

Detection of prodromal Parkinson's disease using a urine proteomics panel and machine learning

Jenny Hällqvist^{1, †}, Sebastian R. Schreglmann^{2, 3, †}, Kristina Kulcsarova^{4, 5, 6}, Matej Skorvanek^{4, 5}, Eva Feketeova^{4, 5}, Mie Rizig⁷, Brit Mollenhauer^{8, 9}, Paola Turano¹⁰, Maria Giulia Bacalini¹¹, Claudio Francheschi¹¹, Nicholas Wood², Kailash P. Bhatia^{2, †}, Kevin Mills^{1, †}

[†]These authors contributed equally to this work.

Abstract

Parkinson's disease is a progressive neurodegenerative disorder and idiopathic REM-sleep behaviour disorder (iRBD) has been identified as its single most specific early symptom. To facilitate the screening of individuals at high risk to develop Parkinson's disease, we developed a multiplexed panel of urine proteomics using machine learning and targeted mass spectrometry to detect iRBD. Random urine samples from clinically and genetically well characterized patients with iRBD, idiopathic and hereditary forms of Parkinson's disease, and matching controls, collected in two academic centres, were analysed in a standardized way.

First, a biomarker discovery and exploratory comparison of samples from randomly selected idiopathic Parkinson's patients and age/sex-matched healthy controls were proteomically profiled and quantified (> 2500 proteins). The most differentially expressed biomarkers were combined into a high-throughput, multiplexed assay using targeted proteomics designed for use of tandem mass spectrometers for potential translation into clinical practise. This was then validated on independent patient and control samples (n=184). After detecting a major influence of sex on the proteome, we focused subsequent analyses on the larger group of available male samples (n = 114) and report results based on iRBD (n = 14), idiopathic (n = 35) and young-onset Parkinson's disease (n = 15), carriers of LRRK2 (n = 10), PARKIN-gene mutations (n = 5), and healthy control subjects (n = 35). After establishing excellent compatibility between the two study sites, orthogonal partial least squares discriminant analysis (OPLS-DA) excluded a relevant effect of aging, but detected significant differences between iRBD and healthy controls (ANOVA-CV $P = 0.002$), as well as the combination of iPD/iRBD and healthy controls (ANOVA-CV $P = 0.01$).

Uni- and multivariate analyses detected a shared expression pattern for the protein biomarkers UBC, NCAM1, MIEN1, SPP2, REG1B, ITIH2, BCHE and C3 between iRBD and idiopathic Parkinson's disease. Utilizing split train/test-datasets in a multiple-regression classifier model resulted in a mean accuracy of

It is made available under a [CC-BY-NC-ND 4.0 International license](#).

78% to detect iRBD, matching iRBD's conversion rate to Parkinson's disease. Hierarchical clustering revealed greater similarities in urine proteomic changes between iRBD and idiopathic than monogenic Parkinson's disease. Several proteins identified correlated either with clinical severity (e.g. VCAM1, MSN, HPX), or risk for future conversion to Parkinson's disease (VCAM1, MSN, MYO10, HSPAIL).

This demonstrates the power of machine learning and urine biomarkers to identify iRBD patients. As we develop new therapies and interventions, the ability to detect individuals at-risk of neurodegeneration in very early disease stage will be invaluable for treatment success.

Author affiliations:

- ¹ UCL Institute of Child Health and Great Ormond Street Hospital, WC1N 1EH London, UK
- ² Clinical and Movement Neurosciences, UCL Queen Square Institute of Neurology, WC1N 3BG London, UK
- ³ Department of Neurology, University Hospital Würzburg, 97080 Würzburg, Germany
- ⁴ Department of Neurology, Faculty of Medicine, P. J. Safarik University and University Hospital L. Pasteur, 040 11 Kosice, Slovak Republic
- ⁵ Department of Neurology, University Hospital of L. Pasteur, 040 11 Kosice, Slovak Republic
- ⁶ Department of Clinical Neuroscience, Scientific Park MEDIPARK, P. J. Safarik University, 040 11 Kosice, Slovak Republic
- ⁷ Department of Neuromuscular Diseases, UCL Queen Square Institute of Neurology, London WC1N 3BG, United Kingdom.
- ⁸ Department of Neurology, University Medical Center Goettingen, Goettingen, Germany
- ⁹ Paracelsus-Elena Klinik, Kassel, Germany
- ¹⁰ Consorzio Interuniversitario Risonanze Magnetiche di Metallo Proteine (CIRMMP) Sesto Fiorentino 50019 Italy
- ¹¹ IRCCS Istituto Delle Scienze Neurologiche di Bologna, Bologna, Italy

Correspondence to: Prof Kevin Mills

Full address **Inborn Errors of Metabolism Section**
UCL Great Ormond Street Institute of Child Health
30 Guilford Street
London
WC1N 1EH

E-mail kevin.mills@ucl.ac.uk

Running title: Maximum 40 characters including spaces

Keywords: Parkinson's disease; biomarker; mass-spectrometry; prodromal; pre-motor phase;

It is made available under a [CC-BY-NC-ND 4.0 International license](#).

Abbreviations: a-Syn = alpha-Synuclein; BCHE = butyrylcholinesterase; CSF = cerebrospinal fluid; ICAM1 = intercellular adhesion molecule-1; IGFBP = Insulin-like-growth-factor-binding-protein; iPD = idiopathic Parkinson's disease; iRBD = idiopathic Rapid Eye Movement sleep behaviour disorder; ITIH2 = Inter-alpha-trypsin inhibitor heavy chain H2; LEDD = levodopa equivalent daily dose; LRRK2 (S) = symptomatic LRRK2 mutation carrier; LRRK2 (AS) = asymptomatic LRRK2 mutation carrier; MTPT = 1-Methyl-4-phenyl-1,2,3,6-tetrahydropyridin; NCAM1 = Neural cell adhesion molecule 1; NfL = neurofilament light chain; OPLS-DA = orthogonal projection to latent structures discriminant analysis; PCSK9 = Proprotein-Convertase-Subtilisin/Kexin Typ 9; PTGDS = prostaglandin D₂ synthase; ROC = receiver operating characteristics; SPP2 = secreted phosphoprotein 2; UBC = ubiquitin C; YOPD = young onset Parkinson's disease;

Introduction

Parkinson's disease is a progressive neurological disorder that affects movement and can cause a wide range of symptoms. Accurate diagnosis allows healthcare professionals to develop an appropriate treatment plan tailored to the individual's specific needs. Early detection and intervention can help manage symptoms more effectively and as we develop new therapies and interventions, the ability to detect individuals at-risk of neurodegeneration in very early disease stage will be invaluable in future clinical trials and outcomes. Idiopathic REM-sleep behaviour disorder (iRBD) is a characteristic parasomnia manifesting by dream-enacting behaviour, unpleasant dreams and loss of REM-sleep muscle atonia¹. Beyond its immediate relevance as a sleep disorder, iRBD importantly has been established as the most relevant non-motor prodromal marker for Parkinson's disease, the second most frequent cause of neurodegeneration², and is hence of particular interest to the Parkinson's disease community.

Parkinson's disease is currently the fastest growing neurodegenerative disorder in man^{4,5}, its prevalence and incidence predominantly increasing in an age-dependent manner³⁻⁶. With an ageing population across the globe, the significant and increasing socio-economic burden⁷ is projected to become an even more pressing burden to individual and public health⁸. Although the clinical syndrome of Parkinson's disease is meanwhile understood to encompass a heterogeneous multitude of aetiologies^{9,10}, the common underlying pathology is defined by a progressive accumulation of alpha-Synuclein (a-Syn)^{11,12}, possibly starting in the enteric or peripheral nervous system¹³⁻¹⁶. Classical Parkinson's disease is hence preceded by an insidious prodromal phase with a sub-threshold degree of neurodegeneration¹⁷ without overt motor symptoms¹⁸.

It is made available under a [CC-BY-NC-ND 4.0 International license](#).

Although the identification and study of genetic risk variants has dramatically improved our understanding of the underlying pathophysiology and genetic architecture of this condition^{10,19}, so far this failed to translate into earlier or more accurate diagnosis²⁰ outside highly dedicated movement disorder centres²¹. The search for a Parkinson's disease biomarker has been ongoing for some time^{22,23}, and while promising candidates such as neurofilament light chain (NfL) have proven non-specific²⁴, α -Syn oligomers detection via seed amplification assays appear promising both in CSF^{25,26}, as well as blood²⁷, but depend on specialised lab expertise²⁴.

Multiple clinical trials have failed²⁸ and continue to fail to establish a disease-modifying therapy despite considerable efforts²⁹⁻³¹. Along several appreciated issues, pertaining from the ideal trial design, patient and outcome measure selection, to the unequivocal proof of neuroprotective efficacy³², the core problem of detecting the condition in the pre-motor phase - before the majority of damage is done – remains^{33,34}. The identification of a biomarker of pre-symptomatic Parkinson's disease would greatly facilitate these efforts.

Meanwhile, iRBD has been recognized as the most specific single clinical prodromal marker of Parkinson's disease to date³⁵⁻³⁷. The risk of pheno-conversion to an alpha-synucleinopathy increases with the time since iRBD diagnosis, ranging from 15-35% within 5 years³⁸, to 81% within 25 years³⁹. Hence, iRBD patient cohorts have increasingly been studied for Parkinson's disease conversion risk and are generally considered an ideal target for future disease-modifying therapy trials^{33,34,40,41}.

The urine proteome is an interesting source for non-invasive biomarker development, as it is readily available in large quantities, the majority of its' content independent of diurnal variability and proteolytic degradation, and accessible to mass spectrometer measurement⁴². First studies have capitalised on these characteristics for biomarker detection in patients with cognitive decline^{43,44}, neural ceroid lipofuscinosis⁴⁵, kidney disease⁴⁶, cancer⁴⁷, dermatomyositis⁴⁸ but also Parkinson's disease^{49,50}.

In this work, we utilised the power of machine learning to create a panel of biomarkers into a scalable, multiplexed, mass spectrometry-based urine proteomics assay that can differentiate iRBD patients from healthy controls. The assay was based on the combination of eight protein biomarkers, identified by an untargeted discovery phase and candidates identified in the literature, and applied as a targeted validation amongst cohorts of iRBD, idiopathic Parkinson's disease, asymptomatic and symptomatic LRRK2 carriers, Parkin carriers and healthy controls. We demonstrate that this panel allows the identification of iRBD and has the potential to predict clinical conversion to Parkinson's disease.

Materials and methods

Study Cohorts

Study participants were recruited from neurology outpatient clinics at the UCL Institute of Neurology, London, and Safarik University and University Hospital, Kosice. In London, recruitment was performed at a specialist movement disorder outpatient clinic and consisted of consecutive patients with clinically established/probable Parkinson's disease according to current diagnostic criteria⁵¹ and established gene carrier status for LRRK2 and PRKN. Young-onset Parkinson's disease (YOPD) was defined by motor symptom onset before 45 years of age⁵². Age- and sex-matched healthy controls without history, signs or symptoms of parkinsonism or other neurodegenerative disease were recruited amongst patient partners. Asymptomatic LRRK2 carriers were recruited from first degree blood relatives of patients who are attending the movement disorders clinic at the National Hospital for Neurology and Neurosurgery, Queen Square, UK and known to be carriers for the LRRK2 G2019S mutation. In Kosice, participants were identified from a prospective data registry and a multistage screening for iRBD (PDBIOM study), which was diagnosed based on one or two nights of video-polysomnography (PSG)^{53,54}. In addition to extensive demographic and clinical details (as outlined below), dopamine transporter scintigraphy (DaT scan), as well as calculated probability of prodromal Parkinson's disease (pPD) based on the updated MDS criteria⁵⁵ as reported previously⁵³ were recorded and assessed. Healthy controls were prospectively recruited amongst deeply phenotyped PDBIOM subjects with negative PSG and normal values of pPD probability based on the updated MDS criteria⁵⁵.

All urine samples from participating study sites were collected according to identical sample processing protocols. Mid-stream urine convenience samples were collected during routine out-patient appointments in high density polyethylene (UCL: R53102 Rocket URIKONE specimen pack; Kosice: 455007 VACUETTE TUBES) containers irrespective of time of day, fasting or medication status. They were aliquoted and frozen at -80°C within 2 hours of collection without prior centrifugation.

