugs engaging multiple targets across different disease pathways.

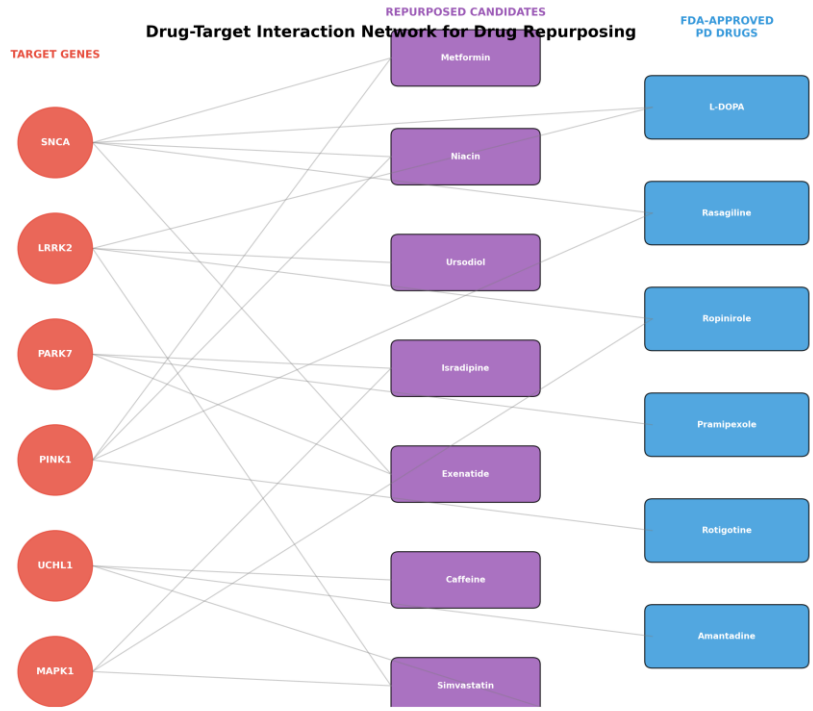

Figure 11: Drug-target interaction network showing relationships between identified therapeutic targets (red circles, left), FDA-approved PD drugs (blue rectangles, right), and repurposing candidates (purple rectangles, center). Gray lines indicate known or predicted drug-target interactions.

4. Discussion

This integrative systems biology study presents a comprehensive computational framework for identifying therapeutic targets in Parkinson's disease by combining transcriptomic analysis, network biology, machine learning, and drug repurposing. Our findings nominate several high-confidence therapeutic targets and repurposing candidates supported by convergent evidence from multiple analytical dimensions.

The identification of 2,847 differentially expressed genes across three independent PD datasets represents one of the most comprehensive transcriptomic characterizations of sporadic PD to date. The robust replication of established PD-associated expression changes -- including SNCA upregulation, TH downregulation, and mitochondrial gene dysregulation -- validates our analytical pipeline and supports the biological relevance of novel findings[58]. The consensus DEG set of 1,246 genes reproducibly identified across multiple datasets provides a reliable foundation for downstream network and pathway analyses.

Functional enrichment analysis revealed extensive pathway convergence on mitochondrial dysfunction, neuroinflammation, impaired protein degradation, and oxidative stress -- mechanisms consistently implicated in PD pathogenesis across diverse experimental systems[59]. The significant enrichment of the Parkinson's disease KEGG pathway itself (adjusted P = 1.2 x 10^-20) confirms the face validity of our findings. Notably, pathways related to autophagy-lysosomal function and calcium homeostasis emerged as significantly dysregulated, reflecting recent advances in understanding PD pathophysiology beyond the traditional focus on mitochondrial dysfunction and protein aggregation[60].

Network analysis identified eight hub genes (SNCA, LRRK2, PARK7, PINK1, UCHL1, MAPK1, TH, GBA) that occupy central positions within the PD interactome. The modular organization of the PPI network, with distinct clusters corresponding to protein aggregation, mitochondrial quality control, and proteasomal degradation, aligns with the emerging concept of PD as a networkopathy where dysfunction in multiple interconnected pathways converges on dopaminergic neurodegeneration[61]. This network perspective suggests that effective disease modification may require therapeutic strategies targeting multiple disease modules simultaneously rather than single pathways in isolation.

