Plasma Proteomics for Parkinson's Disease: Diagnostic Classification, Severity Association, and Therapeutic Hypotheses

Nicholas Minster[1*] and M. Saleet Jafri[1,2]

[1]School of Systems Biology, George Mason University, 4400 University Drive, Fairfax, 22030, Virginia, United States.
[2]Krasnow Institute for Advanced Study and Interdisciplinary Program in Neuroscience, George Mason University, 4400 University Drive, Fairfax, 22030, Virginia, United States.

*Corresponding author(s). E-mail(s): nminster@gmu.edu;
Contributing authors: sjafri@gmu.edu;

Abstract

Background: Although there is no cure, early diagnosis of Parkinson's disease allows effective management and symptomatic relief, potentially delaying the need for more potent medications. Blood-based biomarkers can facilitate early detection, symptom monitoring, and targeted therapy.
Methods: Gene expression and plasma proteomics data from two Parkinson's disease cohorts were integrated. Machine learning models were trained to classify disease status using protein, gene expression, and combined datasets. A severity score was derived by regressing protein levels against clinical motor ratings and tested longitudinally. Enrichment and network analyses identified biological context, and drug perturbation databases were queried for candidate therapies.
Results: Proteomic models outperformed gene-based approaches and generalized well to external data. The severity score correlated with clinical burden and predicted future progression. Enriched pathways involved extracellular signaling, immune response, and post-translational regulation. Several compounds were identified as potential therapeutic candidates based on network targeting and reversal potential.
Conclusions: Peripheral proteomic signatures offer classification, progression, and therapeutic insights in Parkinson's disease, supporting their biological relevance.

Keywords: Parkinson's Disease, Proteomics, Biomarkers, Machine Learning, Peripheral Blood, Disease Severity, Patient Stratification, Drug Repurposing, Network Analysis, Extracellular Signaling

1 Background

Parkinson's disease (PD) represents a looming public health crisis, distinguished not only by its rising global burden but also by the persistent gap between biological onset and symptom onset. Parkinson's Disease is now recognized as the fastest-growing neurodegenerative disorder. Diagnosis relies on motor signs that appear after substantial neurodegeneration has occurred [1, 2]. This delay limits opportunities for disease-modifying therapy. At the molecular level, PD is characterized by the progressive loss of dopaminergic neurons in the substantia nigra pars compacta and the pathological aggregation of α-synuclein into Lewy bodies [3]. Closing the diagnostic gap is a central challenge in Parkinson's disease research [4].

Diagnosis of Parkinson's disease (PD) relies on taking a patient history and neurological exam [5, 6]. Despite active biomarker research, available tests are limited in routine care. Advanced imaging such as DaTScan or high-field MRI is constrained by access and cost [7]. Emerging assays such as CSF α-synuclein seeding remain confined to research settings and are suitable only as a confirmation test [8, 9].

Monogenic forms of PD, although relatively rare, provide critical insight into the molecular pathways disrupted in both familial and sporadic cases. Mutations in GBA1 and LRRK2 together account for approximately 10–15% of cases and contribute to PD pathology through distinct yet intersecting mechanisms [10]. GBA1 mutations compromise lysosomal glucocerebrosidase activity, impairing autophagic clearance and promoting α-synuclein accumulation [11]. LRRK2 gain-of-function mutations alter vesicular trafficking and mitochondrial dynamics via dysregulated kinase signaling [12]. These disruptions converge on lysosomal-autophagic dysfunction and mitochondrial stress, hallmarks increasingly recognized across genetic and idiopathic PD [13]. Studying these mutations has revealed broader disease mechanisms and shared molecular targets for early diagnosis and intervention.

Within this landscape, the circulating secretome, the set of proteins and molecules secreted into the bloodstream, has emerged as a promising biomarker source. It reflects cellular stress, neuroinflammation, and systemic signaling changes that may occur before clinical symptoms arise.

Current therapies are palliative [14]. Levodopa and adjunctive dopaminergic agents alleviate symptoms but fail to halt disease progression and often lead to long-term motor complications [15, 16]. Invasive interventions such as deep brain stimulation offer motor symptom relief in select patients [17]. Multiple recent trials targeting α-synuclein aggregation and neuroinflammation have failed to demonstrate disease-modifying effects, underscoring the need for upstream biomarkers and therapeutic targets [18].

In parallel, lifestyle interventions such as physical exercise are gaining recognition as low-risk strategies that may confer neuroprotective benefits. Exercise has been

shown to modulate DNA methylation patterns associated with aging and neurodegeneration, potentially reversing aspects of the epigenetic clock [19]. Exercise also enhances neuroplasticity, reduces neuroinflammation, and engages molecular pathways related to autophagy, mitochondrial function, and immune regulation, processes central to PD pathogenesis. These molecular effects position physical activity not only as a therapeutic adjunct but also as a window into modifiable disease mechanisms that precede clinical onset.

High-throughput transcriptomic and proteomic platforms enable comprehensive profiling of RNA and protein species, including secreted factors, from small plasma volumes. When paired with machine learning and network biology, these datasets can yield high diagnostic accuracy and identify molecular patterns predictive of severity and progression [20]. However, many studies lack rigorous cross-validation or external replication, raising concerns about generalizability [21, 22]. Interpretable machine learning tools like SHapley Additive exPlanations (SHAP) and Integrated Gradients offer a path forward by linking predictive features to known biology. Integrative multi-omics modeling may further enable the capture of complex molecular phenotypes [23]. These steps translate statistical signals into actionable biological insight by deriving blood-based biomarkers for (i) diagnostic classification (PD vs healthy control), (ii) severity monitoring via a Proteomic Severity Index (PSI), and (iii) hypothesis-generating targets to prioritize therapeutic mechanisms for follow-up.

Study framing and departures from prior work

This work builds directly on the NIA-led study by Makarious and colleagues, which trained multimodal machine-learning models on PPMI and externally validated in PDBP [24]. In that foundation study, combining modalities (e.g., olfaction and polygenic risk alongside molecular data) outperformed single-modality approaches and achieved strong external AUCs. The effort was supported by the Intramural Research Program of the National Institute on Aging and NINDS—underscoring NIH's strategic interest in predictive modeling for PD [24]. Predictors in this study were restricted to molecular features to quantify the stand-alone diagnostic and severity-tracking signal in peripheral molecules. This avoids clinical scales that can dominate models and dilute the biological signal.

To ensure defensible estimates and avoid information leakage, all preprocessing and feature selection are nested within participant-grouped cross-validation, with a prespecified internal hold-out and external testing only once model selection is complete. Early detection is a central motivation: accurate identification during the prodromal phase could enable earlier intervention and better trial enrichment [25]. PPMI is well-suited as a training resource for this aim, as the study explicitly recruits individuals with early, untreated (de novo) Parkinson's disease and healthy controls of similar age and sex; PD eligibility required a positive dopamine transporter (DAT) SPECT scan on central read, and healthy controls showed normal DAT imaging at baseline.

2 Methods

Matched whole-blood RNA sequencing and targeted plasma Olink proteomics data were obtained from the PPMI IR3 and 2023 PDBP releases, available through the AMP-PD Knowledge Platform [26]. These datasets were chosen because they represent two of the largest and most comprehensive longitudinal cohorts for Parkinson's disease, offering deeply integrated clinical and multi-omic data for well-characterized patients. After excluding samples that failed internal quality control, the integrated dataset comprised 971 visit-level samples from 413 unique participants, including 58,294 genes and 1,463 proteins. For cross-sectional modeling, participants were required to have at least one longitudinal instance with RNA-seq, proteomics, and clinical (MDS-UPDRS) data. This criterion yielded 235 participants and 745 visit-level samples for longitudinal analyses (median = 3 visits, IQR = 2–4). Among these, 111 participants contributed a baseline (visit-0) draw with matched MDS-UPDRS scores.

The Olink dataset spanned neurology, inflammation, cardiometabolic, and oncology panels, with protein features identified by UniProt IDs. Only participants with both RNA-seq and proteomics data were retained for integrative analyses. For RNA-seq, differential expression analysis was performed using independent t-tests with Benjamini–Hochberg correction to reduce the number of features. The PPMI cohort comprised three groups: (1) Parkinson's disease (PD) patients diagnosed within two years and not on medication, (2) prodromal individuals over 60 with PD risk factors, and (3) healthy controls without neurological disease or a family history of PD [27]. To ensure diagnostic clarity, participants from the SWEDD and Genetic Registry cohorts were excluded. For external validation, PDBP RNA-seq and proteomics data harmonized to PPMI protocols were incorporated [28].

