Toward Early Detection of Parkinson's Disease: Integrating Novel Proteomic Biomarkers for Predictive Analysis

Joan Jaylani
Business Analytics
Sacred Heart University
Fairfield, CT USA
joan.jaylani@gmail.com

September 2, 2025

Abstract

This study investigates the predictive utility of proteomic biomarkers for identifying individuals at risk for Parkinson's disease. Proteomic datasets were sourced from the Parkinson's Progression Markers Initiative (PPMI) and included protein and peptide expression values derived from biospecimen assays. To assess the contribution of these molecular features to disease classification, machine learning models such as tree-based classifiers and logistic regression were employed. In the first analysis, logistic regression using cerebrospinal fluid-derived alpha-synuclein achieved a predictive accuracy of 96.5%. In the second analysis, Extreme Gradient Boosting (XGBoost) with pooled 10-fold cross-validation identified 83% of positive cases using a panel of blood-based biomarkers. Notably, neuropeptides such as cholecystokinin and amyloid beta 1-42, the proteotypic peptide LINGO1, and small-molecule metabolites homovanillic acid and ornithine emerged as strong predictors. These findings underscore the promise of proteomic biomarkers for early-stage Parkinson's disease screening and highlight their potential integration into future diagnostic strategies.

Keywords—Parkinson's disease, biomarkers, machine learning, peptide, metabolite, proteins, predictors, alpha-synuclein

1. INTRODUCTION

Parkinson's disease, currently affecting between 1-2% of the population, is a complex and progressive neurological disorder, traditionally difficult to diagnose before symptoms like tremors and stiffness show up. The current method of clinical diagnosis involves observing symptoms over several months, often after significant and irreversible neurological deterioration has occurred. As a result, the emergence of the disease can be substantial and less responsive to available medications that can slow this progression. n

However, newer studies are beginning to shed light on diagnosis through biomarkers, genetic and protein-based signals, that may help detect risk earlier. Techniques such as seed amplification assays of alpha-synuclein and mass spectrometry analysis of peptides, along with other studies for understanding metabolites and proteins, are gaining attention for their potential to detect changes long before physical symptoms appear. This project investigates predictive biomarkers for Parkinson's disease, aiming to enable earlier intervention, potentially before clinical symptoms emerge, to slow disease progression and significantly enhance the quality of life for diagnosed individuals, to surpass what current methods allow.

2. CURRENT DIAGNOSTIC PROCESS

The current standard for diagnosing Parkinson’s Disease and other related neurological disorders is through clinical observation, primarily of motor symptoms such as bradykinesia, stiffness, and rigidity, and their progression over time. While motor disorders have been well researched and have extensive diagnostic tools and treatment strategies, non-motor symptoms such as depression, psychosis, sleep disturbances, and falls are often present even before the motor symptoms and are frequently overlooked.

The delay necessitated by the observation time means that by the time the disease is diagnosed, numerous dopamine-producing neurons in the brain have already been irreparably lost. There is no single lab test or scan that can definitively confirm the disease, which means diagnosis is primarily based on clinical judgment (Postuma et al., 2015).

Doctors may use brain imaging, such as MRI, to rule out other conditions, and a DaTscan can sometimes show dopamine activity; however, it is not a definitive part of the diagnosis. One clue that supports Parkinson's is whether the patient improves when given a medication like levodopa. Still, it is a reactive process that often catches the disease late (Postuma et al., 2015). This delay highlights the need to explore biomarkers that can identify the disease earlier, ideally even before symptoms ever appear.

3. BIOMARKER BREAKTHROUGH – ALPHA-SYNUCLEIN

Although not yet clinically deployed, recent studies by organizations such as the Parkinson's Progression Markers Initiative (PPMI) have uncovered breakthrough findings that may help indicate Parkinson's disease and other related neurological disorders collectively known as alpha-synucleinopathies. This umbrella term includes conditions such as Parkinson's Disease (PD), Dementia with Lewy Bodies (DLB), and Multiple System Atrophy (MSA). In these disorders, alpha-synuclein, a protein found in the brain, misfolds and accumulates into clumps known as Lewy bodies. Researchers at the Caughey Laboratory in Hamilton, Montana, using PPMI data and cerebrospinal fluid samples, were able to predict Parkinson's disease with up to 97% accuracy through FMax (Maximum Fluorescence) and TTT (Time to Threshold) tests (Russo et al., 2021). This project replicated these findings using FMax test data and a logistic regression machine learning model, achieving an accuracy rate of 96.5%. Although a cerebrospinal fluid test could provide a more definitive diagnosis, its invasive and specialized requirements make it unsuitable for early, pre-symptomatic detection.

4. PROTEOMIC BIOMARKER RESEARCH – PEPTIDES, METABOLITES, AND PROTEINS

New to the field of predictive biomarkers, proteomic protein fragments, small-molecule metabolites, and proteins are separated by chromatography and detected by mass spectrometry. This approach has also shown promising results. The proteins are broken down into peptides, then characterized by their sequences as follows: [GeneSymbol]_HUMAN|[PeptideSequence]|[ChargeState]. For example, G3P_HUMAN|LISWYDNEFGYSNR|6+ is separated into G3P_HUMAN: Protein name

Glyceraldehyde-3-phosphate dehydrogenase, LISWYDNEFGYSNR: Detected peptide sequence from mass spectrometry, and 6+: Charge state used in MS-based proteomics (Russo, I. et al. 2021). These biomarkers indicate the current state of the body and are often evident even before symptoms appear. In patients with Parkinson's disease, these proteins exhibit differences, including inflammation, oxidative stress, and neurodegeneration at abnormally elevated or reduced levels or harmful alterations. These abnormalities can lead to neurological inflammation, synaptic degradation, oxidative stress, or mitochondrial dysfunction (Russo et al., 2021).

5. DATA AND METHODS

DATA SOURCES

The primary data source for this study is the Parkinson's Progression Markers Initiative, a global research foundation dedicated to studying Parkinson's Disease. (https://www.ppmi-info.org) This foundation has granted permission to download data to further this research project. Two databases were used for this research, including Biospecimen_Analysis_Results.csv and SAA_Biospecimen_Analysis.csv. The first database in this study, SAA_Biospecimen_Analysis.csv, consists of 1,586 participants with 48 variables related to cerebrospinal fluid tests and test values for the presence of alpha-synuclein. The columns most relevant for this analysis were diagnosis, FMax_24hr_Rep 1 (Maximum Fluorescence in 24 hours, Replicate 1), and TTT_24hr_Rep1 (Time to Threshold in 24 hours, Replicate 1). Test times range from February 2022 to February 2025 and were generated based on research by the USA biopharma company Amprion.