Clinical assessments included patient demographics and motor scores (MDS-UPDRS part III) for all participants, and REM sleep behaviour screening questionnaire (RBD-SQ), Sniffing-sticks 16 test (SS16), Epworth sleepiness scale (ESS), Montreal Cognitive assessment (MoCA), Beck depression inventory II (BDI II), Non-motor symptoms questionnaire (NMSS) and Parkinson anxiety scale (PAS) for iRBD patients.

The study was approved by the local institutional ethics committees of participating centres in accordance with the Declaration of Helsinki (UCL: 10/H0721/87. IRAS project ID: 62859; Kosice: 2017/EK/11024, 2015/EK/900016). All participants provided written informed consent prior to study participation.

It is made available under a [CC-BY-NC-ND 4.0 International license](#).

The skew of males to females was significant in the patient cohorts (3:1 male to female ratio). Furthermore, the detection of a possible influence of sex on urine proteomics (Supplementary Figure 1) combined with insufficient numbers of female samples to confirm and perform a statistical correction for this phenomenon, all analyses were therefore concentrated on the analyses of male only samples ($n = 114$; Table 1). Healthy control subjects were recruited at both study centres and age-matched to idiopathic Parkinson's disease patients recruited in London ($P = 0.29$) as well as iRBD patients recruited in Kosice ($P = 0.69$). Additional patients recruited in London with YOPD ($n = 15$; $P < 0.0001$), symptomatic and asymptomatic LRRK2 carriers ($n = 10$; $P = 0.008$), as well as Parkin carriers ($n = 5$; $P = 0.001$) were younger than respective controls, as well as idiopathic Parkinson's disease patients ($P < 0.0001$). Parkin carriers had longer motor disease duration than YOPD ($P = 0.0007$), LRRK2 carriers ($P = 0.001$) and idiopathic patients ($P < 0.0001$), while motor severity as assessed by the MDS-UPDRS III, and Levodopa-equivalent daily dose (LEDD)⁵⁶ did not differ between groups.

Discovery Phase Cohort

The discovery cohort consisted of 9 random and well characterised patients (44% female) with idiopathic Parkinson's disease and had been confirmed by clinical assessment and positive DaTScan (mean age 68.3 ± 7.9 years; mean motor disease duration 11.7 ± 4.1 years, MDS-UPDRS III ON medication 30.7 ± 15.1), and 10 healthy controls, matched for age (69.8 ± 10.7 years; $P = 0.73$) and sex (50% female), recruited in London.

Targeted Proteomics Cohort

Targeted proteomic analyses were performed on samples independent of the discovery phase cohort. To detect and compare the proteomic profiles between iRBD, sporadic and two of the most frequent monogenic forms of Parkinson's disease (LRRK2 and Parkin), as well as controls, urine samples of $n = 184$ patients from two independent cross-sectional cohorts were included in the analyses (Supplemental Table 1).

Discovery urine proteomics

Sample preparation

Random urine samples were thawed at room temperature, followed by vortex for five seconds. 4 mL of urine was aliquoted into 5 mL centrifuge tubes (Sarstedt, 72.701.500). The 4 mL aliquots were centrifuged at room temperature at $3760 \times g$ for 30 minutes to separate the urinary sediment from solution using a Sorvall Legend RT centrifuge. 2 mL of the supernatant was transferred to Amicon Ultra-4 10 kDa molecular weight

It is made available under a [CC-BY-NC-ND 4.0 International license](#).

cut-off filters (Merck Millipore, UFC8010) and 2 mL Milli-Q water was added to give a final volume of 4 mL. To concentrate the urinary proteins, the samples were centrifuged at room temperature for one hour at 4445 x g (Sorvall Legend RT). The concentrate was transferred to a 1.5 mL centrifuge tube (Eppendorf). To ensure maximum recovery, the filters were washed with 100 µL 50 mM ammonium bicarbonate which was pooled with the concentrate. 800 µL ice-cold acetone was added to the pooled concentrate and the samples were vortexed for five seconds before overnight incubation in -20 °C. To separate the supernatant from the protein pellet, the samples were centrifuged for ten minutes at +4 °C and 16900 x g using a Micro-centrifuge 5424 R (Eppendorf). The supernatant was carefully pipetted off and discarded. The pellet was air dried in a fume hood for 20 minutes to evaporate residual acetone. 100 µL Milli-Q water was added to the samples and the protein pellet broken up by vigorous vortexing. The samples were thereafter freeze dried overnight, followed by tryptic digestion and solid phase extraction (SPE) as described by Bennett et al ⁵⁷. The digested and SPE-cleaned samples were reconstituted in 50 µL 3% acetonitrile, 0.1% TFA and a peptide assay (Pierce Quantitative Colorimetric Peptide Assay from Thermo Fisher Scientific) was performed as per the manufacturer's instructions. The peptide concentration in the samples was normalised to 1000 ng/µL before instrumental analysis by 2D-LC-MS.

Proteomics and biomarker discovery analyses of urine from control and Parkinson's Disease patients.

Tryptic peptides were fractionated using a 2D-NanoAquity liquid chromatography system (Waters, Manchester, UK). All samples were fractionated in-line into ten fractions and analysed by label free quantitative proteomics over a 12-hour period as previously described ⁵⁸. Fractionation was performed by reversed phase chromatography, utilising the orthogonality in peptide separation between high and low pH mobile phases. The eluted peptides were detected using a Synapt-G2-Si (Waters) equipped with a nano-electrospray ion source. The mass spectrometer was a Synapt-G2-Si (Waters). Data were acquired in positive MS^E mode from 0 to 60 minutes within the m/z range 50 - 2000. The capillary voltage was set to 3 kV and the source temperature to + 100 °C. The cone gas consisted of nitrogen with a flow of 50 L/h, the desolvation temperature was set to + 200 °C. The purge and desolvation gas consisted of nitrogen, operated at a flow rate of 600 mL/h and 600 L/h respectively. The gas in the IMS cell was helium with a flow rate of 90 mL/h. The low energy acquisition was performed applying a constant collision voltage of 4 V with a 1 second scan time. High energy acquisition was performed by applying a collision energy ramp, from 15 to 40 V, the scan time was 1 second. The lock mass consisted of 500 fmol/µL [glu1]-fibrinopeptide B, continuously infused at a flow rate 0.3 µL/min and acquired every 30 seconds. The doubly charged precursor ion, m/z 785.8426, was utilised for mass correction.

Untargeted data processing

After acquisition, data were imported to Progenesis QI for proteomics (Waters) and the individual fractions processed before all results were merged into one experiment. The Ion Accounting workflow was utilised, with UniProt Canonical Human Proteome as a database (build 2016). The digestion enzyme was set as trypsin. Carbamidomethyl on cysteines was set as a fixed modification; deamidation of glutamine and asparagine, and oxidation of tryptophan and pyrrolidone carboxylic acid on the N-terminus were set as variable modifications. The identification tolerance was restricted to at least two fragments per peptide, three fragments per protein, and one peptide per protein. An FDR of 4% or less was accepted. The resulting identifications and intensities were exported and variables with a confidence score less than 15 and only one unique peptide were filtered out.

Development of a multiplexed and targeted proteomic assay to validate any potential biomarkers as a test for predicting Parkinson's Disease

The peptides included in the targeted assay were selected from the proteomic screening study as in previous work⁵⁹. Furthermore, due to the suggested involvement of inflammation in neurodegenerative diseases, several known pro- and anti-inflammatory proteins identified from the literature were also included in the multiplexed assay (Supplementary Figure 2). This validity test in the first instance consisted of 87 proteins, out of which a number were measured with two peptides, leading to a total of 131 unique peptides. When possible, the peptides were chosen to have an amino acid sequence length between 7 and 20. The amino acid sequences were confirmed to be unique to the proteins by using the Basic Local Alignment Search Tool (BLAST) provided by UniProt⁶⁰. Synthetic peptide standards were purchased from GenScript (Amsterdam, Netherlands). To establish the most optimal transitions, repeated injections of 1 pmol peptide standard onto a Waters Acquity ultra-performance liquid chromatography (UPLC) system coupled to a Waters Xevo-TQ-S triple quadrupole MS were performed. The most high-abundant precursor-to-product ion transitions and their optimal collision energies were determined manually or using Skyline⁶¹. Two transitions were chosen for each peptide, one quantifier for relative concentration determination and one qualifier for identification, totally rendering 266 analyte transitions. Cone and collision energies varied depending on the optimal settings for each peptide. Each peptide was measured with a minimum of 12 points per peak and a dwell time of 10 milliseconds or more to ensure adequate data acquisition. The optimised transitions were distributed over two multiple reaction monitoring (MRM) methods, always keeping the quantifier and qualifier for each peptide in the same MRM segment.

The samples were prepared for instrumental analysis in the same way as the untargeted sample set, except for 150 ng whole protein yeast ENO1 being added prior to filtration, and that the peptide concentrations were not normalised prior to instrumental analysis. The samples were reconstituted in 50 µL 3% acetonitrile,

It is made available under a [CC-BY-NC-ND 4.0 International license](#).

0.1% trifluoroacetic acid, containing 0.1 μM of isotope labelled internal standards from the following proteins (annotated by gene name): ALDOA, C3, GSTO1, RSU1, and TSP1.

Targeted data processing

After acquisition, peak picking and integration were performed using TargetLynx (version 4.1, Waters) or an in-house application (“mrmIntegrate”) written in Python (version 3.8). mrmIntegrate is publicly available to download via the GitHub repository <https://github.com/jchallqvist/mrmIntegrate>. The application takes text files as input (.raw files are transformed into text files through the application “MSConvert” from ProteoWizard⁶² and applies a LOWESS filter over 5 points of the chromatogram. The integration method to produce areas under the curve is trapezoidal integration. The application allows for retention time alignment and simultaneous integration of the same transition for all samples. Peptide peaks were identified by the blank and matrix calibration curves. The integrated peak areas were exported to Microsoft Excel where first the ratio between quantifier and qualifier peak areas were evaluated to ensure that the correct peaks had been integrated. The digestion efficiency was evaluated by monitoring the presence of baker’s yeast ENO1 in the samples, all samples without a signal were excluded from further analysis. The instrumental stability was assessed by inspecting the levels of the isotope labelled internal standard peptides. After the initial quality assessment, the urine concentrations in the samples were harmonised by probabilistic quotient normalisation (PQN)⁶³. Pooled urine quality control samples were additionally evaluated to assess the robustness of the instrumental analysis.

Bioinformatic analysis

Sample quality control

The untargeted and targeted datasets were inspected for outliers and instrumental drift using principal component analysis (PCA) and orthogonal projection to latent variables (OPLS) in SIMCA, version 17 (Umetrics Sartorius Stedim, Umeå, Sweden). After normalization of urine concentrations using probabilistic quotient normalization (PQN), outliers exceeding six median deviations from each variable’s median were excluded. To equalise small instrumental differences, samples within each run batch were z-scored before merging the datasets. Each peptide was thereafter filtered or averaged so that each protein was represented by a single variable. The data were evaluated for normal distribution using D’Agostino and Pearson’s method from SciPy (version 1.11.0). The dataset was inspected for outliers and instrumental drift using principal component analysis (PCA) and orthogonal projection to latent variables (OPLS) in SIMCA, version 17 (Umetrics Sartorius Stedim, Umeå, Sweden).

Statistical analysis, visualisation and modelling

Most of the statistical analyses were performed in Python (version 3.8.16). The untargeted dataset was analysed for differences between the groups by applying Student's two-tailed t-test. Due to the limited sample numbers, the significance threshold was set nominally to $P < 0.05$. The targeted data were evaluated for normal distribution using D'Angostino and Pearson's method from SciPy. Significance testing between the independent groups was performed by Student's two-tailed t-test for normally distributed variables and by Mann-Whitney's non-parametric U-test, both from SciPy's stats package, for non-normally distributed variables. The Benjamini-Hochberg multiple testing correction false discovery rate (FDR) procedure (Statsmodels version 0.14.0) was applied with alpha set to 0.1. Fold-changes were calculated by dividing the means or medians of the affected groups by the control group. Correlation analyses in the targeted data were performed by Spearman's correlation (SciPy). Box plots and hierarchical clusters were produced using the Seaborn (version 0.12.2) and Matplotlib (version 3.7.0) libraries. All multivariate analyses were performed in SIMCA version 17 (Sartorius Umetrics). OPLS and OPLS-discriminant analysis (OPLS-DA) models were evaluated for significance by ANOVA p-values and by permutation tests applying 1000 permutations, where $p < 0.05$ and $p < 0.001$ were deemed significant, respectively. Data were analysed for pathway enrichment and Gene Ontology (GO) annotations using DAVID Bioinformatics Resources (2021 build). Networks were built in Cytoscape (version 3.8.0) applying the "Organic layout" from yFiles. Logistic regression (scikit-learn, version 1.2.2,⁶⁴) was utilised for multiple-regression classifier modelling. The data was split into two 50% parts, one for model training and one for prediction, using the function train-test-split from Scikit-learn. Cross-validation was performed using a stratified k-fold with five splits and the cross validated scores for the model were calculated as mean and standard deviation. ROC curves were constructed from each predictor and from the combined predictors. Plots of the data were constructed using the Matplotlib package (version 3.6.0)⁶⁵.