RNA-seq and Olink proteomic features were concatenated as predictors of binary disease status (Parkinson's disease versus healthy control), with the class label assigned accordingly for each sample. Within PPMI, 20% of PD samples were held out (stratified by patient ID) for internal testing, and all PDBP samples were reserved for external validation. To correct for class imbalance, the Random Forest classifier was configured with $class_weight =' balanced'$. Random Forest was compared with candidate classifiers (XGBoost, Gradient Boosting, SVM, Logistic Regression, Extra Trees), and was selected based on the highest ROC-AUC. In each cross-validation fold (10-fold GroupKFold on participant ID), variance thresholding, SelectKBest (ANOVA F-statistic), and standard scaling were applied prior to classification. Random Forest hyperparameters were tuned using 50 Latin Hypercube–sampled parameter sets in RandomizedSearchCV, optimizing for ROC-AUC. Latin Hypercube sampling was chosen for its balance of efficiency and reproducibility, offering uniform coverage of the hyperparameter space while reducing the number of required parameter sets compared to grid or random search [29].

2.1 Analysis of Selected Features

Model explanations were generated with SHapley Additive exPlanations (SHAP; Python package `shap` v0.44.1), quantifying each feature's contribution to PD-versus-control predictions [30]. SHAP values were computed relative to control class

probability; negative values indicated features that increased PD probability. SHAP summary plots provided global and local interpretability, while waterfall plots visualized probability shifts for individual cases. Force plots highlighted correctly and incorrectly classified samples.

2.2 Gene Ontology Enrichment

Gene Ontology (GO) enrichment analysis was performed with the `g:Profiler` Python API, querying UniProt IDs against GO:BP, GO:MF, and GO:CC terms. Significant terms ($p < 0.05$) were identified, gene ratios were calculated as the proportion of input proteins mapping to each term, and exported ranked results as CSV files for downstream analysis.

2.3 PPI Network Construction

Protein–protein interaction (PPI) networks were constructed by querying STRING DB v11.5 (accessed January 2025) with the `STRINGdb` R package (v2.8.5). The search was restricted to *Homo sapiens*, required experimental or curated evidence, set the confidence threshold to 0.4, and excluded self-loops and disconnected nodes. Cytoscape 3.10.0 using the STRING App 2.0 was used to visualize the network.

2.4 Hub Identification

Network hubs were identified using the CytoNCA plugin in Cytoscape. Nodes with a degree z-score ≥ 2 were defined as hubs. Hub robustness was verified through sensitivity analyses using betweenness and closeness centrality.

2.5 Connectivity Map Query

Differentially regulated proteins ($|\log_2 \mathrm{FC}| \geq 0.5$, $\mathrm{FDR} < 0.05$) were mapped to coding genes and queried the CLUE platform (CMap build 02) via the `cmapR` Python package (v4.0.2). Compounds with connectivity scores (tau) $\leq \pm 90$ were retained as candidate perturbagens.

2.6 Perturbagen Target Retrieval

Validated human targets were retrieved for each compound from DrugBank 5.3 (downloaded January 2025) and ChEMBL 32 via the `bioservices` Python API. Targets were collapsed to unique gene symbols and mapped onto the STRING PPI network.

2.7 Target Overlay and Enrichment Testing

Drug-target genes were overlaid on the PPI network using the enhancedGraphics Cytoscape plugin, representing compound–target edges as dashed lines [31]. Enrichment was assessed with one-sided Fisher's exact tests, comparing observed target hits among PD-associated nodes to expectations across all ∼19,000 STRING proteins. Benjamini–Hochberg correction was applied and $q < 0.05$ was considered significant.

2.8 Pathway Co-localization Testing

Compounds with significant PPI enrichment were tested to determine whether their targets were over-represented in GO biological processes and Reactome pathways previously enriched in the PD biomarker gene set (e.g., "cytokine–cytokine receptor interaction"). Hypergeometric tests were used to evaluate whether the observed overlap between compound targets and pathway genes exceeded random expectation. FDR correction using the Benjamini-Hochberg procedure was applied to account for multiple pathway testing [32].

2.9 Composite Severity Index Derivation

A severity index was derived by modeling baseline protein expression against symptom burden. MDS-UPDRS Parts I–IV scores were extracted from AMP-PD clinical datasets and summed to obtain total scores [33], which were then matched to selected baseline features. A linear regression model was trained with five-fold `GroupKFold` cross-validation, using standardized feature coefficients as weights. The resulting severity index represented a weighted sum of baseline features [34, 35].

2.10 Longitudinal Progression Modeling

The severity index was evaluated as a predictor of progression using longitudinal samples. Baseline-derived PSI values were projected onto follow-up visits, and a linear mixed-effects model was fitted with total MDS-UPDRS as the dependent variable [36]. Fixed effects included severity index, time (years), and their interaction [37], with random intercepts and slopes specified for each participant. The model was implemented in `statsmodels` (`MixedLM`).

2.11 Statistical Evaluation

Spearman correlations were computed between each selected feature and total MDS-UPDRS, with multiple testing correction applied using the Benjamini–Hochberg method. Scatterplots with fitted lines were used to visualize severity index versus MDS-UPDRS, and residuals were assessed to evaluate model fit. Heteroskedasticity was tested using the Breusch–Pagan test.

All analyses were conducted in Python 3.11.4. Core packages included `pandas` (2.2.2), `numpy` (1.26.4), `scikit-learn` (1.4.2), and `statsmodels` (0.14.1). Figures were generated with `matplotlib` (3.8.3).

3 Results

3.1 Model Comparison and Classification Performance

Proteomic features outperformed other modalities for Parkinson's disease (PD) classification. Figure 1 summarizes the analytical workflow. Three data configurations were evaluated using the same pipeline: (1) proteomics-only, (2) gene expression–only, and (3) combined proteomics and gene expression. Each configuration was used to

train and tune models using a 10-fold `GroupKFold` cross-validation on an 80% training subset and assessed performance on an external PDBP hold-out.

The gene expression–only model performed worst, whereas the proteomics-only model delivered the strongest cross-validated and external performance (Table 1). A paired bootstrap comparison (1,000 resamples) confirmed that the proteomics-only model significantly exceeded the combined model (Wilcoxon signed-rank test, $p = 1.22 \times 10^{-103}$).

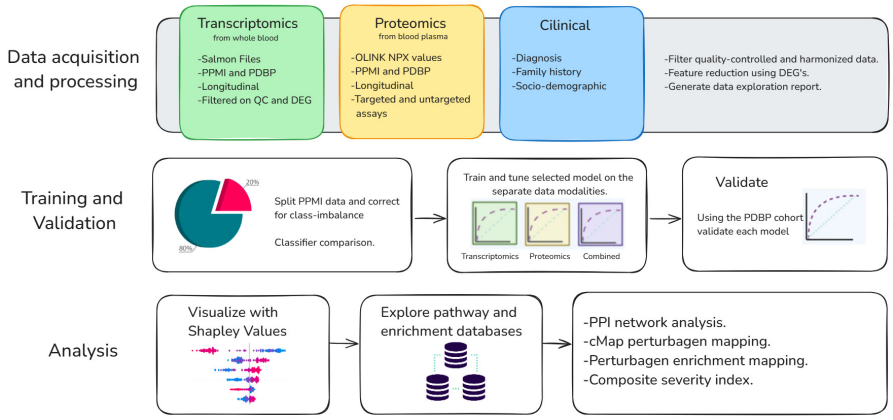

Fig. 1 Workflow diagram.

Given its accuracy, parsimony, and cross-cohort robustness, the proteomics-only model was selected for downstream analyses. On the external validation set, this model achieved an ROC–AUC of 0.8879 and an accuracy of 84%. The confusion matrix indicated 36 false negatives and 13 false positives, consistent with high sensitivity and specificity (Figure 2).

Models were optimized with Latin Hypercube Sampling over 50 `RandomizedSearchCV` iterations. The best proteomics-only Random Forest used $n_estimators = 1074$, $max_depth = 2$, $min_samples_split = 2$, $min_samples_leaf = 1$, and $class_weight =$ 'balanced'. Feature selection retained 32 protein markers that captured the most discriminative signal for PD.