The second database, Biospecimen_Analysis_Results.csv, consists of 487,386 rows of data related to 262 participants. The most relevant columns are the patient ID, diagnosis, biomarker type, test name, and test value. Each patient has multiple records that can include various tests within different biomarkers, such as a plasma test for amino acid, asparagine, or a cerebrospinal fluid test for positive wells of alpha-synuclein. Collection times range from October 2014 to August 2024 and originate from 33 participating academic neurology outpatient practices worldwide, including Austria, Canada, France, Germany, Greece, Israel, Italy, the Netherlands, Norway, Spain, the UK, and the USA (Russo et al., 2021).

OUTCOME VARIABLES FOR ALPHA-SYNUCLEIN ANALYSIS

Using the database SAA_Biospecimen_Analysis.csv, the first approach was to determine if Parkinson's disease could be predicted by tests performed on the biomarker protein, alpha-synuclein, sourced from cerebrospinal fluid. The analysis includes data from patients with Parkinson's disease, along with healthy controls and prodromal subjects. Although 30 different tests were performed related to this biomarker, not all tests were performed on each subject. Additionally, due to the nature of many tests being replications or versions of the same test with different timing, the test variables were highly correlated. Accordingly, of the 30 tests, just FMax_24hr_Rep1 and TTT_24hr_Rep1 were used for the first classification approach. To take a conservative approach and avoid overestimating a positive result, results from prodromal subjects were considered negative for the disease.

OUTCOME VARIABLES FOR PROTEOMIC ANALYSIS

Using the Biospecimen_Analysis_Results.csv database, the analysis focused on protein, metabolite, and protein peptide fragment levels obtained from biospecimen assays conducted by the Parkinson's Progression Markers Initiative (PPMI). The dataset included test results from individuals diagnosed with Parkinson's disease, healthy controls, and subjects at risk of developing the disease (prodromal subjects).

Data was structured by patient ID and biomarker test type, with each patient represented across multiple rows corresponding to different assay results. Consistent with prior analyses, binary classification was applied to offset the overestimation of positive findings: patients diagnosed with Parkinson's disease were assigned a value of 1. In contrast, both healthy controls and prodromal subjects were labeled 0.

MODELING APPROACH FOR ALPHA-SYNUCLEIN CLASSIFICATION

Given that over 80% of participants in the dataset were classified as Parkinson's-positive, the data was imbalanced with a pronounced skew towards positive cases. To handle this disparity, Synthetic Minority Over-sampling Technique (SMOTE) was implemented within each fold to generate synthetic data for the minority class, Parkinson's-negative, to balance the classes and avoid model bias. To further improve the robustness of the model's performance

and mitigate potential variability from a smaller quantity of negative cases, stratified K-fold cross-validation (K=10) was adopted. A logistic regression classifier was then trained on the balanced datasets to predict Parkinson's disease status based on the FMax and TTT variables. The same algorithm was subsequently applied using only the FMax feature to assess whether model accuracy could be further refined.

To evaluate the model's effectiveness, recall, and accuracy were identified as the most clinically relevant metrics. Recall measures the proportion of true positive cases that the model correctly identifies. In the context of disease detection, maximizing true positives and minimizing false negatives is critical since failing to identify a positive case may result in missed opportunities for early intervention or treatment. Accuracy assesses the overall performance of the model by measuring the proportion of all correctly classified positive and negative cases relative to the total number of predictions.

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \qquad \text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}}$$

A second technique, a probability-based threshold loop, was used to identify the optimal decision threshold for the FMax_24h_Rep1 variable. This method evaluated 101 percentile-based cutoff values, ranging from 0% to 100% in 1% increments, to determine the FMax value that best separated Parkinson's-positive from Parkinson's-negative cases. For each threshold, binary classifications were generated by assigning cases above the threshold as positive and those below as negative. The receiver operating characteristic area under the curve (ROC-AUC) was then calculated to assess the ability to discern the cutoff value at each step. The optimal threshold was selected based on the highest ROC-AUC score, indicating the best balance between sensitivity and specificity. The final value serves as a diagnostically significant cutoff, where FMax levels above the threshold indicate a higher chance of Parkinson's disease, while values below suggest a lower likelihood.

MODELING APPROACH FOR PROTEOMIC CLASSIFICATION

To evaluate the proteomic biomarkers in the Biospecimen_Analysis_Results.csv database as a collective tool for diagnostic prediction, the data was pivoted so that each row represented

a single patient, and each column corresponded to either a diagnostic label or a biomarker-related feature. This transformation reshaped the database from 487,386 rows, each representing an individual biomarker test, to rows, each representing a patient, with 4,753 columns capturing biomarker measurements, diagnostic outcomes, and other test-related variables.

To reduce data variance and enhance classification performance, an initial preprocessing step isolated all peptide fragment biomarkers into a new dataframe, narrowing the feature space to 688 columns. Following the approach used in the alpha-synuclein analysis, a 10-fold Stratified K-fold cross-validation strategy was employed to improve model robustness. The data was standardized and then classified using Random Forest, Support Vector Classifier (SVC), Logistic Regression, and Extreme Gradient Boosting (XGBoost). Among these, the SVC model initially performed best, achieving 68.2% accuracy with a recall of 78.1% for true positives (TP) and 55% for true negatives (TN). Notably, reaching the 80% cumulative importance threshold required approximately 300 features. To optimize performance, the analysis was repeated using only the top 50 features identified through feature importance ranking. This refinement yielded a modest performance improvement, with Logistic Regression emerging as the best-performing model, achieving 71% accuracy, 74.5% recall for true positives (TP), and 66% recall for true negatives (TN).

Because the primary goal was to predict a Parkinson's diagnosis as accurately as possible, additional blood-based biomarkers, proteins, and metabolites were incorporated alongside the peptide fragments. Using XGBoost's feature selection capability, the feature set was narrowed to the top 30 predictors across these three biomarker types. In this final iteration, the XGBoost classifier achieved the highest performance, with 81% accuracy and recall rates of 83% for true positives (TP) and 71% for true negatives (TN), representing a significant improvement in diagnostic reliability.

Proteomic biomarkers were next assessed using the p-value calculated by the Mann-Whitney U test, which determines whether the distribution of a particular biomarker is statistically significantly different between individuals with Parkinson's disease and healthy controls. The Mann-Whitney U test is a nonparametric test used to determine whether two

independent groups (e.g., PD vs. controls) have different distributions for a particular feature. (e.g., FMax values) This test is pertinent for this evaluation because it does not assume the data will follow a normal distribution, and it is well-suited for small datasets.