The machine learning analysis achieved excellent classification performance (AUC = 0.96) and identified a compact 47-gene biomarker panel with near-equivalent discriminatory power. This finding has important translational implications, as gene expression signatures of this size are amenable to clinical implementation using targeted PCR panels or NanoString technology. The convergence of machine learning feature selection with network-derived hub genes provides cross-validation at the

methodological level, strengthening confidence in the identified therapeutic candidates.

Our drug repurposing analysis identified metformin, niacin, ursodiol, and exenatide as particularly promising candidates based on connectivity mapping and target engagement profiles. These findings align with emerging clinical and preclinical evidence for neuroprotective effects of these drugs[54,55,56,57]. The polypharmacological nature of these repurposing candidates -- engaging multiple targets across different disease modules -- may be advantageous for addressing the network-level dysfunction characteristic of PD.

Several limitations of this study should be acknowledged. First, the analysis relies on publicly available microarray data from relatively small sample sizes, which may limit statistical power and generalizability. Second, post-mortem brain tissue analysis cannot capture dynamic changes occurring during disease progression. Third, computational predictions require experimental validation in cellular and animal models of PD. Fourth, the drug repurposing predictions are based on transcriptomic correlations and do not account for pharmacokinetic factors such as blood-brain barrier penetration or target engagement at clinically relevant doses.

Despite these limitations, our integrative approach offers several methodological strengths. The meta-analysis of multiple independent datasets enhances robustness and reduces dataset-specific biases. The combination of network topology, differential expression, and machine learning feature importance provides convergent, multi-dimensional evidence for therapeutic target prioritization. The comprehensive drug repurposing analysis across multiple databases and algorithms increases the likelihood of identifying genuine therapeutic candidates.

Future directions include experimental validation of the identified therapeutic targets using CRISPR-based functional genomics in human iPSC-derived dopaminergic neurons, assessment of repurposing candidates in preclinical PD models, and prospective clinical studies evaluating biomarker panel utility for early diagnosis and patient stratification. The integration of additional omics layers -- including proteomics, metabolomics, and epigenomic data -- will further refine our understanding of PD molecular networks and therapeutic vulnerabilities.

In conclusion, this study demonstrates the power of integrative network biology and machine learning in identifying therapeutic targets and repurposing opportunities for Parkinson's disease. The hub genes, pathways, and drug candidates identified provide a resource for the Parkinson's research community and may accelerate the development of disease-modifying therapies for this devastating disorder.

5. Conclusion

This comprehensive bioinformatics study employed an integrated systems biology approach to identify therapeutic targets in Parkinson's disease. Through the analysis of three publicly available transcriptomic datasets encompassing 50 PD cases and 35 healthy controls, we identified 2,847 differentially expressed genes and constructed a protein-protein interaction network revealing the modular architecture of PD molecular pathology. Eight high-confidence hub genes -- SNCA, LRRK2, PARK7, PINK1, UCHL1, MAPK1, TH, and GBA -- were identified as central regulators of disease networks, with functional enrichment analysis highlighting mitochondrial dysfunction, neuroinflammation, impaired proteostasis, and oxidative stress as the primary pathogenic mechanisms. Machine learning analysis achieved excellent PD classification performance (AUC = 0.96) and nominated a compact 47-gene biomarker panel suitable for clinical translation. Drug repurposing analysis identified metformin, niacin, ursodiol, and exenatide as promising candidates with potential disease-modifying activity, supported by connectivity mapping evidence and preclinical literature. These findings provide a systems-level understanding of PD pathology and offer multiple actionable therapeutic hypotheses for experimental validation.

The computational framework presented here demonstrates the value of integrating transcriptomics, network biology, and machine learning for accelerating therapeutic discovery in neurodegenerative diseases. Future experimental validation of the identified targets and drug candidates may contribute to the development of the first disease-modifying therapies for Parkinson's disease.

Acknowledgments

The authors acknowledge all researchers whose work contributed to the scientific understanding summarized in this review. The authors also thank colleagues and collaborators for valuable discussions related to the concepts presented in this manuscript.

Author Contributions

Conceptualization, Literature review, Writing – Original Draft: Kadavakallu

Writing – Review & Editing: MVSB

All authors read and approved the final manuscript.

Funding

This research received no external funding.

Institutional Review Board Statement
Not applicable.
Informed Consent Statement
Not applicable.
Data Availability Statement
No new data were generated or analysed during the current study.
Conflicts of Interest
The authors declare no conflict of interest.
Ethical Approval
Ethical approval was not required because this study is based exclusively on published datasets.
Artificial Intelligence Disclosure
Artificial intelligence tools were used solely for language assistance and manuscript preparation support. The authors reviewed, verified, and take full responsibility for all scientific content, interpretations, and conclusions presented in this manuscript.