3.2 Model Interpretability via SHAP

SHapley Additive exPlanations (SHAP) were used to assess the proteomics-only Random Forest (Figure 3). Because the classifier outputs control-class probability, negative SHAP values indicate features that increase PD probability. `P20711` (*DDC*) showed the largest magnitude contribution (mean SHAP $= -0.063 \pm 0.082$), implicating higher levels with increased PD likelihood. `O60760` (*HPGDS*) (0.0121 ± 0.0518), `Q16853` (*AOC3*) (0.0075 ± 0.0502), and `P01236` (*PRL*) (-0.0182 ± 0.0470) also ranked among the most influential features, with additional moderate effects from `Q7Z5L0` (*VMO1*) (0.0026 ± 0.0266) and `Q5T2D2` (*TREML2*) (0.0117 ± 0.0251). Features with the most

Table 1 Model performance across data configurations. ROC-AUC values are reported for cross-validation on the training set and held-out performance on the external validation set. Cross-validation used participant-wise GroupKFold with all preprocessing nested to prevent leakage. Higher AUC indicates better PD-vs-control discrimination; 95% CIs were computed by stratified bootstrap (1,000 resamples).

Model	Cross-Validation ROC-AUC	Validation ROC-AUC	95% CI Lower	95% CI Upper
Combined	0.9663	0.8825	0.8469	0.9177
Proteomics only	**0.9710**	**0.8879**	**0.8514**	**0.9236**
RNA-seq only	0.5248	0.5869[1]	0.5507	0.6235

Source: Results from 10-fold `GroupKFold` cross-validation and external validation on PDBP cohort.

[1]Proteomics-only model significantly outperforms the Combined model on the validation set (Wilcoxon signed-rank test, $p = 1.22 \times 10^{-103}$).

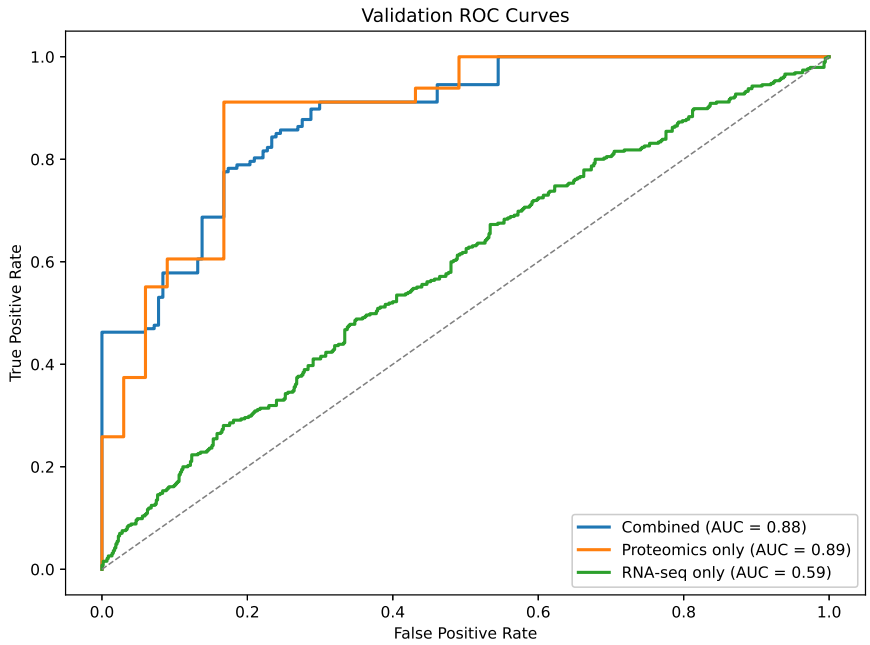

Fig. 2 ROC curves for proteomics-only, combined, gene-only models on hold-out set; proteomics-only highest, gene-only lowest AUC.

stable (low-variance) contributions included `Q8NCC3` (*PLA2G15*) (-0.0010 ± 0.0070), `P21860` (*ERBB3*) (0.0014 ± 0.0069), and `P35052` (*GPC1*) (-0.0005 ± 0.0052).

3.3 Gene Ontology Enrichment

Gene Ontology enrichment analysis was conducted on the 32 selected proteins to provide context for their biological functions. Significant over-representation was observed

for extracellular-related terms, including "extracellular space" ($p = 3.0 \times 10^{-7}$), "extracellular region" ($p = 1.4 \times 10^{-6}$), and "secreted proteins" ($p = 1.5 \times 10^{-8}$) (Figure 4). Additional enrichments included disulfide bonds ($p = 3.0 \times 10^{-9}$), glycosylation ($p = 2.5 \times 10^{-4}$), calcium ion binding ($p = 9.0 \times 10^{-4}$), and EGF-like domains ($p = 9.8 \times 10^{-4}$), pointing to signaling programs relevant to neurodegeneration.

Proteins with disulfide bonds ($p = 3.0 \times 10^{-9}$), glycosylation ($p = 2.5 \times 10^{-4}$), calcium ion binding ($p = 9.0 \times 10^{-4}$), and EGF-like domains ($p = 9.8 \times 10^{-4}$) were also enriched, suggesting involvement in signaling pathways relevant to neuronal function and neurodegeneration.

3.4 Protein–Protein Interaction Network

STRING analysis of the 32 proteins yielded a significantly enriched protein–protein interaction (PPI) network ($p = 0.0438$) with 32 nodes and 10 edges (average degree = 0.625). Functional annotation recapitulated extracellular themes—"extracellular region" ($p = 2.48 \times 10^{-6}$), "extracellular space" ($p = 5.40 \times 10^{-5}$)—and highlighted "cytokine–cytokine receptor interaction" ($p = 0.0020$). Central nodes were predominantly secreted, disulfide-bonded, and glycosylated proteins, suggesting roles in immune signaling and inflammation. Enrichment of proteoglycans ($p = 0.0151$) and heparan sulfate–related proteins ($p = 0.0430$) was also observed, implicating extracellular matrix pathways.

3.5 Drug Perturbation Mapping

Differential expression analysis confirmed that all 32 selected proteins were significantly altered in PD versus controls. Overexpressed proteins included `P20711` (*DDC*) (fold change = 2.17), `P35443` (*THBS4*) (2.08), and `P01130` (*LDLR*) (1.47), while underexpressed proteins included `P01236` (*PRL*) (−5.02), `Q96N03` (*VSTM2L*) (−6.12), and `Q9BYF1` (*ACE2*) (−2.00), consistent with dysregulated immune and metabolic processes.

This analysis step is hypothesis-generating; results depend on the CLUE build and filters used. Connectivity Map (CMap) analysis identified three compounds with strongly negative connectivity scores, suggesting potential reversal of PD-associated expression patterns (Appendix Table A). Each candidate's validated targets were projected onto a curated Parkinson's disease protein–protein interaction (PPI) network. Among these, only the top-ranked perturbagen, SAL-1 (alloxazine/isoalloxazine), was retained. Annotated in CMap as an adenosine receptor antagonist targeting ADORA2A/ADORA2B and profiled in L1000 (Broad Repurposing Hub ID BRD-K40213712). SAL-1 overlapped with PD-associated nodes. GO and Reactome analyses of SAL-1 targets revealed 45 enriched pathways at $q \leq 0.05$, including circadian rhythm regulation, purinergic signaling, and glutamate/GABA transmission, supporting SAL-1 as a mechanistically plausible candidate for follow-up.

3.6 Protein Associations with Clinical Severity

Plasma protein levels were correlated with total MDS-UPDRS scores. `P20711` (*DDC*) showed the strongest positive association ($\rho = 0.61$, $q = 5.9 \times 10^{-11}$), followed by

P07196 (*NEFL*) ($\rho = 0.44$, $q = 2.3 \times 10^{-5}$). Prolactin (P01236, *PRL*; $\rho = -0.43$) and O60760 (*HPGDS*; $\rho = -0.30$) were negatively correlated (Figure 5).

3.7 Derivation of Proteomic Severity Index (PSI)

A Proteomic Severity Index (PSI) was constructed as a weighted linear combination of the 32 proteins. Among 111 participants with baseline data, PSI explained 28.2% of the variance in total MDS-UPDRS (mean cross-validated $R^2 = 0.282$). The largest positive weights were assigned to P20711 (*DDC*) (+9.70) and P35443 (*THBS4*) (+7.66), whereas Q16853 (*AOC3*) (−5.73), P01229 (*LHB*) (−4.80), and P01236 (*PRL*) (−3.74) carried negative weights, indicating inverse associations with symptom burden.

3.8 Longitudinal prediction of disease progression

A total of 745 longitudinal observations from 235 participants were modeled using linear mixed-effects. Baseline PSI independently predicted total MDS-UPDRS (sum of Parts I–IV; $\beta = 0.277$, $p < 0.0001$), and time was also positively associated ($\beta = 1.80$, $p < 0.0001$). The PSI×time interaction was negative and significant ($\beta = -0.043$, $p = 0.027$), suggesting that individuals with higher baseline PSI experienced slightly slower symptom progression over follow-up.