The Mann-Whitney U test statistics are calculated as follows: $U = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1$ where n_1 and n_2 are the sizes of the two groups, and R_1 is the sum of the ranks for group 1. The p-value is calculated from the U statistic and represents the probability of observing a difference in ranks as extreme as the one seen in the data, assuming the null hypothesis of no true difference between the groups is correct. (Conover, 1999). In the Mann-Whitney U test, a small p-value of < 0.05 is used to determine statistical significance. **EXHIBIT F** lists the top five most statistically significant biomarkers, with a total of 30 biomarkers measured as either statistically significant or relevant to feature selection or both. (see **APPENDIX A**)

6. STUDY RESULTS

EVALUATION OF MAXIMUM FLUORESCENCE LEVELS FOR ALPHA-SYNUCLEIN USING LOGISTIC REGRESSION

The logistic regression model demonstrated strong overall performance, achieving an accuracy of 96.5%. In terms of recall, it correctly identified 96.7% of individuals with Parkinson's disease, missing only 3.3% of true positive cases. The model also performed well in identifying true negatives, with a 96% success rate, meaning it incorrectly classified only 4% of negative cases as positive. These results indicate that the model is highly reliable across both classes, exhibiting consistent accuracy and balanced recall. (See **EXHIBIT A**)

EXHIBIT A.

Logistic Regression Classification Report (Pooled 10-Fold CV):

	precision	recall	f1-score	support
0	0.871	0.960	0.914	176
1	0.990	0.967	0.978	747
accuracy			0.965	923
macro avg	0.931	0.963	0.946	923
weighted avg	0.968	0.965	0.966	923

Confusion Matrix:
[[169 7]
[25 722]]
ROC-AUC Score: 0.982

EXHIBIT A displays the logistic regression classifier, which demonstrates excellent performance in identifying Parkinson's disease cases using 10-fold cross-validation. The model achieved an overall accuracy of 96.5%, correctly classifying both Parkinson's-positive and Parkinson's-negative subjects with high reliability.

The classification demonstrated excellent diagnostic performance. Recall for Parkinson's-positive cases (class 1) was 96.7%, indicating that the model correctly identified nearly all true positive cases, with only 3.3% missed. Recall for Parkinson's-negative cases (class 0) was also high at 96.0%, which points to a low rate of false positives. Precision was also notably high for both classes, particularly for class 1 at 99.0%, reflecting minimal false positives in disease prediction. The F1 scores reflect this model's performance, with a score of 97.8% for class 1 and a score of 91.4% for class 0, indicating a good balance between detecting positive cases and avoiding misclassification. Overall, these results suggest that the logistic regression model is effective at classifying patients using alpha-synuclein biomarker data.

Visually observing FMax values split by positive and negative results, we can see a pronounced difference between relative fluorescent unit (RFU) levels. (see **EXHIBIT B**) Following the probability-based threshold loop, it was determined that the split point for a higher chance of Parkinson's disease is 25,219.4 RFU (Relative Fluorescence Units), which corresponds to 20% of the value range. (see **EXHIBIT C**)

EXHIBIT B.

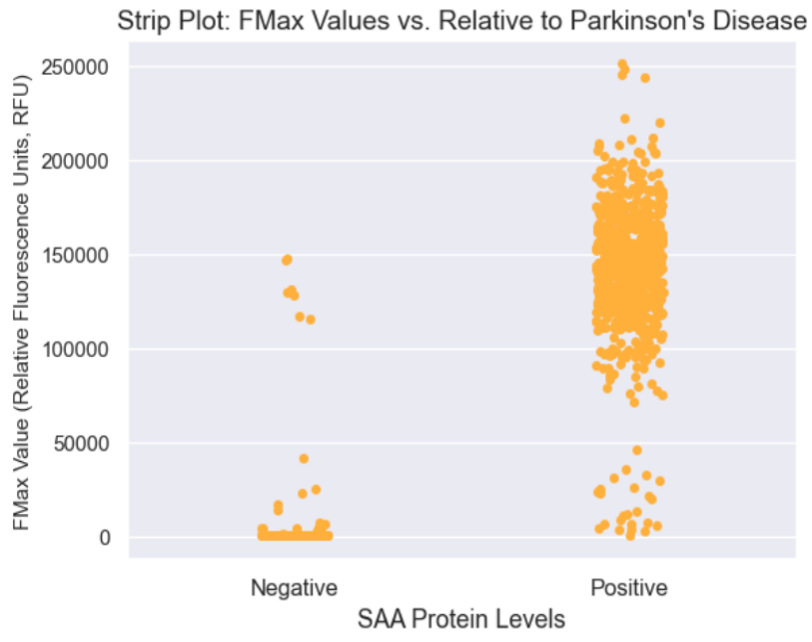


EXHIBIT B conveys a strip plot illustrating the variability in FMax values (measured in Relative Fluorescence Units, RFU) associated with alpha-synuclein levels in patients with Parkinson's disease compared to healthy controls. Each dot represents an individual patient. Notably, the graph shows a higher concentration of data points for individuals with Parkinson's disease, along with consistently elevated FMax values. This pattern suggests both a greater number of detectable alpha-synuclein aggregates and stronger signal intensity among Parkinson 's-positive patients, reinforcing the biomarker's potential diagnostic utility.

EXHIBIT C.

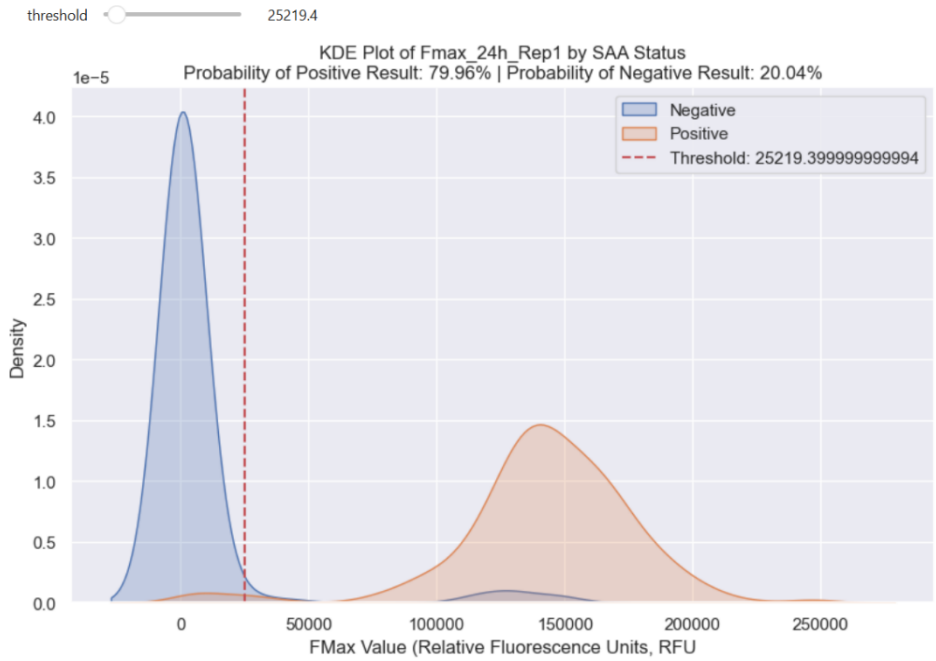


EXHIBIT C. displays a Kernel Density Estimate (KDE) plot comparing FMax values between individuals with and without Parkinson's disease, as determined by Seed Amplification Assay (SAA) status. This analysis focuses on the Fmax_24h_Rep1 feature, which represents the fluorescence intensity after 24 hours of alpha-synuclein aggregation.