Data availability

Chromatograms and raw data files from the targeted proteomic analysis are available via the public Panorama repository https://panoramaweb.org/urine_irbd_trgt_proteomics.url. Further data are available from the corresponding author by reasonable request.

Results

Proteomic discovery phase and development of a targeted assay

Label free quantitative proteomics was performed using two-dimensional in-line nano-liquid chromatography in ten fractions using QTOF MS^E (Waters, UK). This deep phenotyping proteomic analysis

It is made available under a [CC-BY-NC-ND 4.0 International license](#).

of the nine idiopathic PD (iPD) and 10 Control urine samples identified > 2500 proteins, of which 89 were differentially expressed between groups on a significance level of 95% (Figure 1A). Pathway analysis of the differentially expressed proteins demonstrated that several inflammatory pathways were enriched, along with pathways relating to mitochondrial dysfunction and Parkinson's disease (Figure 1B). Based on the discovery phase findings, we constructed a targeted mass spectrometric assay in which the included proteins were selected by a mixture of greatest fold-change, p-value significance and top candidates discovered by other research groups identified via literature studies⁵⁹. This resulted in a panel of 87 proteins (Figure 1C).

Targeted proteomics

A targeted and scheduled proteomic assay was developed using ultra-high performance liquid chromatography (UPLC) system coupled to a triple quadrupole mass spectrometer operating in multiple reaction monitoring mode (MRM). The targeted assay was applied to a new cohort of urine samples consisting of controls, LRRK2 mutation carriers, YOPD, Parkin mutation carriers, idiopathic PD and iRBD. In total, we could reliably detect and quantitate 50 proteins in each of the urine samples. Quality control samples, consisting of pooled control and patient samples, demonstrated a coefficient of variation ranging from 3% to 28%, with 94% of the measured proteins attributed with a coefficient of variation of 15% or less (Supplementary Fig. 3).

Control sample compatibility and influence of age

To determine potential inter-centre variation, age-matched healthy control subjects from the two independent participating centres were compared (London: $n = 49$; Kosice: $n = 17$, 64.4 ± 10.6 years vs. 65.5 ± 9.1 years; $P = 0.7$). We compared the multivariate expression in the control group by orthogonal projection to latent structures discriminant analysis (OPLS-DA) with centre as the discriminant variable. The model was non-significant and demonstrated that no discernible multivariate expression difference could be identified between the centres ($P = 1.0$). In addition, univariate analysis applying Student's two-tailed t-test, followed by the Benjamini-Hochberg procedure for multiple testing correction (FDR = 10%), showed that none of the proteins demonstrated a significant difference between the two centres. Taken together, the inability to discriminate between the control samples both multi- and univariately demonstrates excellent compatibility between the centres. We next explored if there was an effect of age on protein expression through an OPLS model, where age was set as y. This model was non-significant with ANOVA-CV and permutations $P > 0.05$, also demonstrated that no age-related covariation between the variables could be found.

Proteomic identification of idiopathic REM-sleep behaviour disorder

Principle Component Analyses (PCA) was used for an initial overview of the targeted proteomic data and demonstrated that no patient sample clusters were readily identifiable (Figure 2). Investigating the discrimination between iRBD samples and healthy controls, an OPLS-DA model of these two sample groups was constructed. After visual inspection of the cross validated (CV) scores and permutations evaluation ($P < 0.001$), the model was found significant (ANOVA-CV $P = 0.002$). The proteins with strongest influence on the discrimination between the groups were UBC, VCAM1, MIEN1, GOLM1 and ICAM1 – upregulated in iRBD, and ITIH2, APOE, BCHE, CCL4 and PLD3 – downregulated in iRBD (Figure 3B). We further utilised the discriminating OPLS-DA model of iRBD *versus* control to evaluate the level of shared protein expression, compared to control, with the other Parkinsonian sample groups by prediction testing. This resulted in rather low proportions of the groups predicted as iRBD (57% for iPD; 60% for YOPD and LRRK2; 80% for Parkin), thus implying a limited degree of shared expression between the proteins typical for iRBD and Parkinsonian disorders in this sample set.

We similarly constructed an OPLS-DA model comparing individuals with iPD *versus* healthy controls, which proved no significant differences ($P = 0.099$). However, comparing iPD + iRBD as one group *versus* healthy controls showed weakly significant differences ($P = 0.01$), whereas an OPLS-DA comparison between iPD *versus* iRBD could not differentiate between the two groups ($P = 1$). This indicates that although the urine proteomic expression profile was not distinct enough to differentiate iPD individually from healthy controls, there is a relevant overlap in urine protein expression between iPD and iRBD. Assessing the shared similarities between these groups and the differences to healthy controls, a shared-and-unique-structures plot (Figure 3C) of the models (i) iRBD *versus* control, and (ii) iPD *versus* control, was created. This showed that despite differences between iPD and iRBD, they indeed shared similarities in comparison to healthy controls – with, among others, UBC, NCAM1, MIEN1, SPP2 and REG1B being positively correlated with the disease groups, and ITIH2, BCHE and C3 negatively correlated.

Finally, logistic regression was used to build a multiple-regression classifier model to differentiate between iRBD and healthy control samples. To achieve this, we utilised 50% of the samples for model training and predicted the remaining samples. This resulted in a classification accuracy of 80% using ten of the most significant variables from the comparison between iRBD and controls (UBC, BCHE, ITIH2, NCAM1, REG1B, PKM, ANXA2, SPP2, MIEN1, and ICAM1). Cross-validation of the model using five different splits of the data resulted in a mean accuracy of $78\% \pm 7.4\%$. A receiver operating characteristics (ROC) curve was then constructed and demonstrated that the combined panel of proteomic predictors resulted in greater classification accuracy than each predictor alone (Figure 4). To assess the similarities between the comparators in this iRBD model and Parkinsonian samples, we evaluated how the clinically diagnosed Parkinsonian groups were predicted in the logistic regression model. Prediction as belonging to the iRBD

It is made available under a [CC-BY-NC-ND 4.0 International license](#).

class was again generally low in all groups (YOPD 20%; LRRK2 30%; Parkin 20%, iPD 23%). Similar to before, this suggests that there is limited shared protein expression between these groups.

Univariate proteomic differences and clustering

The analyses of univariate protein expression differences between groups demonstrated that the iRBD group, representing prodromal PD, showed the largest differences compared to healthy controls. In total, 14 proteins were differentially expressed in the iRBD group, seven in iPD, four in LRRK2, three in Parkin and six in YOPD (Supplementary Figure 4). When assessed together, seven out of 50 measured proteins were differentially expressed in at least two of the sample groups when compared to healthy controls: ANXA2, BCHE, ITIH2, PKM, REG1B, SPP2 and UBC. All seven proteins were identified in the iRBD group, and all showed a reduced expression compared to controls (Supplementary Figure 5).

In order to explore similarities and differences on a proteomic level, a hierarchical cluster analysis was performed using the signed p-values resulting from the univariate comparisons of the disease groups (Figure 5). The resulting clustering identified two top-level clusters which grouped the proteins by direction of expression change (higher versus lower). Consistent with the above OPLS-DA analyses, the largest similarities in the overall proteomic expression profile were found between iRBD and iPD sample groups. According to this, the proteomic profile in Parkin carriers was distinct, with YOPD and LRRK2 samples clustering in between.

Clinical correlates of proteomic changes

To aid the interpretation of whether changes in the proteome are clinically meaningful, we explored any potential correlation with clinical metrics in the respective patient groups. In the Parkinsonian patient groups (iPD, YOPD and Parkin), correlations were evaluated between demographics, clinical metrics, and protein expression (Table). Unsurprisingly, motor disease severity (MDS-UPDRS III) correlated positively with age and motor disease duration, as well as levodopa equivalent daily dose (LEDD). Similar positive correlations were established for age with LEDD, Motor disease duration with MDS-UPDRS III & LEDD, and MDS-UPDRS III score with LEDD.

Relating proteomic changes with clinical metrics (Table 2), signals with at least two indicators of PD severity were identified for VCAM1, MSN and HPX. VCAM1 was found to positively correlate with MDS-UPDRS III and motor disease duration, MSN negatively correlated with motor disease duration and LEDD, HPX negatively correlated with MDS-UPDRS III and LEDD. Furthermore, MDS-UPDRS III negatively correlated with ICAM1. Motor disease duration was found positively correlated with ITIH2, and negatively with SERPING1, NCAM1, and TXN. Age was positively correlated with APOA4, BCHE and IGHA2, and

It is made available under a [CC-BY-NC-ND 4.0 International license](#).

negatively correlated with MASP2 and REG1B. Several of the identified proteins in the Parkinsonian samples were found to correlate with each other (Supplementary Figure 6).

Among iRBD patients, we in particular evaluated clinical metrics representing the risk for future conversion to PD (Figure 6A). In this regard, we found that dopamine transporter (DaT) scan results negatively correlated with DKK3, TNF β , MYO10, HSPA1L, MSN, and positively correlated with VCAM1. The total likelihood ratios (LR total), as calculated based on the updated MDS research criteria for pPD, were negatively correlated with MSN, C3, SPP2, MYO10, HSPA1L and MSN, and positively correlated with PLA2, CST3, HPX and VCAM1. The calculated probability to develop PD (pPD; %) was negatively correlated with SPP2, GOLM1, MYO10, HSPA1L, C3, UBC, PGAM1, and PPP3CB, and positively correlated with IGHA2, CYCS, APOA4, PLA2, CST3, HPX and VCAM1. Across these analyses, four proteins were found to be overlapping in their correlation with risk for future conversion to PD, namely VCAM1 (positive) as well as MYO10, HSPA1L and MSN (negative; Figure 6B).

Discussion

Development of proteomic panel for iRBD identification

In this work we present a panel of urine proteomic markers that is able to detect iRBD but also shows overlap with similar proteomic changes in idiopathic and hereditary forms of PD. This demonstrates the power of using a combination or panels of biomarkers with machine learning. By analysing multiple biomarkers simultaneously, machine learning models can uncover intricate patterns that may not be evident to human observers. These models can learn the associations between biomarker profiles and disease states, allowing for more accurate diagnosis and prognosis. Our two-step biomarker panel development approach, capitalizes on the analytical power of untargeted proteomics followed by a validity assay performed on triple quadrupole mass spectrometers. This strategy has the added advantage of potential for the rapid translation of these biomarkers into clinical laboratories, as most large centres have this mass spectrometry platform available.

Based on the full proteomic profiling of a small exploratory PD cohort in combination with protein targets of inflammation and neurodegeneration identified from the literature, the unbiased urine proteomic panel developed in this work overcomes the limited scope in targeted biomarker approaches based on a-priori assumptions^{22,23,25,26}. Our exploratory untargeted analysis resulted in a much deeper proteome coverage (>2500 proteins), quantifying roughly twice as many peptides as previous studies^{49,66}. Our two-step approach follows previous own work in the field of neurometabolic diseases⁴⁵, and Parkinson's disease, where we recently developed a multiplexed, mass spectrometry-based plasma proteomics test essay to differentiate de-novo Parkinson's disease patients from healthy controls with excellent classification

It is made available under a [CC-BY-NC-ND 4.0 International license](#).

accuracy (ref BioRxiv?). The current work expands this to urine proteomics and pre-symptomatic testing, and serves as proof-of-concept for a two-step approach for proteomic biomarker development.