4 Discussion

Our findings demonstrate that proteomic features can outperform transcriptomic and multimodal features in distinguishing Parkinson's disease (PD) from healthy controls. As the functional endpoints of gene expression, proteins may more accurately reflect ongoing biological processes. Because proteins are the effectors of cellular function, proteomic measurements more directly reflect disease activity and can reveal changes not captured by RNA-seq. Additionally, proteomic profiles exhibited greater cross-cohort consistency, underscoring their translational potential for PD detection and patient stratification.

Strong enrichment for extracellular and secreted proteins was observed, indicating that Parkinson's disease–related changes extend beyond intracellular signaling into circulating compartments accessible in plasma. Many of these proteins mediate intercellular communication, immune modulation, and extracellular matrix (ECM) remodeling, processes increasingly implicated in neurodegeneration [38]. Enrichment for post-translational features such as disulfide bonds and glycosylation supports specialized extracellular functions, including stabilization, trafficking, and receptor engagement [39]. The protein–protein interaction network contained more edges than expected by chance, which is consistent with coordinated activity among circulating factors. Together, these observations support a model in which PD involves measurable disruption of the plasma proteome, particularly among extracellular signaling and structural proteins that could serve as biomarkers or therapeutic entry points.

The prominence of EGF-like domains suggests additional ligand–receptor pathways that remain underexplored in PD [40]. Concurrent enrichment for cytokine–cytokine receptor interactions reinforces the role of neuroinflammation, including microglial

activation and cytokine release, in dopaminergic neuron loss [41]. Calcium-binding terms and synaptic pathway annotations further point to perturbed neurotransmission, consistent with reports of altered synaptic proteins in brain-derived extracellular vesicles from PD patients [42]. Reactome analysis converged on `GPC1`, `DCN`, `THBS4`, `MATN3`, `ERBB3`, and `PDGFC`, implicating ECM proteoglycan remodeling, glycosaminoglycan metabolism, and receptor tyrosine kinase signaling. Although typically linked to tissue remodeling and fibrosis, these pathways may impair neurotrophic signaling and amplify inflammatory tone in PD [43], and they merit targeted investigation.

Several of these blood-based proteins have been repeatedly linked to PD risk or progression in prior literature, including neurofilament light chain (NfL), glial fibrillary acidic protein (GFAP), total/oligomeric α-synuclein, DJ-1 (PARK7), and proinflammatory cytokines (e.g., IL-6, TNF-α).[44–46]. Large proteomic cohorts have also highlighted plasma proteins associated with incident or prodromal PD [47]. Within our panel, aromatic L-amino acid decarboxylase (AADC/DDC), a dopamine-synthesis enzyme, has recently emerged as a fluid biomarker measurable in plasma and linked to Lewy body disorders and parkinsonian syndromes.[48, 49]. By contrast, circulating prolactin (PRL) shows mixed evidence in PD and is sensitive to dopaminergic therapy effects,[50, 51] while hematopoietic prostaglandin D synthase (HPGDS) has mechanistic support in neuroinflammation but limited validation as a PD blood biomarker.[52]. Our findings both recapitulate emerging plasma signals (e.g., DDC) and nominate less-explored extracellular proteins that align with immune/ECM pathways for further validation.

Potential therapeutics

Our *in silico* perturbation analysis identified small molecules with strongly negative connectivity to the PD proteomic signature. SAL-1 (alloxazine/isoalloxazine) emerged as a notable candidate, consistent with reported adenosine receptor antagonism at A_{2B} and potential A_{2A} activity [53]. While dopaminergic therapies remain standard of care, they do not slow disease progression and can cause dyskinesia [54]. Modulating adenosine signaling, particularly A_{2A} pathways, may improve motor symptoms and reduce dyskinesia, though SAL-1 has not been studied in PD. These results are hypothesis-generating and require experimental validation.

Fostamatinib, a SYK inhibitor, and homoharringtonine (HHT), a protein synthesis inhibitor, also showed negative connectivity. Prior integrative analyses have implicated fostamatinib through interactions with targets such as `DCLK1`, `KCNJ6`, and `SNCA` [55]. HHT reduced amyloid burden and neuroinflammation in Alzheimer's disease models via STAT3 signaling [56], which suggests relevance to shared inflammatory mechanisms in PD. Conversely, compounds with positive connectivity, including the proteasome inhibitor MG-132, are known to exacerbate proteostatic stress and promote α-synuclein aggregation [57]. PKC activators such as prostratin and phorbol-12-myristate-13-acetate may similarly amplify maladaptive kinase cascades [55, 58]. These contrasts highlight therapeutic strategies that restore protein homeostasis, dampen inflammation, and normalize dysregulated receptor signaling.

Proteomic Severity Index

Several proteins showed directionally consistent associations with total MDS-UPDRS. `P20711` (*DDC*), an enzyme in dopamine biosynthesis, was the top positive correlate and the strongest contributor to our severity index, aligning with biological expectations. Prolactin (`P01236`, *PRL*) and hematopoietic prostaglandin D synthase (`O60760`, *HPGDS*) correlated negatively with symptom burden, which may indicate protective or compensatory responses early in disease.

The Proteomic Severity Index (PSI) discriminated moderate-to-severe cases with an AUROC of 0.964 and explained 28.2% of the variance in total MDS-UPDRS, which is notable in a multifactorial disorder [59, 60]. In longitudinal models, PSI and time each showed independent effects on severity. The negative PSI-by-time interaction suggests that individuals with high baseline PSI exhibited slightly slower subsequent worsening. This pattern could reflect regression toward the mean, therapeutic exposure, ceiling effects, or underlying biology. Prospective validation is needed to distinguish among these explanations.

Overall, PSI emerges as a reproducible, blood-based index of motor symptom burden with potential utility for patient stratification and progression monitoring, including trial enrichment and endpoint refinement.

Limitations

This study uses multi-site PPMI and PDBP datasets, which introduce variability from demographics, sampling protocols, and batch effects [61]. Such heterogeneity can dilute effect sizes and complicate generalization across settings [62]. Peripheral blood biomarkers provide practical accessibility but may incompletely capture central nervous system processes central to PD pathogenesis. The proteomics sample size and feature count constrain model complexity. Although rigorous cross-validation, external testing, and conservative feature selection were employed, data-driven selection may still introduce overfitting or selection bias. Finally, the primary analysis is cross-sectional, which limits causal inference about temporal dynamics.

Biological validation

Computational robustness does not replace experimental proof. Independent cohorts, targeted protein assays, and mechanistic work in cellular and animal models are necessary to establish causality. Candidate therapeutics from the connectivity analysis require preclinical testing for target engagement, efficacy, and safety before any clinical consideration.

Future work

Future studies should validate and extend the drug-repurposing signals and pathway-level hypotheses reported here. SAL-1 and fostamatinib merit particular attention, given potential links to receptor tyrosine kinases, ECM remodeling, and immune signaling. The observed extracellular signatures, including proteoglycan and glycosaminoglycan pathways, support the rationale for multi-tissue profiling in cerebrospinal fluid,

brain tissue, and induced pluripotent stem cell–derived neuronal models. Integrating spatial transcriptomics and phosphoproteomics could clarify how extracellular changes intersect with neuronal signaling at cellular resolution.

Clinically, PSI is a scalable, minimally invasive tool that warrants evaluation for predicting conversion from prodromal to clinical PD and for responsiveness to pharmacologic interventions. Embedding PSI into adaptive trial designs could enable biomarker-guided enrichment and more sensitive endpoints [63]. Defining molecular subtypes may further align therapeutic mechanisms with specific patient groups [64]. Extending this framework to other neurodegenerative diseases may reveal shared or distinct biomarker programs that inform cross-disease repurposing and systems-level understanding.

5 Conclusion

Proteomics-based models distinguished PD from healthy controls, revealed extracellular and immune-linked biology, and yielded a blood-based severity index with prognostic value. These findings highlight the translational promise of plasma proteomics for PD. Focused experimental validation and collaborative clinical studies are the next steps toward therapeutic development and biomarker deployment.

6 Availability of data and materials

Data statement here. Code statement here.