Considering the Kernel Density Estimate (KDE) graph above (**EXHIBIT C**), we can see two clearly delineated distributions with the blue curve indicating individuals with a negative SAA result, meaning negative for Parkinson's disease, and the orange curve indicating individuals with a positive SAA result, meaning positive for Parkinson's disease. The vertical dashed line at approximately 25,219 RFU represents the threshold determined by probability, where values below the threshold of 20.04% are more likely to be negative for Parkinson's, and values above the threshold of 79.96% are more likely to be positive for Parkinson's.

The classification at **EXHIBIT D** below reports the optimal maximum fluorescence (FMax) threshold value with an accuracy level of 97.3%. The true positive recall of 97.7% and true negative result of 95.5% indicate the model performed consistently well across both Parkinson's-positive and negative classes, with strong metrics. These results reinforce the clinical potential of FMax as a differentiating biomarker for the detection of Parkinson's disease.

EXHIBIT D

Best Percentile: 20.00%
Best Threshold (FMax): 25219.400
Best ROC-AUC: 0.966

Classification Report:

	precision	recall	f1-score	support
0	0.908	0.955	0.931	176
1	0.989	0.977	0.983	747
accuracy			0.973	923
macro avg	0.949	0.966	0.957	923
weighted avg	0.974	0.973	0.973	923

EXHIBIT D. displays a Classification Report showing the optimal threshold FMax value for a positive or negative determination of Parkinson's disease and the likelihood of accuracy for this probability model. The high precision and recall results indicate that this biomarker shows potential for clinical application.

ANALYSIS OF PROTEOMIC BIOMARKERS FOR PREDICTING PARKINSON'S DISEASE

The XGBoost model achieved an overall accuracy of 81% in differentiating Parkinson's disease cases from controls using a combination of protein peptide, metabolites, and protein biomarkers. The model, sourced from the 30 most closely related biomarkers, demonstrated strong recall and sensitivity, correctly identifying 84% of individuals with Parkinson's disease. However, 16% of true cases were incorrectly classified as negative, indicating that some positive cases were missed. For the healthy control cases, 76% of individuals were correctly identified as negative. Precision in accurately categorizing the disease was 82% for positive predictions and 79% for negative predictions, reflecting a balanced and reliable classification performance across both groups. However, further validation and refinement are needed before the model can be considered for definitive diagnostic use. (see **EXHIBIT E**)

EXHIBIT E

XGBoost - Classification Report (Pooled 10-Fold CV):

	precision	recall	f1-score	support
0	0.79	0.76	0.77	83
1	0.82	0.84	0.83	109
accuracy			0.81	192
macro avg	0.80	0.80	0.80	192
weighted avg	0.81	0.81	0.81	192

EXHIBIT E presents a classification report summarizing the performance of the XGBoost model. The observed precision and recall values suggest that the use of proteomic biomarkers holds promise for clinical application in Parkinson's disease detection.

EXHIBIT F lists the top five most statistically significant biomarkers, according to the Mann-Whitney p-value test, with a total of 10 biomarkers determined to be statistically significant with a p-value of < 0.05%. Using the XG Boost classifier, 20 additional biomarkers were selected based on their level of importance to run the classification algorithm, resulting in a total of 30 biomarkers analyzed. (see **APPENDIX A**)

EXHIBIT F

Highest Statistical Significance:

Biomarker	Category	p-value	PD Association / Explanation
Homovanillic_acid	Metabolite	2.73E-16	Major dopamine metabolite in CSF; reduced levels are consistent with dopaminergic deficits in PD.
CCKN_HUMAN|AHLGALLAR|y6+	Protein peptide	6.39E-05	Fragment of cholecystokinin (CCK); altered CCK signaling and levels reported in PD may influence non-motor symptoms.
Ornithine	Metabolite	0.00013	Metabolite in the urea cycle; altered levels linked to oxidative stress in PD
LIGO1_HUMAN|ATVPFPFDIK|y5+	Protein Peptide	0.000237	Fragment of LINGO1; LINGO1 is implicated in neurodegeneration, myelination, and is a potential target for neuroprotection in PD.
Abeta1-42	Protein peptide	0.000362	Amyloid-beta peptide; decreased CSF levels observed in PD with dementia, a classic AD marker.

EXHIBIT F details observations regarding the top five most statistically significant biomarkers, which include both protein peptide fragments and metabolites. In addition to this analysis, each biomarker has been separately observed to vary among individuals with Parkinson's disease. (Wishart et al., 2023), (UniProt Consortium, 2024)

The graphs below (**EXHIBIT G & EXHIBIT H**) provide a visual representation of the variation between test values for the individual biomarkers. The differences in Homovanillic acid values between individuals with Parkinson's disease and healthy controls are notable, with even casual observation revealing a high density of values within a narrow range for the healthy group. While values falling outside this range alone are not sufficient to diagnose Parkinson's disease, analyzing this variance, especially when considered alongside other significant differences in biomarkers, can help identify cases that may warrant further evaluation.

EXHIBIT G

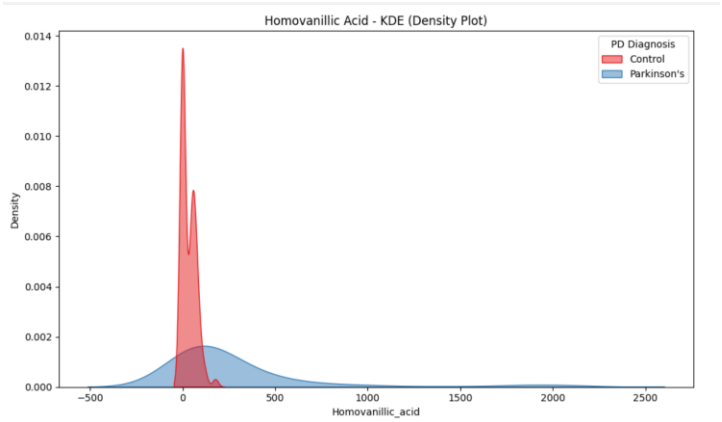


EXHIBIT G The blue values represent healthy control individuals compared to the widely dispersed values in red associated with Parkinson's disease. Because all control-related values overlap with disease-positive values, this metabolite can not be an indicator on its own. However, values outside the narrow band of healthy individuals suggest a need for further investigation.