In its ideal form, a biomarker is easily accessible, stable, correlates with disease severity and is scalable for a standardized adaptation across institutions. As a potential biomarker source, our laboratory has demonstrated that urine has several conceptual advantages, along ease of sampling⁴² for the use in targeted proteomic analyses and the stratification of disease progression⁶⁷. In particular, due to the high salt/urea concentration, the urine proteome is relatively stable in comparison to blood-based assays regarding circadian variability and influence of food intake⁶⁸. It is easy to ship requiring no precautions and is easily obtainable in large amounts and, in contrast to plasma and serum, urine requires no phlebotomy or sample preparation such as centrifugation.

Our results demonstrate a certain overlap in the urine proteomic changes in iRBD and Parkinson's disease with respect to healthy controls. This is in line with the well documented epidemiological association between the two clinical diagnoses, with over 80% of iRBD patients going to develop Parkinson's disease^{1,37,69,70}. In particular, we detected that between iPD and iRBD, UBC, NCAM1, MIEN1, SPP2, REG1B, TXN, ICAM1, PTGDS, FABP5 and SERPING1 were positively, and ITIH2, BCHE and C3 negatively correlated with disease. From the above, NCAM1, ITIH2, and BCHE were also found to correlate with clinical markers of disease severity in Parkinson's disease, strengthening a likely mechanistic link: NCAM1, involved in cell adhesion, axon development and neuron projection, has been previously found to be downregulated in plasma of drug-naïve Parkinson's disease patients⁷¹. ITIH2, a plasma serine protease inhibitor involved in extracellular matrix stabilization, has been found upregulated in the substantia nigra in a 1-Methyl-4-phenyl-1,2,3,6-tetrahydropyridin (MPTP) mouse model of PD⁷². BCHE is a known bioscavenger for pesticides - its serum activity has previously been shown to be a moderate biomarker for Parkinson's disease⁷³, while the K-variant BCHE polymorphism increases disease risk in pesticide-exposed individuals via reduced serum activity levels⁷⁴.

Overlapping with the above, in the iRBD samples we were able to detect a correlation between at least one measure for future conversion to Parkinson's disease with UBC, SPP2, and C3. None of these proteins have previously been implicated with iRBD pathophysiology, and for SPP2 the protein function is unknown so far. UBC is of major importance for protein homeostasis through its role in the ubiquitin-proteasome system, which is impeded early on in the Parkinson's disease process⁷⁵. C3 plays a central role in the activation of the microglia-complement-phagosome pathway, and has been reported to be elevated in the CSF⁷⁶, as well as reduced in plasma in Parkinson's disease⁷⁷.

This further strengthens a likely mechanistic link between these proteins and the underlying developing α-Syn pathology. Mere correlation between a clinical outcome and a biomarker does not constitute a valid surrogate⁷⁸. However, the fact that in addition to the above-described proteomic convergence between

It is made available under a [CC-BY-NC-ND 4.0 International license](#).

diseases we detected half of the above listed proteins (ICAM1, PTGDS, C3 and SERPING1) similarly in plasma of drug-naïve Parkinson's disease patients (ref Brit's paper), serves as external validation of the identified targets and their likely pathophysiological relevance. Importantly, these proteins have also been implicated in the Parkinson's disease process, and in particular related neuroinflammation – ICAM1 as a prominent marker of neuroinflammation, involved in sustained substantia nigra inflammation⁷⁹ and T-cell infiltration⁸⁰. ICAM1 has been found elevated in CSF in Parkinson's⁸¹, but also Alzheimer's disease⁸² and thus might not be specific to a-Syn pathology. Prostaglandin 2, the product of PTGDS, has been implied in the circadian system and regulation of the sleep/wake cycle^{83,84}, offering a pathophysiological link to sleep-related symptoms. Interestingly, PTGDS has also been identified to correlate with abnormal immune cell composition in Parkinson's disease⁸⁵ and regulate the anti-inflammatory function of astrocytes, a mechanism involved in DJ-1 related hereditary Parkinson's⁸⁶. SERPING1 has similarly been described to regulate the loss of dopaminergic neurons by regulating astrocyte expression levels in the substantia nigra of MPTP-treated mice⁸⁷.

Overall, the majority of marker proteins identified between iRBD and Parkinson's disease hence point at a major involvement of the immune system – well established and reviewed in⁸⁸ - signs of which have recently been also detected in iRBD patients' peripheral blood immune cell composition⁸⁹ and substantia nigra immune activation⁹⁰. The role of the immune system in iRBD, but also non-manifesting carriers of Parkinson's disease causing mutations is increasingly recognized⁹¹, and our findings are in line with this.

Our results establish urine proteomic differences between idiopathic, young-onset and different hereditary forms of Parkinson's disease, as previously described for LRRK2 in CSF samples⁶⁶. Interestingly, although LRRK2 and PARKIN pathophysiology is overlapping in mitochondrial dysfunction, the proteomic profile for PARKIN was distinct and clustered separately from all other entities. The limited number of included monogenic cases warrants cautious interpretation, and greater sample numbers are needed for definitive conclusions.

The ability of the test to detect individuals who are at risk of developing Parkinson's Disease

Research into biomarkers in the field of Parkinson's follows diagnostic (detecting established disease), monitoring (detecting change in severity) and prognostic (prediction of future disease development) aims. The 'shutting the stable door after the horse has bolted' hypothesis has gained much traction in neurodegenerative disease research. As it is appreciated that the majority of damage through neurodegeneration has already occurred once motor symptoms are present³³, and interventions hence less likely to salvage neuronal function, the pre-motor phase of the disease has become the focus for biomarker

It is made available under a [CC-BY-NC-ND 4.0 International license](#).

development^{22,37}. Along mutation carriers in established Parkinson's risk genes, such as LRRK2^{92,93}, iRBD has moved into the focus of biomarker development due to its high specificity for later conversion to alpha-synucleinopathies³⁷. There is significant consensus in the field that early and accurate detection of iRBD is likely the key to maximise the potential for disease modification³⁰.

Established diagnostic tools for iRBD so far involve questionnaires, which have been reported with accuracies below 70% when compared prospectively to polysomnography⁹⁴ and have been shown not to be valid in *de novo* Parkinson's disease patients⁹⁵, or expensive and not readily available, gold-standard video-polysomnography⁷⁰. Until recently, only clinical, autonomic, electrophysiological and imaging biomarkers had been identified to predict pheno-conversion from iRBD to Parkinson's disease^{37,69}. Meanwhile, CSF-based a-Syn seeding assays have been reported as one possibly accurate predictor^{25,26}. However, the invasive nature of the sampling procedure, labour-intense processing and necessary standardisation of the quantification technique is significant hindrance to a wide-spread application as a screening tool. Thus, there are currently no readily available screening tools to identify individuals at-risk of iRBD on a population level. In this context, urine proteomics appear as an ideal form of a non-invasive, easily collectable biomarker, which can be quantified in any accredited mass spectrometry lab. Sampling convenience, scalability and standardization would lend itself for the application in a population screening program.

Summary

This study demonstrates the power of using targeted proteomics to measure multiple biomarkers at once and when combined with machine learning, a panel of biomarkers can enhance both sensitivity and specificity of tests to identify neurodegenerative disorders. This work presents a fully developed and translational targeted urine proteomic panel for the detection of iRBD. Furthermore, the assay may be run in less than 5 min/sample and can be run in any accredited, clinical mass-spectrometry laboratory and when run at scale, could cost less than \$10 per sample. This panel might offer a way for future screening programs to identify individuals at high risk to develop Parkinson's disease in very early disease stages.

Acknowledgements

This section is not mandatory.

It is made available under a [CC-BY-NC-ND 4.0 International license](#).

Jenny Hällqvist,	orcid: 0000-0001-6709-3211
Sebastian R. Schreglmann,	orcid: 0000-0002-4129-5808
Kristina Kulcsarova,	orcid: 0000-0003-3135-1826
Matej Skorvanek,	orcid: 0000-0001-5497-8715
Eva Feketeova,	orcid: 0000-0002-7746-3797
Mie Rizig,	orcid: 0000-0003-2420-4449
Brit Mollenhauer	orcid: 0000-0001-8437-3645
Paola Turano	orcid: 0000-0002-7683-8614
Maria Giulia Bacalini	orcid: 0000-0003-1618-2673
Claudio Francheschi	orcid:
Nicholas Wood,	orcid: 0000-0002-9500-3348
Kailash P. Bhatia,	orcid: 0000-0001-8185-286X
Kevin Mills,	orcid: 0000-0003-0763-8288

Funding

The authors would like to thank The European Union for their sponsorship of this work through the Horizon 2020 Framework Programme (grant number 634821, PROPAG-AGEING), The Peto Foundation for their kind donations and the BRC at Great Ormond Street Hospital. SRS received support from the Advanced Clinician Scientist Program by the Interdisciplinary Centre for Clinical Research, Wuerzburg, Germany – he is a Fellow of the Thiemann Foundation. KK, MS and EF received funding from the Slovak Grant and Development Agency under contracts no. APVV-18-0547 and APV-22-0279, and by the Slovak Scientific Grant Agency under contract no. VEGA 1/0712/22.

Competing interests

The authors report no competing interests.

Supplementary material

Supplementary Table 1. Demographics of samples analysed for targeted proteomics. The number of samples, percentages of males and females, and age \pm standard deviation are reported for the sample groups control, iPD, iRBD, LRRK2 (S), LRRK2 (AS), YOPD and PARKIN

Group	n	M	F	% F/M	Age \pm SD
Controls (iPD and iRBD combined)	66	35	31	47/53	64.7 \pm 10.1
Control (iPD)	49	21	28	57.1/42.9	64.4 \pm 10.6
Control (iRBD)	17	14	3	17.6/82.4	65.5 \pm 9.1
iPD	59	35	24	40.7/59.3	67.1 \pm 8.4
iRBD	18	14	4	22.2/77.8	67.2 \pm 9.2
LRRK2 (S)	13	9	4	30.8/69.2	57.6 \pm 7.5
LRRK2 (AS)	3	1	2	66.7/33.3	51.5 \pm 26.2
YOPD	19	15	4	21.1/78.9	48.2 \pm 7.7
Parkin	6	5	1	16.7/83.3	49.2 \pm 5.0

iPD: idiopathic Parkinson's disease, iRBD: Rapid Eye Movement sleep disturbance disorder, LRRK2 (S): symptomatic LRRK2 mutation carrier, LRRK2 (AS): asymptomatic LRRK2 mutation carrier, YOPD: young onset Parkinson's disease, Parkin: mutations in the PRKN gene cause autosomal recessive juvenile Parkinson's disease, M: male, F: female, SD: standard deviation.

Supplementary Table 2. Results from the comparison between the parkinsonian groups iRBD, iPD, YOPD, PARKIN and LRRK2 versus control. The table shows p-value significances as ** $P < 0.01$ and * $P < 0.05$. The arrows following the asterisks represent if the protein was up- or downregulated in the disease group compared to control.

	iRBD	iPD	YOPD	PARKIN	LRRK2
ANXA2	** (↓)	* (↓)			* (↓)
APOE	* (↓)				
BCHE	** (↓)	* (↓)		** (↓)	
BST1	* (↓)				
ICAM1	* (↑)				
ITIH2	** (↓)	* (↓)			* (↓)
MIEN1	* (↑)				
MYO10	* (↑)				
NCAM1	** (↑)				
PKM	** (↓)		* (↓)		
REG1B	** (↑)	* (↑)	* (↑)		

It is made available under a [CC-BY-NC-ND 4.0 International license](#).