Abbreviations

- AMP-PD: Accelerating Medicines Partnership Parkinson's Disease
- AUC: Area Under the Curve
- CLUE: Connectivity Map User Environment
- CMap: Connectivity Map
- CNS: Central Nervous System
- ECM: Extracellular Matrix
- EV: Extracellular Vesicle
- FDR: False Discovery Rate
- GO: Gene Ontology
- LHS: Latin Hypercube Sampling
- MAO-B: Monoamine Oxidase B
- MDS-UPDRS: Movement Disorder Society Unified Parkinson's Disease Rating Scale
- MRI: Magnetic Resonance Imaging
- NGS: Next-Generation Sequencing
- NPX: Normalized Protein Expression
- PD: Parkinson's Disease
- PDBP: Parkinson's Disease Biomarkers Program
- PEA: Proximity Extension Assay
- PPI: Protein–Protein Interaction
- PSI: Proteomic Severity Index

- q: Adjusted p-value (FDR-corrected)
- RNA-seq: RNA Sequencing
- ROC: Receiver Operating Characteristic
- SHAP: SHapley Additive exPlanations
- STRING: Search Tool for the Retrieval of Interacting Genes/Proteins
- SPECT: Single Photon Emission Computed Tomography
- TPM: Transcripts Per Million

Declarations

Ethics approval and consent to participate
A secondary analysis was conducted using de-identified, controlled-access individual-level data from the AMP-PD Knowledge Platform (PPMI IR3 and PDBP 2023 releases). Access required a Data Use Agreement (DUA) authorized by George Mason University. AMP-PD does not provide direct identifiers or any re-identification key to investigators, and the DUA prohibits any attempt to re-identify participants. All AMP-PD data-handling requirements and George Mason University research data policies were followed.

The IRB at George Mason University determined that IRB ID: STUDY00000792 is not research involving human subjects as defined by DHHS and FDA regulations.

The source cohorts (PPMI and PDBP) obtained ethics approval at participating institutions and collected written informed consent from all participants before data collection and de-identification. No contact or interaction with participants occurred, and no interventions were performed. **Consent for publication**

Not applicable. This study includes no individual person data (images or identifiable details).

Availability of data and materials
Data used in this study were obtained from the Accelerating Medicines Partnership Parkinson's Disease (AMP-PD) Knowledge Platform (https://amp-pd.org). Access to AMP-PD requires registration and approval by the respective data access committees.

All code used for preprocessing, modeling, and statistical analysis is openly available at [https://github.com/NMinster/pd-multiomics-analysis]. Processed data files sufficient to reproduce the figures and tables in this manuscript are provided in the same repository, subject to the AMP-PD and PDBP data use agreements.

Competing interests
The authors declare that they have no competing interests.

Funding
This work was supported by the National Science Foundation Research Traineeship (NSF-NRT) program and the George Mason University Dissertation Completion Grant. The authors and their institution did not receive payment or services from a third party for any other aspect of the submitted work.

Authors' contributions
Author 1 conceptualized the study, performed data analysis, interpreted results, and

wrote the manuscript. Author 2 supervised the project, contributed to the study design and interpretation, and revised the manuscript for critical content. Both authors read and approved the final manuscript.

Acknowledgements

Data used in the preparation of this article were obtained from the Accelerating Medicines Partnership® (AMP®) Parkinson's Disease (AMP® PD) Knowledge Platform. For up-to-date information on the study, visit https://www.amp-pd.org.

The AMP® PD program is a public–private partnership managed by the Foundation for the National Institutes of Health and funded by the National Institute of Neurological Disorders and Stroke (NINDS) in partnership with the Aligning Science Across Parkinson's (ASAP) initiative; Celgene Corporation, a subsidiary of Bristol-Myers Squibb Company; GlaxoSmithKline plc (GSK); The Michael J. Fox Foundation for Parkinson's Research; Pfizer Inc.; AbbVie Inc.; Sanofi US Services Inc.; and Verily Life Sciences.

ACCELERATING MEDICINES PARTNERSHIP and AMP are registered service marks of the U.S. Department of Health and Human Services.

PPMI is sponsored by The Michael J. Fox Foundation for Parkinson's Research and supported by a consortium of scientific partners. For up-to-date information on the study, visit https://www.ppmi-info.org.

The Parkinson's Disease Biomarker Program (PDBP) consortium is supported by the National Institute of Neurological Disorders and Stroke (NINDS) at the National Institutes of Health. A full list of PDBP investigators can be found at https://pdbp.ninds.nih.gov/policy. The PDBP investigators have not participated in reviewing the data analysis or content of the manuscript.

Authors' information

Not applicable.

Appendix A cMap Results

References

[1] Dorsey ER, Sherer T, Okun MS, Bloem BR. The emerging evidence of the Parkinson pandemic. Journal of Parkinson's disease. 2018;8(s1):S3–S8.

[2] Moustafa AA, Chakravarthy S, Phillips JR, Gupta A, Keri S, Polner B, et al. Motor symptoms in Parkinson's disease: A unified framework. Neuroscience & Biobehavioral Reviews. 2016;68:727–740.

[3] Reich SG, Savitt JM. Parkinson's disease. Medical Clinics of North America. 2019;103(2):337–+.

[4] Nalls MA, McLean CY, Rick J, Eberly S, Hutten SJ, Gwinn K, et al. Diagnosis of Parkinson's disease on the basis of clinical and genetic classification: a population-based modelling study. The Lancet Neurology. 2015;14(10):1002–1009. https://doi.org/10.1016/S1474-4422(15)00178-7.

Table A1 High-scoring compounds from screening (sideways layout)

Rank	Score	Type	ID	Name	Description
14	98.20	cp	BRD-K70577657	H-9	PKA inhibitor
19	97.89	cp	BRD-K72895815	SSR-69071	Leukocyte elastase inhibitor
24	97.74	cp	BRD-K74402642	NSC-632839	Ubiquitin specific protease inhibitor
30	97.60	cp	BRD-K04546108	JAK3-inhibitor-VI	JAK inhibitor
38	97.01	cp	BRD-K84987553	MDM2-inhibitor	MDM inhibitor
46	96.72	cp	BRD-K81225797	SCH-58261	Adenosine receptor antagonist
47	96.62	cp	BRD-K96799727	pifithrin-mu	HSP inhibitor
49	96.53	cp	BRD-K49669041	BX-912	Pyruvate dehydrogenase kinase inhibitor
51	96.39	cp	BRD-A55484088	BNTX	Opioid receptor antagonist
53	96.17	cp	BRD-K63630713	etacrynic-acid	$Na^+/K^+/Cl^-$ transporter inhibitor
54	96.15	cp	BRD-A03216249	mepivacaine	Potassium channel blocker
55	96.07	cp	BRD-K61829047	7b-cis	Exportin antagonist
59	95.69	cp	BRD-A43150328	penicillic-acid	Other antibiotic
61	95.67	cp	BRD-A45333398	periplocymarin	Apoptosis stimulant
64	95.60	cp	BRD-K78659596	MLN-2238	Proteasome inhibitor
65	95.56	cp	BRD-K67844266	MLN-4924	NEDD activating enzyme inhibitor
66	95.51	cp	BRD-K63606607	bufalin	ATPase inhibitor
67	95.40	cp	BRD-K98548675	parthenolide	$NF-\kappa B$ pathway inhibitor
69	95.28	cp	BRD-K13169950	NSC-3852	HDAC inhibitor
71	95.24	cp	BRD-K56334280	amonafide	Topoisomerase inhibitor
77	95.06	cp	BRD-K51967704	BIIB021	HSP inhibitor
81	94.42	cp	BRD-K18518344	digitoxigenin	ATPase inhibitor
82	94.40	cp	BRD-K60230970	MG-132	Proteasome inhibitor
84	94.29	cp	BRD-A11007541	BCI-hydrochloride	Protein phosphatase inhibitor
85	94.29	cp	BRD-K80970344	pyrrolidine-dithiocarbamate	$NF-\kappa B$ pathway inhibitor
91	93.73	cp	BRD-A80502530	cinobufagin	ATPase inhibitor
92	93.70	cp	BRD-K22010301	JLK-6	γ-Secretase inhibitor
93	93.67	cp	BRD-A34806832	proscillaridin	ATPase inhibitor
96	93.56	cp	BRD-A71459254	cymarin	ATPase inhibitor
102	92.99	cp	BRD-A58955223	sulforaphane	Antineoplastic
103	92.96	cp	BRD-K99749624	linifanib	PDGFR receptor inhibitor
104	92.92	cp	BRD-K33583600	isoliquiritigenin	Guanylate cyclase activator
107	92.78	cp	BRD-K26669427	WR-216174	PFMRK inhibitor
108	92.70	cp	BRD-A56020723	CA-074-Me	Cathepsin inhibitor
109	92.65	cp	BRD-K36737713	AG-957	Protein tyrosine kinase inhibitor
112	92.35	cp	BRD-K15935639	z-leu3-VS	Proteasome inhibitor
116	91.93	cp	BRD-K17497770	butein	EGFR inhibitor
117	91.88	cp	BRD-K05396879	15-delta-prostaglandin-j2	PPAR receptor agonist
120	91.66	cp	BRD-A15914070	4-hydroxy-2-nonenal	Cytotoxic lipid peroxidation product
123	91.47	cp	BRD-K13566078	BMS-345541	IKK inhibitor
124	91.45	cp	BRD-K84595254	strophanthidin	ATPase inhibitor
127	91.41	cp	BRD-K78373679	RO-3306	CDK inhibitor
128	91.31	cp	BRD-K07762753	aminopurvalanol-a	Tyrosine kinase inhibitor
132	91.12	cp	BRD-K91145395	prostratin	PKC activator
133	91.07	cp	BRD-A07395371	esmolol	Adrenergic receptor antagonist
134	90.59	cp	BRD-K66792149	quinoclamine	Algicide
136	90.52	cp	BRD-U51951544	ZG-10	JNK inhibitor
137	90.43	cp	BRD-A15079084	phorbol-12-myristate-13-acetate	PKC activator
138	90.38	cp	BRD-K37691127	hinokitiol	Tyrosinase inhibitor
141	90.27	cp	BRD-A41451487	PK-11195	Benzodiazepine receptor antagonist
143	90.05	cp	BRD-A94756469	digoxin	ATPase inhibitor
8522	–91.00	cp	BRD-K20285085	fostamatinib	SYK inhibitor
8527	–91.91	cp	BRD-K76674262	homoharringtonine	Protein synthesis inhibitor
8556	–98.73	cp	BRD-K40213712	SAL-1	Adenosine receptor antagonist