EXHIBIT H

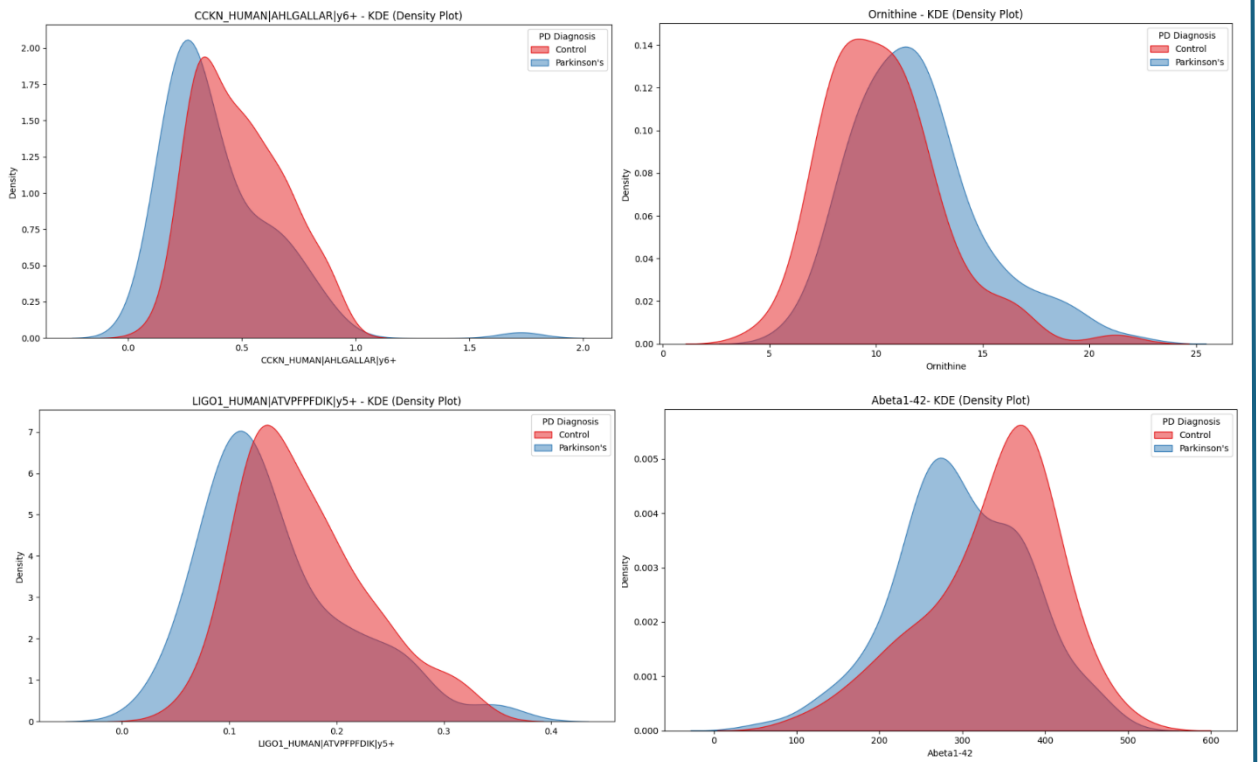


EXHIBIT H The Kernel Density Estimate graphs associated with these next four biomarkers do not show as marked of a difference in density as the Homovanillic acid graph. (Exhibit G) There are statistically significant differences between control-related values and Parkinson 's-related values within each graph; however, taken together, they all create a more substantial probability of being able to predict the disease accurately.

In addition to selecting statistically significant features, the XGBoost classifier analyzed how features interacted with each other to classify the data accurately. The biomarkers were ranked

according to their ability to assist the model's prediction. Some biomarkers do not have a known correlation with Parkinson's disease; however, because they contribute to a moderately accurate predictor, further study of these biomarkers and their interactions may be warranted. Please refer to **APPENDIX A** for a comprehensive list of biomarkers, their corresponding biological categories, and their functions within the body.

7. STUDY LIMITATIONS

The initial study analyzing the alpha-synuclein biomarker demonstrates an excellent ability to predict Parkinson's based on FMax levels. However, the data available in the analyzed database, SAA_Biospecimen_Analysis.csv, was highly skewed, with 745 positive Parkinson's cases and only 21 negative cases after data cleansing. The study attempts to mitigate this imbalance by generating synthetic data (SMOTE) for the negative class; however, this may be insufficient to offset the disproportion in the data. On a favorable note, as similar observations have been documented in other studies (Russo et al., 2021), these results likely remain accurate. It is also important to note that the majority of PPMI study participants are of European genetic ancestry, which may limit how well these findings extend to people of different racial and ethnic origins.

The proteomic analysis in this study included 109 Parkinson's-positive cases and 83 negative controls, providing a more balanced dataset. However, the relatively small sample size may limit the robustness of the findings. TTo improve the robustness of the findings, K-fold cross-validation was used, enabling the model to be evaluated repeatedly across various data subsets. Although the participant pool was relatively small, the dataset itself was highly complex, containing 487,386 variables representing biomarkers and other test-related features. This high dimensionality introduced considerable variance, and over 300 variables were needed to achieve even moderate predictive performance. This challenge was addressed in part by focusing the analysis on protein peptides and a small subset of metabolites and proteins. However, this manual selection process may have further constrained the results.

It should be acknowledged that all biomarker measurements in the Biospecimen_Analysis_Results.csv file were derived from cerebrospinal fluid. To prioritize

clinical feasibility, only biomarkers that could also be measured in blood were considered for inclusion in this study. The motivation for prioritizing blood-based biomarkers stems from the fact that lumbar puncture is an invasive procedure that requires specialized expertise and training. In contrast, blood collection is a routine and accessible procedure for most patients. While these selected biomarkers are clinically relevant, blood-based assays generally have lower sensitivity and specificity compared to cerebrospinal fluid testing.

Another challenge with proteomic biomarkers is that while they are collectively predictive, many may be too general individually. Because their values can be related to a wide range of unrelated conditions, these markers may lack the specificity needed for accurate diagnosis of a particular disease. For example, impaired glucose metabolism is associated with mitochondrial dysfunction in Parkinson's (Wishart et al., 2023) but is also associated with Alzheimer's Disease, Amyotrophic Lateral Sclerosis (ALS), Multiple Sclerosis (MS) (Chase, 2014), and several conditions that are not even neurologically related such as Type 2 Diabetes. Another example is the protein peptide hormone and neuropeptide Cholecystokinin (CCK), which is found in the brain, and is associated with dopamine suppression and may relate to neuropsychiatric symptoms in patients with Parkinson's disease. (Brimblecombe KR, Cragg SJ., 2015) However, CCK is also associated with disparate other conditions such as schizophrenia (Harada et al., 2021) and pancreatic cancer (Smith, J. P. et al., 2014).