SPP2	** (↑)	* (↑)		
TFEB	** (↓)			
UBC	** (↑)	** (↑)	** (↑)	
FGA		* (↓)		
IGHA2			* (↓)	
TXN			** (↑)	
VCAM1			** (↓)	
HSPA5				** (↑)
PGAM1				* (↓)
GOLM1				* (↑)
COL6A3				* (↓)

References

1. Dauvilliers Y, Schenck CH, Postuma RB, et al. REM sleep behaviour disorder. *Nat Rev Dis Primers*. 2018;4(1):19. doi:10.1038/s41572-018-0016-5
2. Armstrong MJ, Okun MS. Diagnosis and Treatment of Parkinson Disease. *Jama*. 2020;323(6):548-560. doi:10.1001/jama.2019.22360
3. Pringsheim T, Jette N, Frolkis A, Steeves TDL. The prevalence of Parkinson's disease: A systematic review and meta-analysis. *Mov Disord*. 2014;29(13):1583-1590. doi:10.1002/mds.25945
4. Collaborators G 2016 N, Feigin VL, Nichols E, et al. Global, regional, and national burden of neurological disorders, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Neurology*. 2019;18(5):459-480. doi:10.1016/s1474-4422(18)30499-x
5. Dorsey ER, Sherer T, Okun MS, Bloem BR. The Emerging Evidence of the Parkinson Pandemic. *J Park Dis*. 2018;8(s1):S3-S8. doi:10.3233/jpd-181474
6. Savica R, Grossardt BR, Bower JH, Ahlskog JE, Rocca WA. Time Trends in the Incidence of Parkinson Disease. *JAMA Neurol*. Published online 2016. doi:10.1001/jamaneurol.2016.0947
7. Dorsey ER, Elbaz A, Nichols E, et al. Global, regional, and national burden of Parkinson's disease, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Neurology*. 2018;17(11):939-953. doi:10.1016/s1474-4422(18)30295-3
8. Yang W, Hamilton JL, Kopil C, et al. Current and projected future economic burden of Parkinson's disease in the U.S. *Npj Park Dis*. 2020;6(1):15. doi:10.1038/s41531-020-0117-1
9. Greenland JC, Williams □ Gray CH, Barker RA. The clinical heterogeneity of Parkinson's disease and its therapeutic implications. *Eur J Neurosci*. 2019;49(3):328-338. doi:10.1111/ejn.14094
10. Poewe W, Seppi K, Tanner CM, et al. Parkinson disease. *Nature reviews Disease primers*. 2017;3:17013. doi:10.1038/nrdp.2017.13

It is made available under a [CC-BY-NC-ND 4.0 International license](#).

11. Braak H, Ghebremedhin E, Rüb U, Bratzke H, Tredici KD. Stages in the development of Parkinson's disease-related pathology. *Cell and Tissue Research*. 2004;318(1):121-134. doi:10.1007/s00441-004-0956-9
12. McCann H, Cartwright H, Halliday GM. Neuropathology of α -synuclein propagation and braak hypothesis. *Mov Disord*. 2016;31(2):152-160. doi:10.1002/mds.26421
13. Holmqvist S, Chutna O, Bousset L, et al. Direct evidence of Parkinson pathology spread from the gastrointestinal tract to the brain in rats. *Acta Neuropathol*. 2014;128(6):805-820. doi:10.1007/s00401-014-1343-6
14. Klingelhöfer L, Reichmann H. Pathogenesis of Parkinson disease—the gut–brain axis and environmental factors. *Nat Rev Neurol*. 2015;11(11):625-636. doi:10.1038/nrneurol.2015.197
15. Doppler K, Jentschke HM, Schulmeyer L, et al. Dermal phospho- α -synuclein deposits confirm REM sleep behaviour disorder as prodromal Parkinson's disease. *Acta Neuropathol*. 2017;133(4):535-545. doi:10.1007/s00401-017-1684-z
16. Kuzkina A, Panzer C, Seger A, et al. Dermal Real-Time Quaking-Induced Conversion Is a Sensitive Marker to Confirm Isolated Rapid Eye Movement Sleep Behavior Disorder as an Early α -Synucleinopathy. *Mov Disord*. Published online 2023. doi:10.1002/mds.29340
17. Mahlknecht P, Seppi K, Poewe W. The Concept of Prodromal Parkinson's Disease. *Journal of Parkinson's Disease*. 2015;5(4):681-697. doi:10.3233/jpd-150685
18. Berg D, Postuma RB, Adler CH, et al. MDS research criteria for prodromal Parkinson's disease. *Mov Disord*. 2015;30(12):1600-1611. doi:10.1002/mds.26431
19. Nalls MA, Blauwendraat C, Vallerga CL, et al. Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. *The Lancet Neurology*. 2019;18(12):1091-1102. doi:10.1016/s1474-4422(19)30320-5
20. 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. doi:10.1212/wnl.0000000000002350
21. Virameteekul S, Revesz T, Jaunmuktane Z, Warner TT, Pablo-Fernández ED. Clinical Diagnostic Accuracy of Parkinson's Disease: Where Do We Stand? *Mov Disord*. 2023;38(4):558-566. doi:10.1002/mds.29317
22. He R, Yan X, Guo J, Xu Q, Tang B, Sun Q. Recent Advances in Biomarkers for Parkinson's Disease. *Front Aging Neurosci*. 2018;10:305. doi:10.3389/fnagi.2018.00305
23. Parnetti L, Gaetani L, Eusebi P, et al. CSF and blood biomarkers for Parkinson's disease. *Lancet Neurology*. 2019;18(6):573-586. doi:10.1016/s1474-4422(19)30024-9
24. Mollenhauer B, Dakna M, Kruse N, et al. Validation of Serum Neurofilament Light Chain as a Biomarker of Parkinson's Disease Progression. *Mov Disord*. 2020;35(11):1999-2008. doi:10.1002/mds.28206
25. Shahnawaz M, Tokuda T, Waragai M, et al. Development of a Biochemical Diagnosis of Parkinson Disease by Detection of α -Synuclein Misfolded Aggregates in Cerebrospinal Fluid. *Jama Neurol*. 2016;74(2):163-172. doi:10.1001/jamaneurol.2016.4547

It is made available under a [CC-BY-NC-ND 4.0 International license](#).

26. Concha Marambio L, Weber S, Farris CM, et al. Accurate Detection of α -Synuclein Seeds in Cerebrospinal Fluid from Isolated Rapid Eye Movement Sleep Behavior Disorder and Patients with Parkinson's Disease in the DeNovo Parkinson (DeNoPa) Cohort. *Mov Disord.* 2023;38(4):567-578. doi:10.1002/mds.29329
27. Kluge A, Bunk J, Schaeffer E, et al. Detection of neuron-derived pathological α -synuclein in blood. *Brain : a J Neurol.* 2021;145(9):3058-3071. doi:10.1093/brain/awac115
28. Olanow CW, Kieburtz K, Schapira AHV. Why have we failed to achieve neuroprotection in Parkinson's disease? *Ann Neurol.* 2008;64(S2):S101-S110. doi:10.1002/ana.21461
29. Whone A, Luz M, Boca M, et al. Randomized trial of intermittent intraputamenal glial cell line-derived neurotrophic factor in Parkinson's disease. *Brain.* 2019;142(3):512-525. doi:10.1093/brain/awz023
30. Mari Z, Mestre TA. The Disease Modification Conundrum in Parkinson's Disease: Failures and Hopes. *Front Aging Neurosci.* 2022;14:810860. doi:10.3389/fnagi.2022.810860
31. Pagano G, Taylor KI, Anzures-Cabrera J, et al. Trial of Prasinezumab in Early-Stage Parkinson's Disease. *New Engl J Med.* 2022;387(5):421-432. doi:10.1056/nejmoa2202867
32. Athauda D, Foltynie T. Challenges in detecting disease modification in Parkinson's disease clinical trials. *Parkinsonism Relat D.* 2016;32:1-11. doi:10.1016/j.parkreldis.2016.07.019
33. Mahlknecht P, Marini K, Werkmann M, Poewe W, Seppi K. Prodromal Parkinson's disease: hype or hope for disease-modification trials? *Transl Neurodegener.* 2022;11(1):11. doi:10.1186/s40035-022-00286-1
34. Crotty GF, Keavney JL, Alcalay RN, et al. Planning for Prevention of Parkinson Disease. *Neurology.* 2022;99(7_Supplement_1):1-9. doi:10.1212/wnl.0000000000200789
35. Tolosa E, Garrido A, Scholz SW, Poewe W. Challenges in the diagnosis of Parkinson's disease. *Lancet Neurology.* 2021;20(5):385-397. doi:10.1016/s1474-4422(21)00030-2
36. Hustad E, Aasly JO. Clinical and Imaging Markers of Prodromal Parkinson's Disease. *Front Neurol.* 2020;11:395. doi:10.3389/fneur.2020.00395
37. Miglis MG, Adler CH, Antelmi E, et al. Biomarkers of conversion to α -synucleinopathy in isolated rapid-eye-movement sleep behaviour disorder. *Lancet Neurology.* 2021;20(8):671-684. doi:10.1016/s1474-4422(21)00176-9
38. Iranzo A, Molinuevo JL, Santamaría J, et al. Rapid-eye-movement sleep behaviour disorder as an early marker for a neurodegenerative disorder: a descriptive study. *Lancet Neurology.* 2006;5(7):572-577. doi:10.1016/s1474-4422(06)70476-8
39. Schenck CH, Boeve BF, Mahowald MW. Delayed emergence of a parkinsonian disorder or dementia in 81% of older men initially diagnosed with idiopathic rapid eye movement sleep behavior disorder: a 16-year update on a previously reported series. *Sleep Med.* 2013;14(8):744-748. doi:10.1016/j.sleep.2012.10.009
40. Fereshtehnejad S, Montplaisir JY, Pelletier A, Gagnon J, Berg D, Postuma RB. Validation of the MDS research criteria for prodromal Parkinson's disease: Longitudinal assessment in a REM sleep behavior disorder (RBD) cohort. *Mov Disord.* 2017;32(6):865-873. doi:10.1002/mds.26989

It is made available under a [CC-BY-NC-ND 4.0 International license](#).

41. Videnovic A, Högl B. Rapid-Eye-Movement Sleep Behavior Disorder. Published online 2018:641-647. doi:10.1007/978-3-319-90152-7_44
42. Decramer S, Peredo AG de, Breuil B, et al. Urine in Clinical Proteomics. *Mol Cell Proteomics*. 2008;7(10):1850-1862. doi:10.1074/mcp.r800001-mcp200
43. Watanabe Y, Hirao Y, Kasuga K, Kitamura K, Nakamura K, Yamamoto T. Urinary proteome profiles associated with cognitive decline in community elderly residents—A pilot study. *Front Neurol*. 2023;14:1134976. doi:10.3389/fneur.2023.1134976
44. Kurbatova N, Garg M, Whiley L, et al. Urinary metabolic phenotyping for Alzheimer's disease. *Sci Rep-uk*. 2020;10(1):21745. doi:10.1038/s41598-020-78031-9
45. Iwan K, Clayton R, Mills P, et al. Urine proteomics analysis of patients with neuronal ceroid lipofuscinoses. *Iscience*. 2021;24(2):102020. doi:10.1016/j.isci.2020.102020
46. Fan G, Gong T, Lin Y, et al. Urine proteomics identifies biomarkers for diabetic kidney disease at different stages. *Clin Proteom*. 2021;18(1):32. doi:10.1186/s12014-021-09338-6
47. Fan H, Li X, Li ZW, et al. Urine proteomic signatures predicting the progression from premalignancy to malignant gastric cancer. *Ebiomedicine*. 2022;86:104340. doi:10.1016/j.ebiom.2022.104340
48. Morales M, Alayi TD, Tawalbeh SM, et al. Urine proteomics by mass spectrometry identifies proteins involved in key pathogenic pathways in patients with juvenile dermatomyositis. *Rheumatology*. Published online 2023. doi:10.1093/rheumatology/kead033
49. Winter SV, Karayel O, Strauss MT, et al. Urinary proteome profiling for stratifying patients with familial Parkinson's disease. *Embo Mol Med*. 2021;13(3):e13257. doi:10.15252/emmm.202013257
50. Wang S, Kojima K, Mobley JA, West AB. Proteomic analysis of urinary extracellular vesicles reveal biomarkers for neurologic disease. *Ebiomedicine*. 2019;45:351-361. doi:10.1016/j.ebiom.2019.06.021
51. Postuma RB, Berg D, Stern M, et al. MDS clinical diagnostic criteria for Parkinson's disease. *Mov Disord*. 2015;30(12):1591-1601. doi:10.1002/mds.26424
52. Gustavsson EK, Trinh J, McKenzie M, et al. Genetic Identification in Early Onset Parkinsonism among Norwegian Patients. *Mov Disord Clin Pract*. 2017;4(4):499-508. doi:10.1002/mdc3.12501
53. Kulcsarova K, Ventosa JR, Feketeova E, et al. Comparison in detection of prodromal Parkinson's disease patients using original and updated MDS research criteria in two independent cohorts. *Parkinsonism Relat D*. 2021;87:48-55. doi:10.1016/j.parkreldis.2021.04.028
54. Ventosa JR, Kulcsárová K, Mertová L, et al. Heart Rate Variability in evaluation of autonomic dysfunction in idiopathic REM-sleep behaviour disorder. *Neurol Neurochir Pol*. Published online 2023. doi:10.5603/pjnns.a2023.0021
55. Heinzel S, Berg D, Gasser T, et al. Update of the MDS research criteria for prodromal Parkinson's disease. *Mov Disord*. 2019;34(10):1464-1470. doi:10.1002/mds.27802
56. Tomlinson CL, Stowe R, Patel S, Rick C, Gray R, Clarke CE. Systematic review of levodopa dose equivalency reporting in Parkinson's disease. *Mov Disord*. 2010;25(15):2649-2653. doi:10.1002/mds.23429

It is made available under a [CC-BY-NC-ND 4.0 International license](#).