[5] Rizzo G, Copetti M, Arcuti S, Martino D, Fontana A, Logroscino G. Accuracy of clinical diagnosis of Parkinson disease: A systematic review and meta-analysis. Neurology. 2016;86(6):566–576. Publisher: Wolters Kluwer Health, Inc. on behalf of the American Academy of Neurology Section: Views & Reviews. https://doi.org/10.1212/WNL.0000000000002350.

[6] Durcan R, Wiblin L, Lawson RA, Khoo TK, Yarnall AJ, Duncan GW, et al. Prevalence and duration of non-motor symptoms in prodromal Parkinson's disease. Eur J Neurol. 2019;26(7):979–985. https://doi.org/10.1111/ene.13919.

[7] Rispoli V, Schreglmann SR, Bhatia KP. Neuroimaging advances in Parkinson's disease. Curr Opin Neurol. 2018;31(4):415–424. https://doi.org/10.1097/WCO.0000000000000584.

[8] Concha-Marambio L, Pritzkow S, Shahnawaz M, Farris CM, Soto C. Seed amplification assay for the detection of pathologic alpha-synuclein aggregates in cerebrospinal fluid. Nat Protoc. 2023;18(4):1179–1196. Number: 4 Publisher: Nature Publishing Group. https://doi.org/10.1038/s41596-022-00787-3.

[9] Cocco C, Manai AL, Manca E, Noli B. Brain-Biomarker Changes in Body Fluids of Patients with Parkinson's Disease. Int J Mol Sci. 2023;24(13):10932. https://doi.org/10.3390/ijms241310932.

[10] Gabbert C, Schaake S, Lüth T, Much C, Klein C, Aasly JO, et al. GBA1 in Parkinson's disease: variant detection and pathogenicity scoring matters. BMC Genomics. 2023;24(1):322. https://doi.org/10.1186/s12864-023-09417-y.

[11] Chen Y, Sam R, Sharma P, Chen L, Do J, Sidransky E. Glucocerebrosidase as a therapeutic target for Parkinson's disease. Expert Opin Ther Targets. 2020;24(4):287. Publisher: NIH Public Access. https://doi.org/10.1080/14728222.2020.1733970.

[12] Kedariti M, Frattini E, Baden P, Cogo S, Civiero L, Ziviani E, et al. LRRK2 kinase activity regulates GCase level and enzymatic activity differently depending on cell type in Parkinson's disease. NPJ Parkinsons Dis. 2022;8(1):1–14. Number: 1 Publisher: Nature Publishing Group. https://doi.org/10.1038/s41531-022-00354-3.

[13] Smith LJ, Lee CY, Menozzi E, Schapira AHV. Genetic variations in GBA1 and LRRK2 genes: Biochemical and clinical consequences in Parkinson disease. Front Neurol. 2022;13. Publisher: Frontiers. https://doi.org/10.3389/fneur.2022.971252.

[14] Reed X, Schumacher-Schuh A, Hu J, Bandres-Ciga S. Advancing Personalized Medicine in Common Forms of Parkinson's Disease through Genetics: Current Therapeutics and the Future of Individualized Management. J Pers Med. 2021;11(3):169. https://doi.org/10.3390/jpm11030169.

[15] Verschuur CVM, Suwijn SR, Boel JA, Post B, Bloem BR, van Hilten JJ, et al. Randomized Delayed-Start Trial of Levodopa in Parkinson's Disease. New England Journal of Medicine. 2019;380(4):315–324. Publisher: Massachusetts Medical Society _eprint: https://doi.org/10.1056/NEJMoa1809983. https://doi.org/10.1056/NEJMoa1809983.

[16] Bogetofte H, Alamyar A, Blaabjerg M, Meyer M. Levodopa Therapy for Parkinson's Disease: History, Current Status and Perspectives. CNS & Neurological Disorders - Drug Targets- CNS & Neurological Disorders). 2020;19(8):572–583. https://doi.org/10.2174/1871527319666200722153156.

[17] Rosqvist K, Horne M, Hagell P, Iwarsson S, Nilsson MH, Odin P. Levodopa Effect and Motor Function in Late Stage Parkinson's Disease. Journal of Parkinson's Disease. 2018;8(1):59–70. https://doi.org/10.3233/JPD-171181.

[18] Lang AE, Espay AJ. Disease Modification in Parkinson's Disease: Current Approaches, Challenges, and Future Considerations. Movement Disorders: Official Journal of the Movement Disorder Society. 2018;33(5):660–677. https://doi.org/10.1002/mds.27360.

[19] Xu M, Zhu J, Liu XD, Luo MY, Xu NJ. Roles of physical exercise in neurodegeneration: reversal of epigenetic clock. Translational Neurodegeneration. 2021;10(1):30.

[20] Issa NT, Byers SW, Dakshanamurthy S. Big data: the next frontier for innovation in therapeutics and healthcare. Expert Review of Clinical Pharmacology. 2014;7(3):293–298. Publisher: Taylor & Francis _eprint: https://doi.org/10.1586/17512433.2014.905201. https://doi.org/10.1586/17512433.2014.905201.

[21] Li C, Gao Z, Su B, Xu G, Lin X. Data analysis methods for defining biomarkers from omics data. Analytical and Bioanalytical Chemistry. 2022;414(1):235–250. https://doi.org/10.1007/s00216-021-03813-7.

[22] Amelia A, Pena-Castillo L, Usefi H. Assessing the Reproducibility of Machine-learning-based Biomarker Discovery in Parkinson's Disease. Comput Biol Med. 2023;arXiv:2304.03239. https://doi.org/10.48550/arXiv.2304.03239. 2304.03239 [q-bio].

[23] Li T, Le W. Biomarkers for Parkinson's Disease: How Good Are They? Neuroscience Bulletin. 2020;36(2):183–194. https://doi.org/10.1007/s12264-019-00433-1.

[24] Makarious MB, Leonard HL, Vitale D, Iwaki H, Sargent L, Dadu A, et al. Multi-modality machine learning predicting Parkinson's disease. npj Parkinson's Disease. 2022;8(35). https://doi.org/10.1038/s41531-022-00288-w.

[25] Yamashita KY, others. Biomarkers in Parkinson's disease: A state of the art review. eNeurologicalSci. 2023;31:100443. https://doi.org/10.1016/j.ensci.2023.100443.

[26] Iwaki H, Leonard HL, Makarious MB, Bookman M, Landin B, Vismer D, et al. Accelerating Medicines Partnership: Parkinson's Disease. Genetic Resource. Movement Disorders. 2021;36(8):1795–1804. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/mds.28549. https://doi.org/10.1002/mds.28549.

[27] Marek K, Jennings D, Lasch S, Siderowf A, Tanner C, Simuni T, et al. The Parkinson progression marker initiative (PPMI). Progress in neurobiology. 2011;95(4):629–635.

[28] Rosenthal LS, Drake D, Alcalay RN, Babcock D, Bowman FD, Chen-Plotkin A, et al. The NINDS Parkinson's disease biomarkers program. Movement Disorders. 2016;31(6):915–923.

[29] Cauwet ML, Couprie C, Dehos J, Luc P, Rapin J, Riviere M, et al. Fully parallel hyperparameter search: Reshaped space-filling. In: International Conference on Machine Learning. PMLR; 2020. p. 1338–1348.