Finally, similar to the alpha-synuclein analysis, the participants in the PPMI study are predominantly composed of individuals with European genetic backgrounds. As a result, the applicability of these findings to other racial and ethnic populations may be limited.

8. DISCUSSION

REGARDING THE ALPHA-SYNUCLEIN BIOMARKER

Alpha-synuclein has been recognized as a key player in Parkinson's disease since 1997, when it was first identified as a significant component of Lewy bodies, the distinguishing pathological characteristic of Parkinson's, as well as other neurological disorders, including dementia with Lewy bodies, multiple system atrophy, and Alzheimer's disease (Wang, 2016). More recently, advances in laboratory techniques have enabled the use of the alpha-synuclein

seed amplification assay on cerebrospinal fluid samples, allowing for the sensitive detection of alpha-synuclein aggregates. This biomarker has further confirmed its central role in these diseases, with a 93% prediction rate (Siderow et al., 2023).

The results of this study are consistent with recent advances in the field of neurology. The association between alpha-synuclein and Parkinson's disease was demonstrated with a high accuracy of 96.5% using logistic regression analysis. Additionally, using incremental iterations of probability analysis, a numerical threshold value for distinguishing between Parkinson's positive and negative cases was established using the widely accepted FMax test.

REGARDING OTHER PROTEOMIC BIOMARKERS

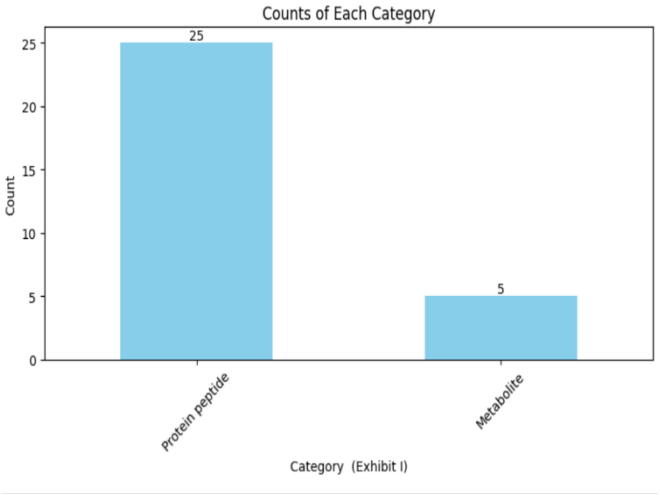
In addition to alpha-synuclein studies, other proteomic biomarkers have also shown promising relationships with Parkinson's disease. Integrating results from protein peptides, metabolites, and protein biomarkers, although not definitive in diagnosing Parkinson's disease, demonstrates strong predictive potential. Using feature selection, the top 30 biomarkers, with a 5:1 ratio of protein peptides to metabolites, achieved 81% accuracy, correctly categorizing 83% of individuals with Parkinson's disease. The methods used for measurement were as follows:

- Mass spectrometry was used for the protein peptides, although two, Abeta1-42 and GPNMB_HUMAN|AYVPIAQVK|y6+, could also be analyzed using ELISA (Enzyme-Linked Immunosorbent Assay).
- Clinical chemistry, typically available in hospitals or larger clinical laboratories, was used to analyze glucose levels.
- (HPLC) High-Performance Liquid Chromatography and Gas Chromatography-Mass Spectrometry (GC-MS) were used in tandem to measure metabolites, including Homovanillic acid, Ornithine, DOPAC, and 24(S)-hydroxycholesterol.

Of these methods, only clinical chemistry is currently available for general diagnostic use. ELISA (Enzyme-Linked Immunosorbent Assay) is used particularly for specialized proteins such as alpha-synuclein and protein peptide, Abeta1-42, and is only available in laboratories, reference labs, or research hospitals. Mass spectrometry and its subcategories, LC-MS/MS

(Liquid Chromatography–Tandem Mass Spectrometry) and GC-MS (Gas Chromatography–Mass Spectrometry), are currently only available in niche research facilities and are unavailable for general diagnostic use (Wenc et al., 2024). Mass spectrometry and ELISA require both state-of-the-art equipment and highly skilled laboratory staff trained to manage the equipment and perform diagnostic procedures. Additionally, strict regulations apply to clinical laboratories that are not required for research laboratories.

However, researchers are actively working to develop automated mass spectrometry analyzers that are suitable for routine clinical use. Efforts are also underway to improve software integration between Laboratory Information Management Systems (LIMS) and MS instrumentation (Wenc et al., 2024). Additionally, new strategies are being explored to streamline sample preparation, with the goal of balancing assay sensitivity and turnaround time with the practicality needed for widespread clinical implementation (Wenc et al., 2024).



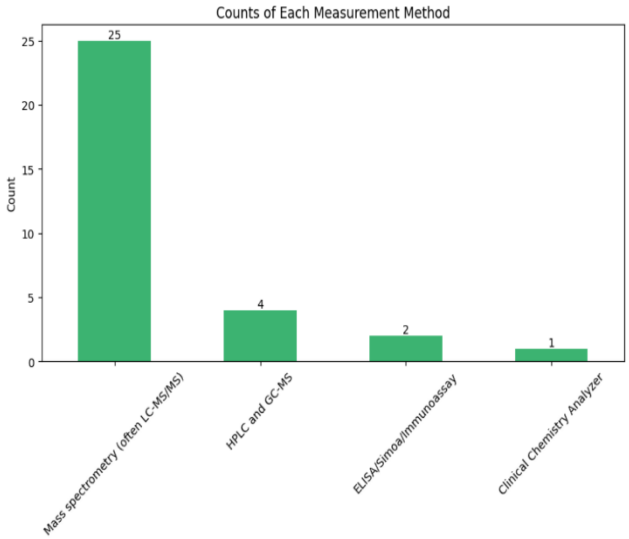


EXHIBIT I present the different types of proteomic features included in the study, along with the number of each type reviewed.

EXHIBIT J details the various testing methods currently required to accurately assess these biomarkers.

9. CONCLUSION

IMPLICATIONS FOR CLINICAL PRACTICE

Accurate early detection of Parkinson's disease is critical for timely intervention and better patient outcomes. The strong sensitivity demonstrated by these models, consistent with findings from other research studies, suggests that they can support early diagnosis by identifying the majority of true positive cases. If clinical laboratories can adopt user-friendly equipment and methods, blood-based testing could potentially become part of routine screening for individuals over the age of 65, the average age of Parkinson's onset (Michael J. Fox Foundation, 2025). This breakthrough could have a significant impact on individuals with Parkinson's disease as it may enable detection up to seven years before symptom onset. Early diagnosis and intervention are critical for delaying symptom progression and improving long-term outcomes.