57. Bennett K, Callard R, Heywood W, et al. New Role for LEKTI in Skin Barrier Formation: Label-Free Quantitative Proteomic Identification of Caspase 14 as a Novel Target for the Protease Inhibitor LEKTI. *J Proteome Res.* 2010;9(8):4289-4294. doi:10.1021/pr1003467
58. Bliss E, Heywood WE, Benatti M, Sebire NJ, Mills K. An optimised method for the proteomic profiling of full thickness human skin. *Biol Proced Online.* 2016;18(1):15. doi:10.1186/s12575-016-0045-y
59. Doykov I, Hällqvist J, Gilmour KC, Grandjean L, Mills K, Heywood WE. 'The long tail of Covid-19' - The detection of a prolonged inflammatory response after a SARS-CoV-2 infection in asymptomatic and mildly affected patients. *F1000Research.* 2020;9:1349. doi:10.12688/f1000research.27287.2
60. Consortium TU, Bateman A, Martin MJ, et al. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.* 2020;49(D1):D480-D489. doi:10.1093/nar/gkaa1100
61. MacLean B, Tomazela DM, Shulman N, et al. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics.* 2010;26(7):966-968. doi:10.1093/bioinformatics/btq054
62. Kessner D, Chambers M, Burke R, Agus D, Mallick P. ProteoWizard: open source software for rapid proteomics tools development. *Bioinformatics.* 2008;24(21):2534-2536. doi:10.1093/bioinformatics/btn323
63. Dieterle F, Ross A, Schlotterbeck G, Senn H. Probabilistic Quotient Normalization as Robust Method to Account for Dilution of Complex Biological Mixtures. Application in 1H NMR Metabonomics. *Anal Chem.* 2006;78(13):4281-4290. doi:10.1021/ac051632c
64. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine Learning in Python. *JMLR.* 2011;(12):2825-2830. doi:10.48550/arxiv.1201.0490
65. Hunter JD. Matplotlib: A 2D Graphics Environment. *Comput Sci Eng.* 2007;9(3):90-95. doi:10.1109/mcse.2007.55
66. Karayel O, Winter SV, Padmanabhan S, et al. Proteome profiling of cerebrospinal fluid reveals biomarker candidates for Parkinson's disease. *Cell Rep Med.* 2022;3(6):100661. doi:10.1016/j.xcrm.2022.100661
67. Doykov ID, Heywood WE, Nikolaenko V, et al. Rapid, proteomic urine assay for monitoring progressive organ disease in Fabry disease. *J Méd Genet.* 2020;57(1):38. doi:10.1136/jmedgenet-2019-106030
68. Good DM, Thongboonkerd V, Novak J, et al. Body Fluid Proteomics for Biomarker Discovery: Lessons from the Past Hold the Key to Success in the Future. *J Proteome Res.* 2007;6(12):4549-4555. doi:10.1021/pr070529w
69. Natale ER de, Wilson H, Politis M. Predictors of RBD progression and conversion to synucleinopathies. *Curr Neurol Neurosci.* 2022;22(2):93-104. doi:10.1007/s11910-022-01171-0
70. Skorvanek M, Feketeova E, Kurtis MM, Ruz J, Sonka K. Accuracy of Rating Scales and Clinical Measures for Screening of Rapid Eye Movement Sleep Behavior Disorder and for Predicting Conversion to Parkinson's Disease and Other Synucleinopathies. *Front Neurol.* 2018;9:376. doi:10.3389/fneur.2018.00376
71. Abdi IY, Bartl M, Dakna M, et al. Cross-sectional proteomic expression in Parkinson's disease-related proteins in drug-naïve patients vs healthy controls with longitudinal clinical follow-up. *Neurobiol Dis.* 2023;177:105997. doi:10.1016/j.nbd.2023.105997

It is made available under a [CC-BY-NC-ND 4.0 International license](#).

72. Yeo S, An KS, Hong YM, et al. Neuroprotective changes in degeneration-related gene expression in the substantia nigra following acupuncture in an MPTP mouse model of Parkinsonism: Microarray analysis. *Genet Mol Biol.* 2015;38(1):115-127. doi:10.1590/s1415-475738120140137
73. Dong MX, Xu XM, Hu L, Liu Y, Huang YJ, Wei YD. Serum Butyrylcholinesterase Activity: A Biomarker for Parkinson's Disease and Related Dementia. *BioMed Res Int.* 2017;2017:1524107. doi:10.1155/2017/1524107
74. Rösler TW, Salama M, Shalash AS, et al. K-variant BCHE and pesticide exposure: Gene-environment interactions in a case-control study of Parkinson's disease in Egypt. *Sci Rep.* 2018;8(1):16525. doi:10.1038/s41598-018-35003-4
75. McKinnon C, Snoo MLD, Gondard E, et al. Early-onset impairment of the ubiquitin-proteasome system in dopaminergic neurons caused by α -synuclein. *Acta Neuropathol Commun.* 2020;8(1):17. doi:10.1186/s40478-020-0894-0
76. Abdi I, Ghanem SS, El-Agnaf OM. Immune-related biomarkers for Parkinson's disease. *Neurobiol Dis.* 2022;170:105771. doi:10.1016/j.nbd.2022.105771
77. Wu HC, Chang KH, Chiang MC, Chen CM. Alterations of Plasma Galectin-3 and C3 Levels in Patients with Parkinson's Disease. *Brain Sci.* 2021;11(11):1515. doi:10.3390/brainsci11111515
78. Califf RM. Biomarker definitions and their applications. *Exp Biol Med.* 2018;243(3):213-221. doi:10.1177/1535370217750088
79. Miklossy J, Doudet DD, Schwab C, Yu S, McGeer EG, McGeer PL. Role of ICAM-1 in persisting inflammation in Parkinson disease and MPTP monkeys. *Exp Neurol.* 2006;197(2):275-283. doi:10.1016/j.expneurol.2005.10.034
80. Li W, Chen S, Luo Y, et al. Interaction between ICAM1 in endothelial cells and LFA1 in T cells during the pathogenesis of experimental Parkinson's disease. *Exp Ther Med.* 2020;20(2):1021-1029. doi:10.3892/etm.2020.8758
81. Lerche S, Zimmermann M, Roeben B, et al. Inflammatory CSF profiles and longitudinal development of cognitive decline in sporadic and GBA-associated PD. *npj Park's Dis.* 2023;9(1):38. doi:10.1038/s41531-023-00476-2
82. Janelidze S, Mattsson N, Stomrud E, et al. CSF biomarkers of neuroinflammation and cerebrovascular dysfunction in early Alzheimer disease. *Neurology.* 2018;91(9):10.1212/WNL.0000000000006082. doi:10.1212/wnl.0000000000006082
83. Jiang F, Wu Q, Sun S, Bi G, Guo L. Identification of potential diagnostic biomarkers for Parkinson's disease. *FEBS Open Bio.* 2019;9(8):1460-1468. doi:10.1002/2211-5463.12687
84. Ahmad AS, Ottallah H, Maciel CB, Strickland M, Doré S. Role of the L-PGDS-PGD2-DP1 receptor axis in sleep regulation and neurologic outcomes. *Sleep.* 2019;42(6). doi:10.1093/sleep/zsz073
85. Huang Y, Liu H, Hu J, et al. Significant Difference of Immune Cell Fractions and Their Correlations With Differential Expression Genes in Parkinson's Disease. *Front Aging Neurosci.* 2021;13:686066. doi:10.3389/fnagi.2021.686066

It is made available under a [CC-BY-NC-ND 4.0 International license](#).

86. Choi DJ, An J, Jou I, Park SM, Joe EH. A Parkinson's disease gene, DJ-1, regulates anti-inflammatory roles of astrocytes through prostaglandin D2 synthase expression. *Neurobiol Dis*. 2019;127:482-491. doi:10.1016/j.nbd.2019.04.003
87. Seo MH, Yeo S. Association of increase in Serping1 level with dopaminergic cell reduction in an MPTP-induced Parkinson's disease mouse model. *Brain Res Bull*. 2020;162:67-72. doi:10.1016/j.brainresbull.2020.05.011
88. Bartl M, Xylaki M, Bähr M, Weber S, Trenkwalder C, Mollenhauer B. Evidence for immune system alterations in peripheral biological fluids in Parkinson's disease. *Neurobiol Dis*. 2022;170:105744. doi:10.1016/j.nbd.2022.105744
89. Zhang H, Wang T, Li Y, et al. Plasma immune markers in an idiopathic REM sleep behavior disorder cohort. *Park Relat Disord*. 2020;78:145-150. doi:10.1016/j.parkreldis.2020.07.017
90. Farmen K, Nissen SK, Stokholm MG, et al. Monocyte markers correlate with immune and neuronal brain changes in REM sleep behavior disorder. *Proc Natl Acad Sci*. 2021;118(10):e2020858118. doi:10.1073/pnas.2020858118
91. Terkelsen MH, Klastrup IH, Hvingelby V, Lauritsen J, Pavese N, Romero-Ramos M. Neuroinflammation and Immune Changes in Prodromal Parkinson's Disease and Other Synucleinopathies. *J Park's Dis*. 2022;12(Suppl 1):S149-S163. doi:10.3233/jpd-223245
92. Pont-Sunyer C, Tolosa E, Caspell-Garcia C, et al. The prodromal phase of leucine-rich repeat kinase 2-associated Parkinson disease: Clinical and imaging Studies. *Mov Disord*. 2017;32(5):726-738. doi:10.1002/mds.26964
93. Soto M, Fernández M, Bravo P, et al. Differential serum microRNAs in premotor LRRK2 G2019S carriers from Parkinson's disease. *Npj Park Dis*. 2023;9(1):15. doi:10.1038/s41531-023-00451-x
94. Stefani A, Serradell M, Holzknecht E, et al. Low Specificity of Rapid Eye Movement Sleep Behavior Disorder Questionnaires: Need for Better Screening Methods. *Mov Disord*. Published online 2023. doi:10.1002/mds.29407
95. Halsband C, Zapf A, Sixel-Döring F, Trenkwalder C, Mollenhauer B. The REM Sleep Behavior Disorder Screening Questionnaire is not Valid in De Novo Parkinson's Disease. *Mov Disord Clin Pract*. 2018;5(2):171-176. doi:10.1002/mdc3.12591

Figure legends

It is made available under a [CC-BY-NC-ND 4.0 International license](#).

Figure 1 Discovery phase analysis and targeted proteomic panel development. (A) Volcano plot depicting 89 proteins differentially expressed between Parkinson's patients and controls in the discovery phase. (B) Significantly enriched pathways based on the findings from the untargeted proteomic analysis. (C) Based on proteins identified through further screening studies, the targeted protein panel was developed, consisting of 87 proteins.

Figure 2 Targeted urine proteomics of Parkinsonian samples. (A) A principal component analysis (PCA) of targeted urine proteomics in the Parkinsonian sample groups iPD, YOPD, Parkin and LRRK2, a prodromal group consisting of iRBD, and healthy controls. The scores demonstrated no clear differences between the sample groups, indicating subtle differences in protein expression between the patient groups and controls. (B) PCA loadings showing the proteins responsible for the samples' scores distribution.