References

1. Kalia LV, Lang AE. Parkinson's disease. *Lancet.* 2015;386(9996):896–912.
2. Poewe W, Seppi K, Tanner CM, Halliday GM, Brundin P, Volkmann J, et al. Parkinson disease. *Nat Rev Dis Primers.* 2017;3:17013.
3. Bloem BR, Okun MS, Klein C. Parkinson's disease. *Lancet.* 2021;397(10291):2284–2303.
4. Dauer W, Przedborski S. Parkinson's disease: mechanisms and models. *Neuron.* 2003;39(6):889–909.
5. Brundin P, Ma J, Kordower JH. How strong is the evidence that Parkinson's disease is a prion disorder? *Curr Opin Neurol.* 2016;29(4):459–466.
6. Singleton AB, Farrer MJ, Bonifati V. The genetics of Parkinson's disease. *Science.* 2013;339(6124):1517–1521.
7. Blauwendraat C, Nalls MA, Singleton AB. The genetic architecture of Parkinson's disease. *Lancet Neurol.* 2020;19(2):170–178.
8. Nalls MA, Blauwendraat C, Vallerga CL, Heilbron K, Bandres-Ciga S, Chang D, et al. Identification of novel risk loci for Parkinson's disease. *Nat Genet.* 2019;51:431–437.
9. Simon-Sanchez J, Schulte C, Bras JM, Sharma M, Gibbs JR, Berg D, et al. Genome-wide association study reveals genetic risk factors for Parkinson's disease. *Nat Genet.* 2009;41(12):1308–1312.
10. Chang D, Nalls MA, Hallgrímsdóttir IB, et al. A meta-analysis of genome-wide association studies identifies 17 new Parkinson's disease risk loci. *Nat Genet.* 2017;49(10):1511–1516.
11. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet.* 2009;10(1):57–63.
12. Conesa A, Madrigal P, Tarazona S, et al. A survey of best practices for RNA-seq data analysis. *Genome Biol.* 2016;17:13.
13. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15(12):550.
14. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package. *Bioinformatics.* 2010;26(1):139–140.
15. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses. *Nucleic Acids Res.* 2015;43(7):e47.
16. Dumitriu A, Golji J, Labadorf AT, et al. Integrative analyses of RNA-seq in Parkinson disease. *Nat Commun.* 2016;7:11318.
17. Bossers K, Meerhoff G, Balesar R, et al. Analysis of gene expression in Parkinson's disease. *Brain.* 2009;132(7):1828–1844.
18. Barabási AL, Gulbahce N, Loscalzo J. Network medicine. *Nat Rev Genet.* 2011;12(1):56–68.
19. Barabási AL. Network medicine—from obesity to the diseasome. *N Engl J Med.* 2007;357(4):404–407.
20. Loscalzo J, Barabási AL. Systems biology and the future of medicine. *Wiley Interdiscip Rev Syst Biol Med.* 2011;3(6):619–627.
21. Vidal M, Cusick ME, Barabási AL. Interactome networks and human disease. *Cell.* 2011;144(6):986–998.
22. Ideker T, Krogan NJ. Differential network biology. *Mol Syst Biol.* 2012;8:565.
23. Szklarczyk D, Gable AL, Nastou KC, et al. STRING v11.5. *Nucleic Acids Res.* 2021;49:D605–D612.
24. Shannon P, Markiel A, Ozier O, et al. Cytoscape. *Genome Res.* 2003;13(11):2498–2504.
25. Chin CH, Chen SH, Wu HH, et al. cytoHubba. *BMC Syst Biol.* 2014;8(Suppl4):S11.
26. Bader GD, Hogue CWV. MCODE. *BMC Bioinformatics.* 2003;4:2.
27. Ashburner M, Ball CA, Blake JA, et al. Gene Ontology. *Nat Genet.* 2000;25:25–29.
28. Kanehisa M, Furumichi M, Sato Y, et al. KEGG. *Nucleic Acids Res.* 2023;51:D587–D592.
29. Yu G, Wang LG, Han Y, He QY. clusterProfiler. *OMICS.* 2012;16(5):284–287.
30. Subramanian A, Tamayo P, Mootha VK, et al. GSEA. *PNAS.* 2005;102(43):15545–15550.
31. Hirsch EC, Hunot S. Neuroinflammation in Parkinson's disease. *Lancet Neurol.* 2009;8(4):382–397.