However, particularly when considering the use of multiple proteomic biomarkers, significant obstacles remain to routine clinical implementation. Regulatory challenges, the need for equipment modifications, and the limited specificity of specific biomarkers remain barriers, along with the necessity for greater model accuracy to enhance predictive performance. The incorrectly classified negative results observed in this study underscore the importance of ongoing refinement to reduce missed diagnoses and ensure that more patients can benefit from early intervention and monitoring.

APPENDIX A (abridged)

Feature	Category	p-value	PD Association / Explanation
Homovanillic_acid	Metabolite	2.73E-16	Major dopamine metabolite in CSF; reduced levels are consistent with dopaminergic deficits in PD.
CCKN_HUMAN|AHLGALLAR|y6+	Protein peptide	6.39E-05	Fragment of cholecystokinin, altered CCK signaling and levels reported in PD, may influence non-motor symptoms.
Ornithine	Metabolite	0.00013	Urea cycle intermediate; some metabolomic studies show altered ornithine in PD, linked to oxidative stress.
LIGO1_HUMAN|ATVPFPFDIK|y5+	Protein peptide	0.00024	Fragment of LINGO1; LINGO1 implicated in neurodegeneration, myelination and potential target for neuroprotection in PD.
Abeta1-42	Protein peptide	0.00036	Amyloid-beta peptide; decreased CSF levels observed in PD with dementia, classic AD marker.
NPTXR_HUMAN|LVEAFGGATK|y5+	Protein peptide	0.00147	Fragment of Neuronal pentraxin receptor (NPTXR); NPTXR is involved in synaptic plasticity, reduced levels linked to synaptic dysfunction in PD.
DOPAC	Metabolite	0.00220	Major dopamine metabolite; reduced CSF DOPAC in PD, reflects dopaminergic neuron loss.
SCG1_HUMAN|GEAGAPGEEDIQGPTK|y4+	Protein peptide	0.00335	Secretogranin-1 fragment; secretogranins may be altered in PD and other neurodegenerative diseases.
SLIK1_HUMAN|SLPVDVFAGVSLSK|y7+	Protein peptide	0.00618	SLAIN motif-containing protein one peptide; not explicitly linked to PD but involved in neuronal microtubule regulation.
SLIK1_HUMAN|SLPVDVFAGVSLSK|y8+	Protein peptide	0.03480	SLAIN motif-containing protein one peptide; not specifically linked to PD but involved in neuronal microtubule regulation.
24(S)-hydroxycholesterol	Metabolite	0.08150	Cholesterol metabolite; altered brain levels seen in PD and neurodegeneration.
Glucose	Metabolite	0.08550	Altered glucose metabolism reported in PD; insulin resistance and metabolic dysfunction may be comorbid.
FIBG_HUMAN|YLQEIYNSNNQK|b4+	Protein peptide	0.08880	Fibrinogen gamma chain peptide; inflammation and coagulation abnormalities observed in PD.
PEDF_HUMAN|DTDTGALLFIGK|y4+	Protein peptide	0.10610	Pigment epithelium-derived factor peptide; PEDF has neuroprotective roles and may be reduced in PD.
KAIN_HUMAN|WADLSGITK|y8+	Protein peptide	0.11900	Kainate receptor subunit peptide (GRIK1); glutamate signaling involved in PD pathophysiology.
GPNMB_HUMAN|AYVPIAQVK|y6+	Protein peptide	0.13200	GPNMB (glycoprotein non-metastatic melanoma protein B) peptide; increased GPNMB expression reported in PD brain.
CFAI_HUMAN|VFSLQWGEVK|y8+	Protein peptide	0.15918	Complement factor I peptide; regulates complement, altered complement activity and neuroinflammation implicated in PD.
PON1_HUMAN|IFFYDSENPPASEVLR|y8+	Protein peptide	0.16600	Paraoxonase 1 fragment; PON1 involved in oxidative stress and reported to be altered in PD.

APOL1_HUMAN|VAQELEEK|b6+	Protein peptide	0.16700	Apolipoprotein L1 peptide; APOL1 polymorphisms linked to neurodegeneration and inflammation.
COCH_HUMAN|GVISNSGGPVR|y7+	Protein peptide	0.17700	Cochlin peptide; cochlin mutations more associated with hearing loss, not directly with PD.
SODC_HUMAN|HVGDLGNVTADK|y3+	Protein peptide	0.22900	Superoxide dismutase [Cu-Zn] (SOD1) peptide; oxidative stress and SOD1 dysregulation implicated in PD.
C1S_HUMAN|SNALDIIFQTDLTGQK|y7+	Protein peptide	0.25800	Complement C1s subcomponent peptide; complement system activation reported in PD brains.
CFAI_HUMAN|VFSLQWGEVK|y6+	Protein peptide	0.52400	Complement factor I peptide; regulates complement, altered complement activity and neuroinflammation implicated in PD.
TRFE_HUMAN|HSTIFENLANK|y5+	Protein peptide	0.54100	Transferrin peptide: Iron metabolism is disrupted in PD, and transferrin is an iron-binding protein.
IPSP_HUMAN|AVVEVDESGTR|y3+	Protein peptide	0.60200	Peptide from serine protease inhibitor; direct role in PD not established.
CBLN1_HUMAN|STFIAPR|y3+	Protein peptide	0.61400	Cerebellin-1 peptide (CBLN1) has synaptic functions and is downregulated in some synucleinopathies.
PTGDS_HUMAN|AQGFTEDTIVFLPQTDK|y7+	Protein peptide	0.76700	Prostaglandin-H2 D-isomerase fragment; prostaglandin pathways altered in PD.
SEM7A_HUMAN|VSLAPNSR|y5+	Protein peptide	0.77000	Semaphorin-7A peptide: Semaphorins are involved in axon guidance and neuroinflammation, with limited literature on PD.
ITIH5_HUMAN|SYLEITPSR|y4+	Protein peptide	0.90400	Inter-alpha-trypsin inhibitor heavy chain H5 peptide; extracellular matrix protein, not directly established in PD.
CYTM_HUMAN|DLSPDDPQVQK|b3+	Protein peptide	0.98200	Cytochrome c fragment; mitochondrial dysfunction is a core feature in PD pathogenesis.