Figure 3 Orthogonal partial least squares discriminant analysis results of targeted proteomic analysis in iRBD and Parkinson's disease. Displayed results represent orthogonal partial least squares discriminant analysis (OPLS-DA) comparisons of idiopathic REM-sleep behaviour disorder (iRBD) patients versus controls, and their communalities with idiopathic PD (iPD) versus controls. (A) The iRBD OPLS-DA model discriminates significantly against controls (ANOVA-CV $P = 0.002$), with differences visualised in a iBiplot (a second orthogonal component has been added for visualisation purposes). (B) Predictive loading plot showing the proteins with strongest effect on the discriminating ability of the model. (C) Shared-and-unique structures plot of the two models (i) iRBD versus control, (ii) and iPD versus control, demonstrated that the two models share characteristics with, among others, UBC, NCAM1, MIEN1, SPP2 and REG1B being positively correlated with the disease groups, and ITIH2, BCHE and C3 negatively correlated.

It is made available under a [CC-BY-NC-ND 4.0 International license](#).

Figure 4 Receiver operating characteristics curve of the targeted urine proteomic panel to differentiate iRBD from healthy controls. ROC plot showing the results from a logistic regression classifier, using the individually significant proteins separately, and in combination.

Figure 5 Hierarchical clustering of the urine proteome by diagnosis/genotype. Hierarchical clustering of the signed log10 p-values from the comparisons of respective parkinsonian groups and iRBD to control. The clustering method was set to “weighted” and the metric was “cosine.” The row-wise clustering of the results from the sample groups demonstrates that iPD and iRBD share characteristics, also to a lesser degree with YOPD and LRRK2. PARKIN does not make part of this general cluster. Column-wise, the proteins are separated into two top-level clusters, generally characterised by higher values in the Parkinsonian conditions compared to control to the left and lower values to the right.

Figure 6 Clinical correlations with proteomic changes in iRBD samples. (A) Network representation of significant correlations with clinical measures for iRBD and proteins measured by targeted proteomics. Pink lines represent negative correlation and green lines positive correlations. Line thickness represents -log10 correlation p-value. (B) Euler diagram showing the overlapping correlations of the proteins significantly correlated with the clinical measures. HSPA1L, MYO10, MSN and VCAM1 were correlated with all three clinical measures. (Spearman correlation, no post-hoc p-value correction)

Table 1. Demographic and clinical details of male only donors of urine samples analysed for targeted proteomics. The number of samples, age, motor disease duration, motor disease severity (MDS-UPDRS III score) and levodopa-equivalent dose (LEDD) \pm standard deviation are reported for healthy controls, iPD, iRBD, LRRK2, YOPD and PARKIN carriers.

	n	Age \pm SD	Motor disease duration \pm SD	MDS-UPDRS III \pm SD	LEDD \pm SD
Combined controls	3	66.1 \pm			
	5	10.4	0	N/A	0
	2	66.4 \pm			
Control (London)	1	11.5	0	N/A	0
Control (Kosice)	1	65.7 \pm 9.2	0	N/A	0

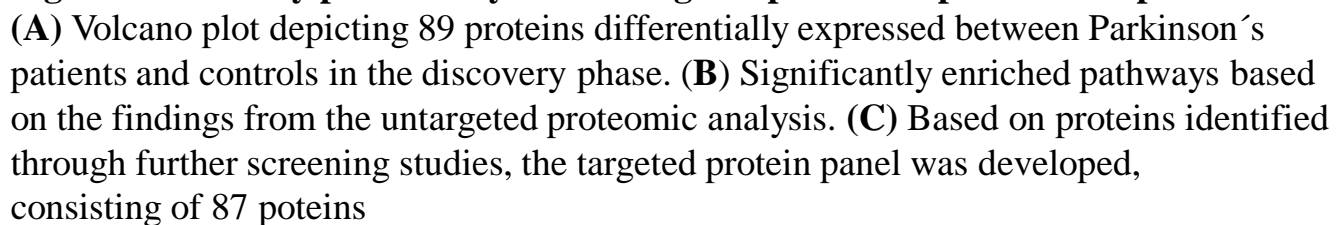
It is made available under a [CC-BY-NC-ND 4.0 International license](#).

	4				
LRRK2 (1AS + 1 9S)	0	55.4 ± 5.9	8.7 ± 4.5	34.3 ± 18.0	437.2 ± 223.9
YOPD	1				
	5	48 ± 8.7	9.3 ± 6.5	23 ± 13.4	809.7 ± 757
Parkin	5	47.6 ± 3.6	22.8 ± 16.2	24.6 ± 12	667.5 ± 702.3
iPD	3				
	5	69.3 ± 8.9	8.1 ± 4.3	30.2 ± 15.5	830 ± 722.9
iRBD	1				
	4	67.1 ± 9.3	N/A	N/A	N/A

iPD: idiopathic Parkinson's disease, iRBD: Rapid Eye Movement sleep disturbance disorder, LRRK2 (S): symptomatic LRRK2 mutation carrier, LRRK2 (AS): asymptomatic LRRK2 mutation carrier, YOPD: young onset Parkinson's disease, Parkin: mutations in the PRKN gene cause autosomal recessive juvenile Parkinson's disease, SD: standard deviation, LEDD: levodopa equivalent daily dose

Table 2. Clinical and proteomic correlations in Parkinsonian patients Correlations with clinical MDS-UPDRS III score, motor disease duration and LEDD. The correlation was evaluated in the pooled group iPD, YOPD, LRRK2 and PARKIN. Spearman correlation was utilised and no post-hoc correction of the *P*-values was performed. The arrows indicate if the correlation was positive or negative.

	Age	Motor disease duration	MDS-UPDRS III	LEDD
Age			0.02 (↑)	0.03 (↑)
Motor disease duration			0.01 (↑)	0.0006 (↑)
MDS-UPDRS III (ON)	0.02 (↑)	0.01 (↑)		0.02 (↑)
LEDD	0.03 (↑)	0.0006 (↑)	0.02 (↑)	
APOA4	0.007 (↑)			
BCHE	0.002 (↑)			
HPX			0.05 (↓)	0.006 (↓)
ICAM1			0.01 (↓)	
IGHA2	0.009 (↑)			
ITIH2		0.005 (↑)		
MASP2	0.02 (↓)			
MSN		0.02 (↓)		0.04 (↓)
NCAM1		0.02 (↓)		
REG1B	0.02 (↓)			
SERPING1		0.03 (↓)		
TXN		0.02 (↓)		
VCAM1		0.02 (↑)	0.05 (↑)	



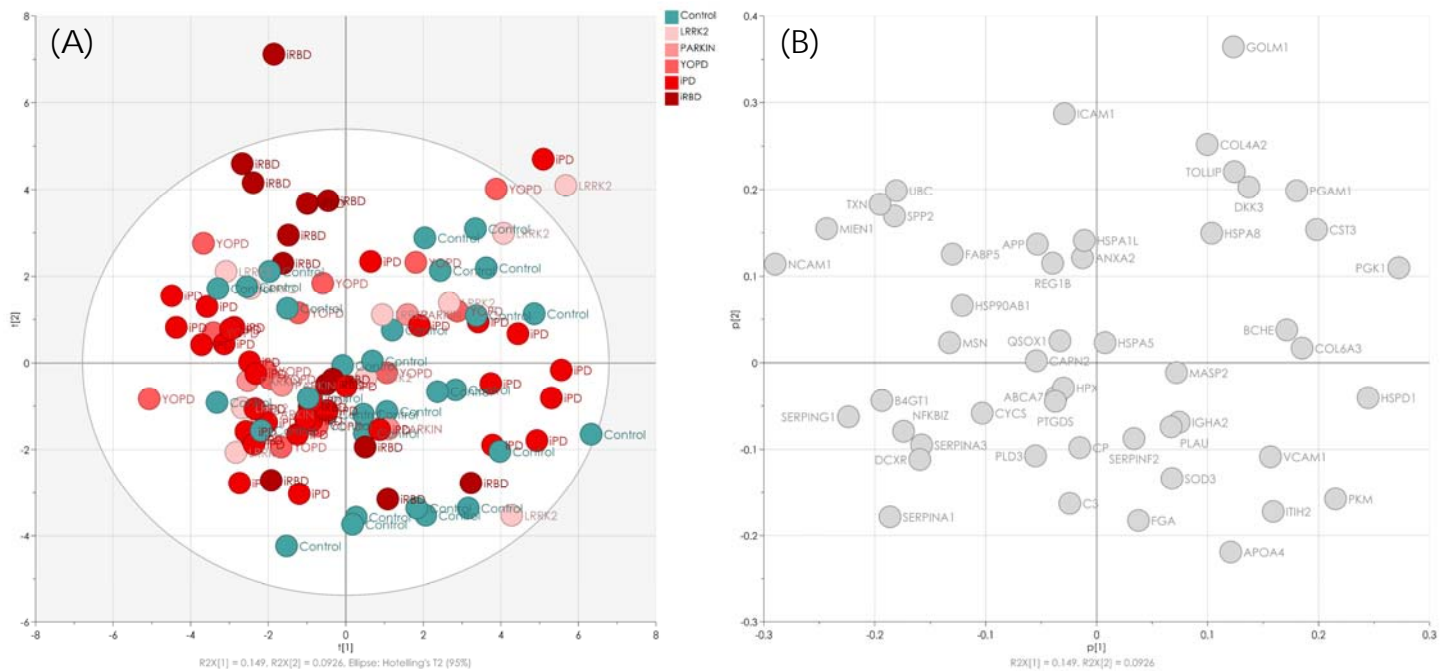


Figure 2 Targeted urine proteomics of Parkinsonian samples.

(A) A principal component analysis (PCA) of targeted urine proteomics in the Parkinsonian sample groups iPD, YOPD, Parkin and LRRK2, a prodromal group consisting of iRBD, and healthy controls. The scores demonstrated no clear differences between the sample groups, indicating subtle differences in protein expression between the patient groups and controls. (B) PCA loadings showing the proteins responsible for the samples' scores distribution.