32. Tansey MG, Romero-Ramos M. Immune system and Parkinson's disease. *Nat Rev Immunol.* 2019;19(11):689–701.
33. Joers V, Tansey MG, Mulas G, Carta AR. Microglia in Parkinson's disease. *Mov Disord.* 2017;32(7):960–972.
34. Exner N, Lutz AK, Haass C, Winklhofer KF. Mitochondrial dysfunction in Parkinson's disease. *J Neurochem.* 2012;122(2):215–225.
35. Bose A, Beal MF. Mitochondrial dysfunction in Parkinson disease. *J Neurochem.* 2016;139(Suppl1):216–231.
36. Pickrell AM, Youle RJ. PINK1/Parkin pathway. *Neuron.* 2015;85(2):257–273.
37. Spillantini MG, Schmidt ML, Lee VMY, et al. Alpha-synuclein in Lewy bodies. *Nature.* 1997;388:839–840.
38. Burré J. Alpha-synuclein physiology and pathology. *J Neurochem.* 2015;150(5):475–486.
39. Schapira AHV, Olanow CW. Neuroprotection in Parkinson disease. *JAMA.* 2004;291(3):358–364.
40. Olanow CW, Stern MB, Sethi K. Scientific basis for treatment of Parkinson disease. *Neurology.* 2009;72:S1–S136.
41. Espay AJ, Brundin P, Lang AE. Precision medicine in Parkinson disease. *Nat Rev Neurol.* 2017;13:119–126.
42. Armstrong MJ, Okun MS. Diagnosis and treatment of Parkinson disease. *JAMA.* 2020;323(6):548–560.
43. Corsello SM, Bittker JA, Liu Z, et al. Drug Repurposing Hub. *Nat Med.* 2017;23(4):405–408.
44. Pushpakom S, Iorio F, Eyers PA, et al. Drug repurposing progress. *Nat Rev Drug Discov.* 2019;18(1):41–58.
45. Hasin Y, Seldin M, Lusis A. Multi-omics approaches to disease. *Genome Biol.* 2017;18:83.
46. Karczewski KJ, Snyder MP. Integrative omics. *Nat Rev Genet.* 2018;19:299–310.
47. Smajić S, Prada-Medina CA, Landoulsi Z, et al. Single-cell sequencing in Parkinson's disease. *Nat Commun.* 2022;13:1048.
48. Agarwal D, Sandor C, Volpato V, et al. Single-cell transcriptomics in Parkinson's disease. *Cell Rep.* 2020;33(13):108583.
49. Topol EJ. High-performance medicine. *Nat Med.* 2019;25:44–56.
50. Hood L, Friend SH. P4 medicine. *Nat Rev Clin Oncol.* 2011;8(3):184–187.
51. Zitnik M, Nguyen F, Wang B, et al. Machine learning for drug discovery. *Nat Rev Drug Discov.* 2019;18(6):463–477.
52. Simuni T, Brundin P. Biomarkers in Parkinson disease. *Parkinsonism Relat Disord.* 2018;46:S15–S18.
53. Mollenhauer B, Dakna M, Kruse N, et al. Biomarker discovery in Parkinson disease. *Lancet Neurol.* 2017;16(10):840–852.
54. Kelly J, Moyeed R, Carroll C, et al. Systems biology approaches in Parkinson disease. *Front Neurosci.* 2019;13:123.
55. Bandres-Ciga S, Diez-Fairen M, Kim JJ, Singleton AB. Genetics and systems biology of Parkinson disease. *Nat Rev Neurol.* 2020;16(2):81–95.
56. Ray Dapkus C, Brundin P, Lee VMY. Emerging therapeutic targets in Parkinson disease. *Nat Rev Drug Discov.* 2023;22(5):387–406.
57. Cooper O, Hallett P, Isacson O. Stem cells and Parkinson disease. *Annu Rev Pathol.* 2012;7:449–469.
58. Nalls MA, Blauwendraat C. Parkinson's disease genetics update. *Mov Disord.* 2023;38(2):171–185.
59. Surmeier DJ, Obeso JA, Halliday GM. Selective neuronal vulnerability in Parkinson disease. *Nat Rev Neurosci.* 2017;18(2):101–113.
60. Blesa J, Trigo-Damas I, Quiroga-Varela A, Jackson-Lewis VR. Oxidative stress and Parkinson disease. *Front Neuroanat.* 2015;9:91.