[30] Lundberg SM, Erion G, et al. From local explanations to global understanding with explainable AI for trees. Nature Machine Intelligence. 2020;2:56–67.

[31] Majeed A, Mukhtar S. Protein–protein interaction network exploration using cytoscape. In: Protein-Protein Interactions: Methods and Protocols. Springer; 2023. p. 419–427.

[32] Reimand J, Isserlin R, Voisin V, Kucera M, Tannus-Lopes C, Rostamianfar A, et al. Pathway enrichment analysis and visualization of omics data using g: Profiler, GSEA, Cytoscape and EnrichmentMap. Nature protocols. 2019;14(2):482–517.

[33] Goetz CG, Tilley BC, Shaftman SR, Stebbins GT, Fahn S, Martinez-Martin P, et al. Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): scale presentation and clinimetric testing results. Movement disorders: official journal of the Movement Disorder Society. 2008;23(15):2129–2170.

[34] Nguyen KP, Raval V, Treacher A, Mellema C, Yu F, Pinho MC, et al. Predicting longitudinal disease severity for individuals with parkinson's disease using functional MRI and machine learning prognostic models. bioRxiv. 2020;p. 2020–04.

[35] Chaudhry F, Kim TW, Elemento O, Betel D. Machine learning analysis of population-wide plasma proteins identifies hormonal biomarkers of Parkinson's

Disease. medRxiv. 2024;p. 2024–12.

[36] Bartl M, Dakna M, Schade S, Wicke T, Lang E, Ebentheuer J, et al. Longitudinal change and progression indicators using the Movement Disorder Society-Unified Parkinson's Disease Rating Scale in two independent cohorts with early Parkinson's Disease. Journal of Parkinson's Disease. 2022;12(1):437–452.

[37] Wang J, Luo S, Li L. Dynamic prediction for multiple repeated measures and event time data: An application to Parkinson's disease. The annals of applied statistics. 2017;11(3):1787.

[38] Shen J, Chen XC, Li WJ, Han Q, Chen C, Lu JM, et al. Identification of Parkinson's disease-related pathways and potential risk factors. Journal of International Medical Research. 2020;48(10):0300060520957197. Publisher: SAGE Publications Ltd. https://doi.org/10.1177/0300060520957197.

[39] Bakshi T, Pham D, Kaur R, Sun B. Hidden relationships between N-glycosylation and disulfide bonds in individual proteins. International Journal of Molecular Sciences. 2022;23(7):3742.

[40] Ou Gy, Lin Ww, Zhao Wj. Neuregulins in neurodegenerative diseases. Frontiers in aging neuroscience. 2021;13:662474.

[41] Bandres-Ciga S, Saez-Atienzar S, Kim JJ, Makarious MB, Faghri F, Diez-Fairen M, et al. Large-scale pathway specific polygenic risk and transcriptomic community network analysis identifies novel functional pathways in Parkinson disease. Acta Neuropathologica. 2020;140(3):341–358. https://doi.org/10.1007/s00401-020-02181-3.

[42] Hong CT, Chung CC, Yu RC, Chan L. Plasma extracellular vesicle synaptic proteins as biomarkers of clinical progression in patients with Parkinson's disease. Elife. 2024;12:RP87501.

[43] Rike WA, Stern S. Proteins and transcriptional dysregulation of the brain extracellular matrix in Parkinson's disease: a systematic review. International Journal of Molecular Sciences. 2023;24(8):7435.

[44] Tönges Lea. Blood-based biomarker in Parkinson's disease. Frontiers in Aging Neuroscience. 2022;.

[45] Yamashita Kea. Biomarkers in Parkinson's disease: A state of the art review. Clinics and Research in Hepatology and Gastroenterology (Parkinson's Disease and Movement Disorders Reports). 2023;.

[46] Zarkali Aea. Neuroimaging and fluid biomarkers in Parkinson's disease. Nature Communications. 2024;.

[47] Hällqvist Jea. Plasma proteomics identify biomarkers predicting Parkinson's disease up to 7 years before symptom onset. Brain. 2024;.

[48] Pereira Jea. DOPA decarboxylase is an emerging biomarker for Lewy body disorders. Nature Aging. 2023;.

[49] Rutledge Jea. Comprehensive proteomics of CSF, plasma, and urine nominate biomarkers for early Parkinson's disease. Science Translational Medicine. 2024;.

[50] Chang Kea. Assessing plasma levels of α-synuclein and neurofilament light chain in Parkinson's disease. Frontiers in Aging Neuroscience. 2021;.

[51] Teodoro Tea. Association between serum prolactin levels and neurodegenerative diseases: a meta-analysis. Neuroimmunomodulation. 2022;.

[52] Rittchen S, Heinemann A. Therapeutic potential of hematopoietic prostaglandin D2 synthase in allergic inflammation. Cells. 2019;8(6):619.

[53] Cieślak M, Komoszyński M, Wojtczak A. Adenosine A2A receptors in Parkinson's disease treatment. Purinergic Signalling. 2008;4(4):305–312. https://doi.org/10.1007/s11302-008-9100-8.

[54] Behari M, Srivastava AK, Pandey RM. Quality of life in patients with Parkinson's disease. Parkinsonism & Related Disorders. 2005;11(4):221–226. https://doi.org/10.1016/j.parkreldis.2004.12.005.

[55] Yergolkar AV, Yalamanchili J, Satish K, GNS H, Saraswathy G. PND24 target identification and drug repurposing for Parkinson's disease: a NOVEL integrative computational approach. Value in Health Regional Issues. 2020;22:S79.

[56] Jiang X, Wu Q, Zhang C, Wang M. Homoharringtonine inhibits Alzheimer's disease progression by reducing neuroinflammation via STAT3 signaling in APP/PS1 mice. Neurodegenerative Diseases. 2021;21(3-4):93–102.

[57] Sun F, Anantharam V, Zhang D, Latchoumycandane C, Kanthasamy A, Kanthasamy AG. Proteasome inhibitor MG-132 induces dopaminergic degeneration in cell culture and animal models. Neurotoxicology. 2006;27(5):807–815.

[58] Abbas Miana G, Riaz M, Shahzad-ul Hussan S, Zafar Paracha R, Zafar Paracha U. Prostratin: An Overview. Mini Reviews in Medicinal Chemistry. 2015;15(13):1122–1130.

[59] Kim SG, Hwang JS, George NP, Jang YE, Kwon M, Lee SS, et al. Integrative Metabolome and Proteome Analysis of Cerebrospinal Fluid in Parkinson's Disease. International Journal of Molecular Sciences. 2024;25(21):11406.

[60] Winchester L, Lawton M, Barber I, Ash J, Liu B, Evetts S, et al. Identification of a novel proteomic Biomarker in Parkinson's Disease: Discovery and Replication

in Blood, brain and CSF. medRxiv. 2021;p. 2021–12.

[61] Kaiser S, Zhang L, Mollenhauer B, Jacob J, Longerich S, Del-Aguila J, et al. A proteogenomic view of Parkinson's disease causality and heterogeneity. npj Parkinson's Disease. 2023;9(1):24.

[62] Wakasugi N, Takano H, Abe M, Sawamoto N, Murai T, Mizuno T, et al. Harmonizing multisite data with the ComBat method for enhanced Parkinson's disease diagnosis via DAT-SPECT. Frontiers in Neurology. 2024;15:1306546.

[63] Evans W, Meslin EM, Kai J, Qureshi N. Precision Medicine—Are We There Yet? A Narrative Review of Precision Medicine's Applicability in Primary Care. Journal of Personalized Medicine. 2024;14(4):418. Number: 4 Publisher: Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/jpm14040418.

[64] Frasier M, Fiske BK, Sherer TB. Precision medicine for Parkinson's disease: The subtyping challenge. Frontiers in Aging Neuroscience. 2022;14:1064057. https://doi.org/10.3389/fnagi.2022.1064057.