** Full appendix with measurement methods and references available in a separate file – please email the author.*

PPMI PARTNER ACKNOWLEDGEMENTS

PPMI – a public-private partnership – is funded by the Michael J. Fox Foundation for Parkinson's Research and funding partners, including 4D Pharma, Abbvie, AcureX, Allergan, Amathus Therapeutics, Aligning Science Across Parkinson's, AskBio, Avid Radiopharmaceuticals, BIAL, BioArctic, Biogen, Biohaven, BioLegend, BlueRock Therapeutics, Bristol-Myers Squibb, Calico Labs, Capsida Biotherapeutics, Celgene, Cerevel Therapeutics, Coave Therapeutics, DaCapo Brainscience, Denali, Edmond J. Safra Foundation, Eli Lilly, Gain Therapeutics, GE HealthCare, Genentech, GSK, Golub Capital, Handl Therapeutics, Insitro, Jazz Pharmaceuticals, Johnson & Johnson Innovative Medicine, Lundbeck, Merck, Meso Scale Discovery, Mission Therapeutics, Neurocrine Biosciences, Neuron23, Neuropore, Pfizer, Piramal, Prevail Therapeutics, Roche,

Sanofi, Servier, Sun Pharma Advanced Research Company, Takeda, Teva, UCB, Vanqua Bio, Verily, Voyager Therapeutics, the Weston Family Foundation and Yumanity Therapeutics.

PPMI DATA ACKNOWLEDGEMENTS

Data used in the preparation of this article were obtained on 2024-09-18 from the Parkinson's Progression Markers Initiative (PPMI) database (https://www.ppmi-info.org/access-data-specimens/download-data), RRID:SCR_006431. For up-to-date information on the study, visit http://www.ppmi-info.org.

REFERENCES

1. Brimblecombe, K. R., & Cragg, S. J. (2015). Peptide neurotransmitters in the basal ganglia: Roles in Parkinson's disease. Frontiers in Synaptic Neuroscience, 7, 4. https://doi.org/10.3389/fnsyn.2015.00004

2. Chase, A. A (2014). A role for serotonergic neurons in levodopa-induced dyskinesia. Nat Rev Neurol 10, 180. https://doi.org/10.1038/nrneurol.2014.41

3. Chaudhuri KR, Sauerbier A. (2016). Parkinson's disease. Unraveling the nonmotor mysteries of Parkinson's disease. Nat Rev Neurol.;12(1):10–1. doi: 10.1038/nrneurol.2015.236. PMID: 26714657.

4. Conover, W. J. (1999). Practical nonparametric statistics (3rd ed.). Wiley.

5. Gibbons CH, Levine T, Adler C, et al. (2024) Skin Biopsy Detection of Phosphorylated α-Synuclein in Patients with Synucleinopathies. JAMA. 331(15):1298–1306. doi:10.1001/jama.2024.0792

6. Hällqvist, J., Bartl, M., Dakna, M. et al. (2024). Plasma proteomics identifies biomarkers predicting Parkinson's disease up to 7 years before symptom onset. Nat Commun 15, 4759. https://doi.org/10.1038/s41467-024-48961-3

7. Harada, S., Tsuboi, D., & Sato, M. (2021). Cholecystokinin in schizophrenia: A mini review. Frontiers in Neuroscience, 15, 711008. https://doi.org/10.3389/fnins.2021.711008

8. Janelidze, S., et al. (2022). Cerebrospinal Fluid α-Synuclein Seed Amplification in Synucleinopathies: A Prospective Cohort Study. JAMA Neurology, 79(4), 399–406. https://jamanetwork.com/journals/jamaneurology/fullarticle/2790015

9. Michael J. Fox Foundation. Parkinson's Progression Markers Initiative (PPMI) dataset. (2025)

10. National Institute of Neurological Disorders and Stroke (NINDS). Parkinson's Disease Biomarkers Program. https://www.ninds.nih.gov

11. Postuma RB, et al., (2015) Clinical diagnostic criteria for Parkinson's disease. Mov Disord. (12):1591–601. doi: 10.1002/mds.26424. PMID: 26474316.

12. Postuma, R. B., et al. (2015). MDS clinical diagnostic criteria for Parkinson's disease. Movement Disorders, 30(12), 1591–1601.

13. Russo, I., et al. (2021). Proteomic and transcriptomic changes associated with PD. Cell Reports.

14. Russo, M. J., et al., (2021). High diagnostic performance of independent alpha-synuclein seed amplification assays for detection of early Parkinson's disease. Acta Neuropathologica Communications, 9(1), 179. https://doi.org/10.1186/s40478-021-01282-8

15. Siderowf, A., et al. (2023). Assessment of heterogeneity among participants in the Parkinson's Progression Markers Initiative cohort using α-synuclein seed amplification: A cross-sectional study. The Lancet Neurology, 22(5), 407–417. https://www.thelancet.com/journals/laneur/article/PIIS1474-4422(23)00109-6/abstract

16. Simuni, T., et al. (2024). A biological definition of neuronal α-synuclein disease: Towards an integrated staging system for research. The Lancet Neurology, 23(2), 178–190. https://www.thelancet.com/journals/laneur/article/PIIS1474-4422(23)00405-2/abstract

17. Smith, J. P., & Solomon, T. E. (2014). Cholecystokinin: An old hormone with a new role in pancreatic cancer. World Journal of Gastroenterology, 20(44), 16060–16068. https://doi.org/10.3748/wjg.v20.i44.16060

18. The UniProt Consortium. UniProt: the Universal Protein Knowledgebase in 2025. Retrieved May 29, 2025, from https://www.uniprot.org/uniprotkb/P06307/entry

19. Wang, C., et al. (2016). Versatile Structures of α-Synuclein. Frontiers in Molecular Neuroscience, 9. https://www.frontiersin.org/journals/molecular-neuroscience/articles/10.3389/fnmol.2016.00048

20. Wang, Z., Gilliland, T., Kim, H.J. et al. (2024). A minimally invasive biomarker for sensitive and accurate diagnosis of Parkinson's disease. *Acta Neuropathol commun* 12, 167 (2024). https://doi.org/10.1186/s40478-024-01873-1 12, 167

21. Wenk D, Zuo C, Kislinger T, Sepiashvili L. (2024). Recent developments in mass-spectrometry-based targeted proteomics of clinical cancer biomarkers. Clin Proteomics.;21(1):6. doi: 10.1186/s12014-024-09452-1. PMID: 38287260; PMCID: PMC10826105.

22. Wishart, D. S., Feunang, et al. (2023). Human Metabolome Database: HMDB0000118. Retrieved May 29, 2025, from https://hmdb.ca/metabolites/HMDB0000118

23. Wishart, D. S., Feunang, et al. (2023). Human Metabolome Database: Ornithine (HMDB0000214). Retrieved May 29, 2025, from https://hmdb.ca/metabolites/HMDB0000214

24. Yazdani, U. et al. (2016). Blood biomarker for Parkinson's disease: peptoids. npj Parkinson's Disease, https://utsouthwestern.elsevierpure.com/en/publications/blood-biomarker-for-parkinson-disease-peptoids/fingerprints/ 2: 16012; doi:10.1038.