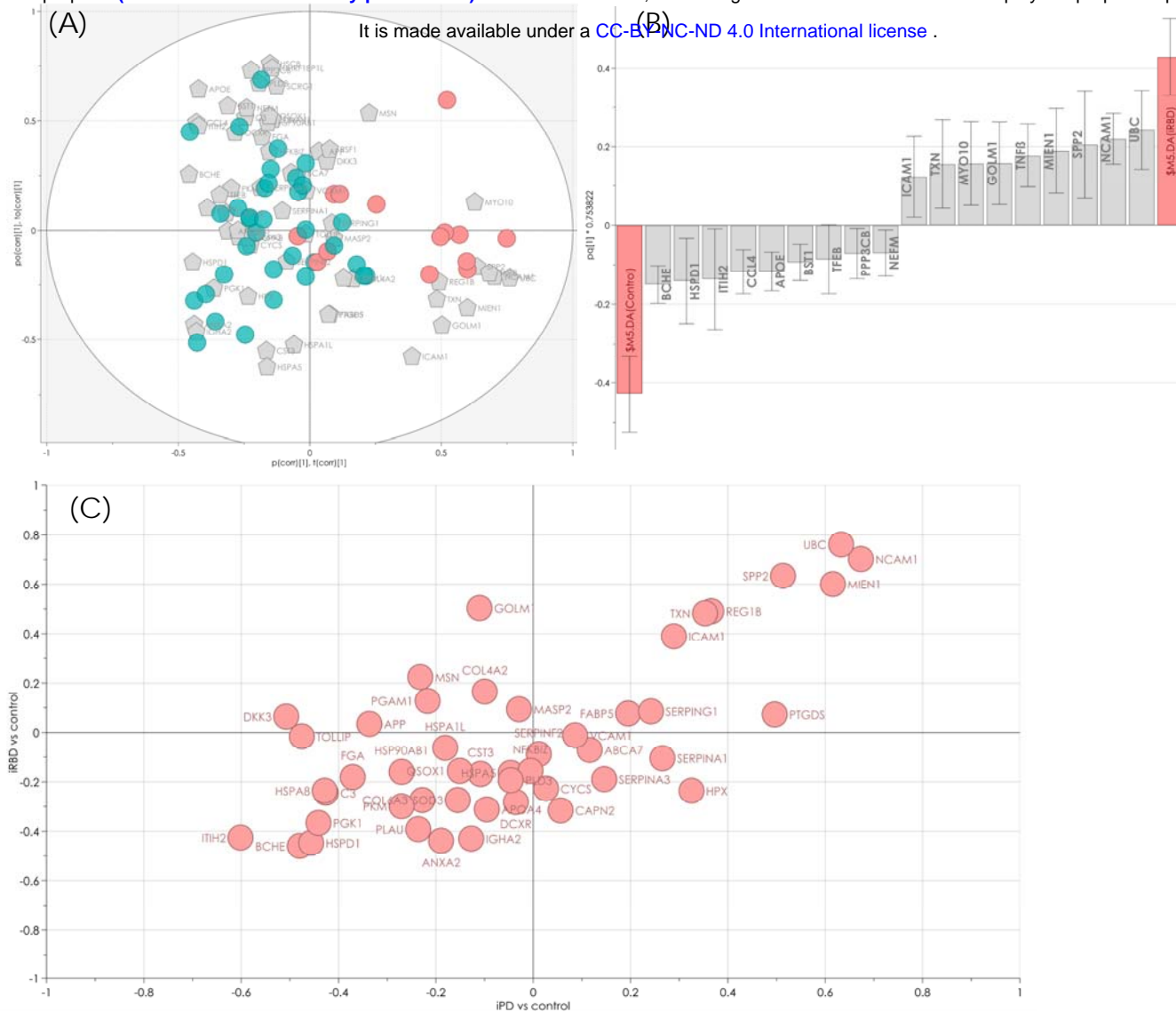


Figure 3 Orthogonal partial least squares discriminant analysis results of targeted proteomic analysis in iRBD and Parkinson's disease. Displayed results represent orthogonal partial least squares discriminant analysis (OPLS-DA) comparisons of idiopathic REM-sleep behaviour disorder (iRBD) patients versus controls, and their communalities with idiopathic PD (iPD) versus controls. (A) The iRBD OPLS-DA model discriminates significantly against controls (ANOVA-CV $P = 0.002$), with differences visualised in a iBiplot (a second orthogonal component has been added for visualisation purposes). (B) Predictive loading plot showing the proteins with strongest effect on the discriminating ability of the model. (C) Shared-and-unique structures plot of the two models (i) iRBD versus control, (ii) and iPD versus control, demonstrated that the two models share characteristics with, among others, UBC, NCAM1, MIEN1, SPP2 and REG1B being positively correlated with the disease groups, and ITIH2, BCHE and C3 negatively correlated.

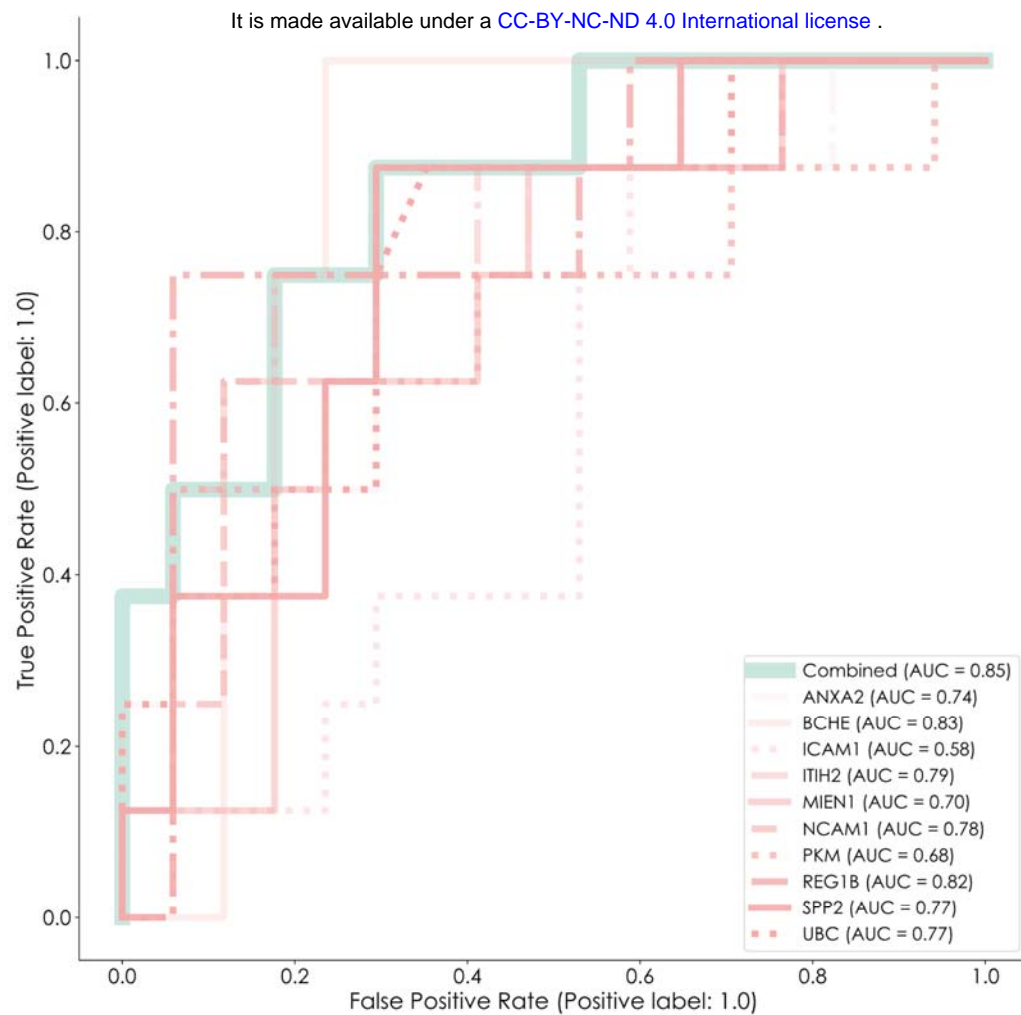


Figure 4 Receiver operating characteristics curve of the targeted urine proteomic panel to differentiate iRBD from healthy controls. ROC plot showing the results from a logistic regression classifier, using the individually significant proteins separately, and in combination.

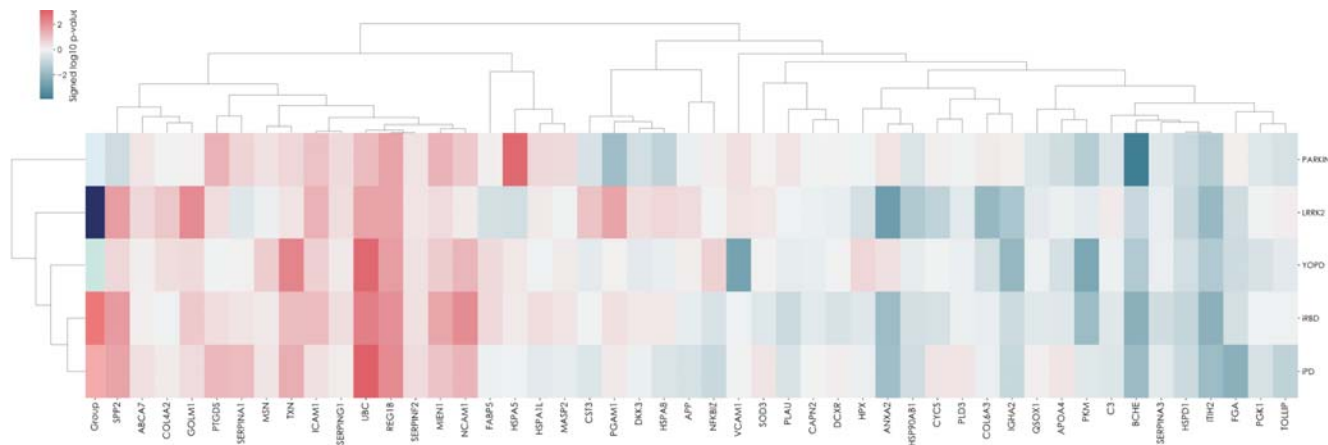


Figure 5 Hierarchical clustering of the urine proteome by diagnosis/genotype. Hierarchical clustering of the signed log₁₀ p-values from the comparisons of respective parkinsonian groups and iRBD to control. The clustering method was set to “weighted” and the metric was “cosine.” The row-wise clustering of the results from the sample groups demonstrates that iPD and iRBD share characteristics, also to a lesser degree with YOPD and LRRK2. PARKIN does not make part of this general cluster. Column-wise, the proteins are separated into two top-level clusters, generally characterised by higher values in the Parkinsonian conditions compared to control to the left and lower values to the right.

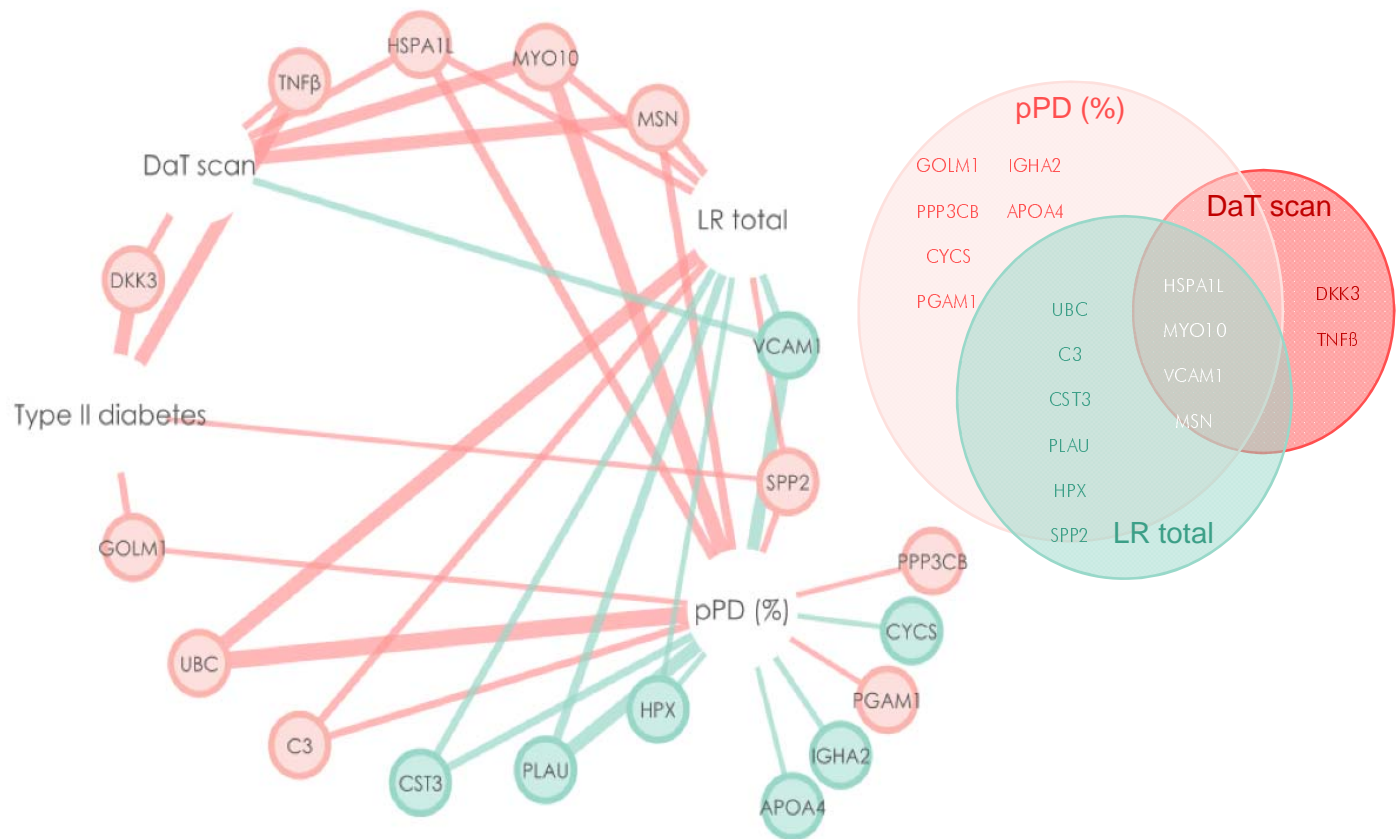


Figure 6 Clinical correlations with proteomic changes in iRBD samples. (A) Network representation of significant correlations with clinical measures for iRBD and proteins measured by targeted proteomics. Pink lines represent negative correlation and green lines positive correlations. Line thickness represents $-\log_{10}$ correlation p-value. **(B)** Euler diagram showing the overlapping correlations of the proteins significantly correlated with the clinical measures. HSPA1L, MYO10, MSN and VCAM1 were correlated with all three clinical measures. (Spearman correlation, no post-hoc p-value correction)