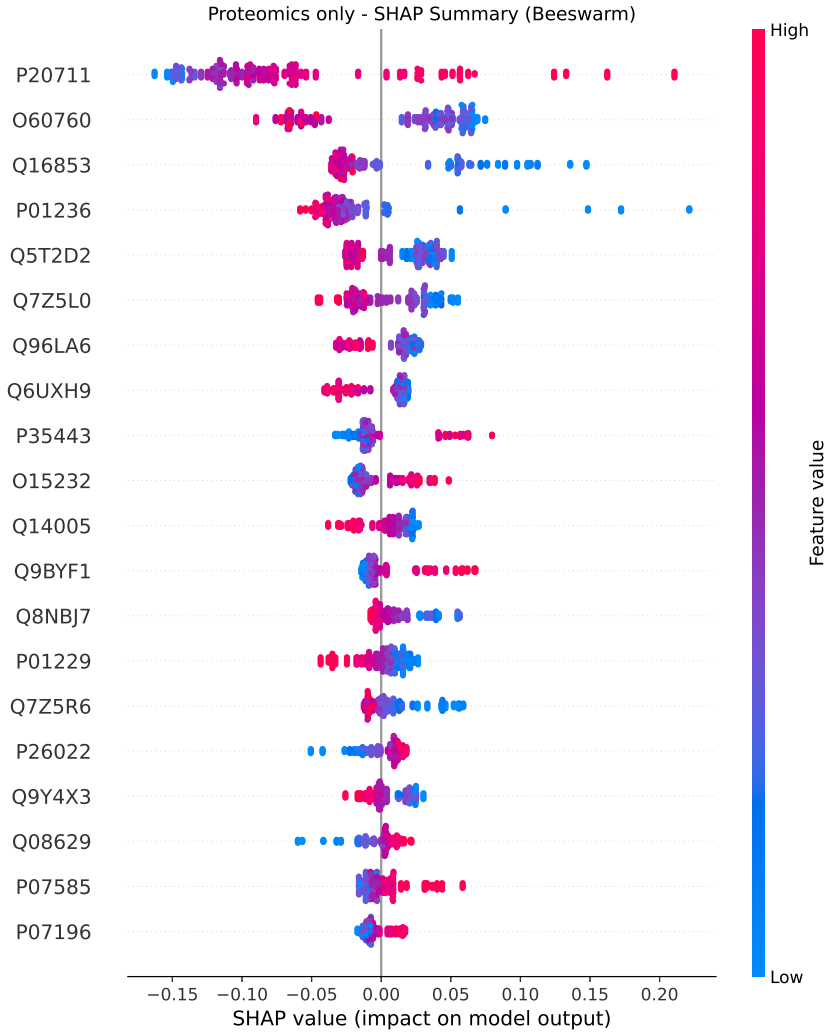

Fig. 3 Global feature importance for the proteomics-only classifier. SHAP summary ("beeswarm") plot for the Random Forest model trained on plasma proteins. Each point is a sample-level SHAP value for a given protein (x-axis: impact on model log-odds). Features are ordered by mean absolute SHAP across participants (top = most influential). Point color encodes the standardized NPX value of the protein in that sample (higher = warmer). Because the classifier outputs control-class probability, negative SHAP values indicate proteins that increase the probability of PD in that sample. Notable contributors include DDC (P20711) and THBS4 (P35443), which have positive associations with PD, and PRL (P01236) and HPGDS (O60760), which have negative associations. Preprocessing (variance filtering, SelectKBest, scaling) and model fitting were nested within 10-fold GroupKFold by participant to prevent leakage; 32 proteins were retained for the final model. See Methods for details.

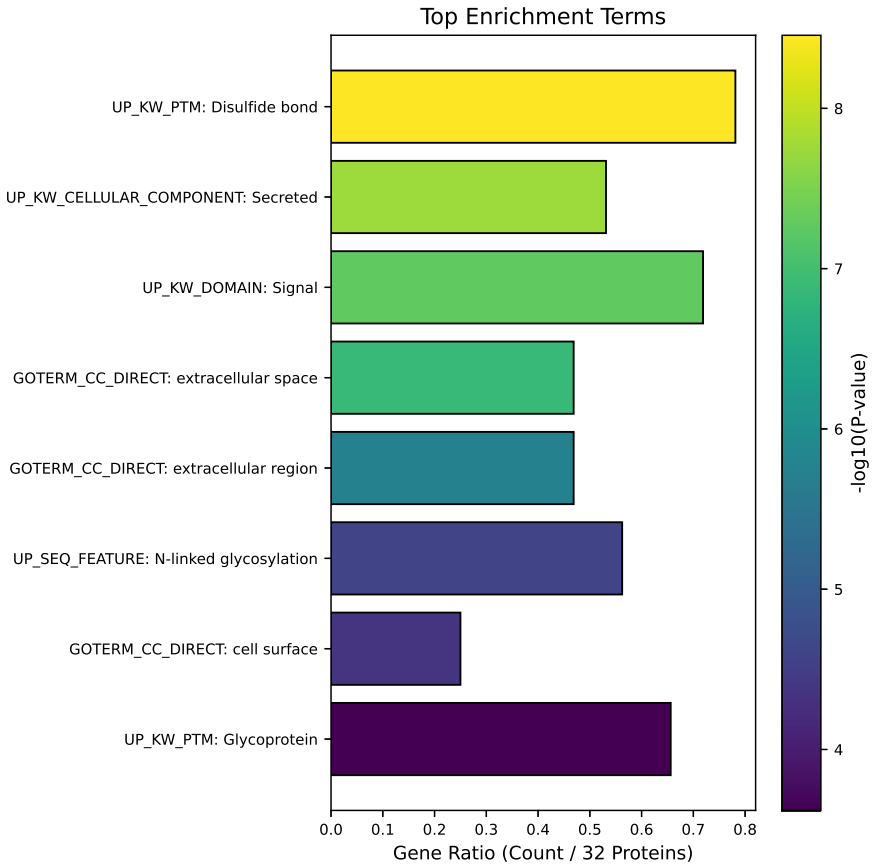

Fig. 4 Functional enrichment of the 32 proteins retained by the proteomics classifier. Bar length shows the gene ratio (hits per 32 proteins); color encodes the enrichment significance as $-\log_{10}(p)$. Terms are drawn from UniProt Keywords (UP_KW), GO Cellular Component (GOTERM_CC_DIRECT), and UniProt sequence features. The top enrichments (e.g., *disulfide bond*, *secreted*, *signal peptide*, *extracellular space/region*, *N-linked glycosylation*, *cell surface*, *glycoprotein*) indicate a cystine-rich, glycosylated, extracellular/secreted proteome consistent with circulating plasma biology. Over-representation tests were performed against the assayed protein background with Benjamini–Hochberg FDR control ($q < 0.05$).

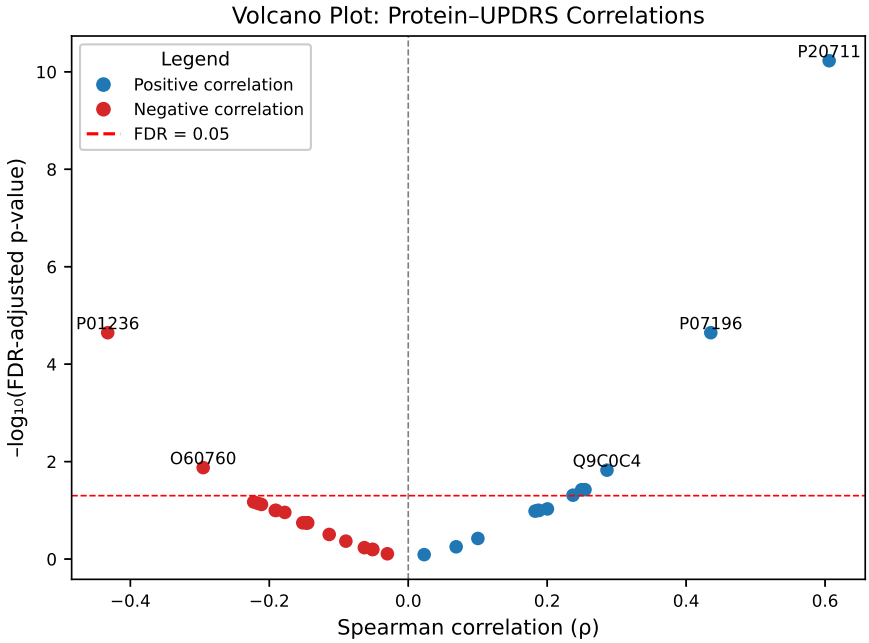

Fig. 5 Plasma proteins correlated with clinical severity (MDS-UPDRS total). Volcano plot of protein–severity associations at baseline (Spearman ρ on the x-axis; $-\log_{10}$(FDR-adjusted p) on the y-axis). Associations were computed within PD participants (baseline draw; multiple testing controlled by Benjamini–Hochberg). Proteins to the right are positively correlated with greater motor or non-motor burden; to the left are negatively correlated (putative protective/compensatory patterns). Representative signals are annotated: DDC (P20711) and NEFL (P07196) show positive correlations, whereas PRL (P01236) and HPGDS (O60760) are negatively correlated. Full statistics are provided in Supplementary Tables; analysis choices and quality controls are described in Methods.