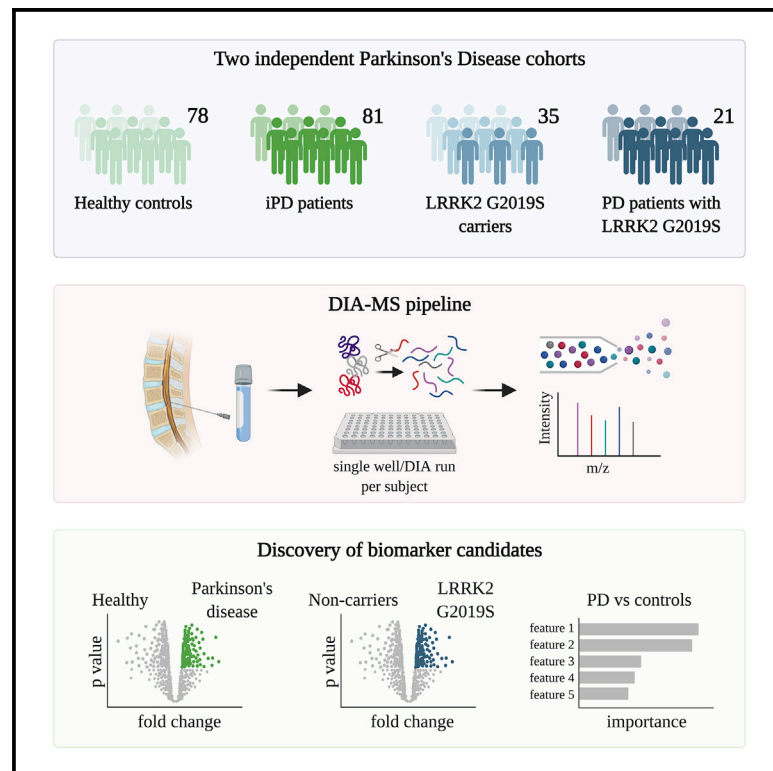


Proteome profiling of cerebrospinal fluid reveals biomarker candidates for Parkinson's disease

Graphical abstract



Authors

Ozge Karayel, Sebastian Virreira Winter, Shalini Padmanabhan, ..., Anne-Marie Wills, Clemens R. Scherzer, Matthias Mann

Correspondence

virreirawinter@omicera.com (S.V.W.),
mmann@biochem.mpg.de (M.M.)

In brief

Karayel et al. present a scalable, sensitive, and reproducible mass spectrometry-based proteomics workflow for profiling of cerebrospinal fluid and demonstrate its promise for biomarker discovery for idiopathic and familial forms of Parkinson's disease.

Highlights

- Reproducible and scalable MS-based proteomics workflow for proteome profiling of CSF
- Quantification of over 1,700 proteins collected from more than 200 individuals
- Lysosomal and immune-related biomarker signatures specific to LRRK2 G2019S carriers
- Overlapping PD- and LRRK2-associated alterations in CSF and urine proteomes



Article

Proteome profiling of cerebrospinal fluid reveals biomarker candidates for Parkinson's disease

Ozge Karayel,^{1,8,10} Sebastian Virreira Winter,^{1,9,10,*} Shalini Padmanabhan,² Yuliya I. Kuras,^{3,4} Duc Tung Vu,¹ Idil Tuncali,^{3,4} Kalpana Merchant,⁶ Anne-Marie Wills,⁵ Clemens R. Scherzer,^{3,4,5} and Matthias Mann^{1,7,11,*}

¹Department of Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, Martinsried, Germany

²The Michael J. Fox Foundation for Parkinson's Research, New York, NY, USA

³APDA Center for Advanced Parkinson Research, Harvard Medical School, Brigham and Women's Hospital, Boston, MA, USA

⁴Precision Neurology Program, Harvard Medical School, Brigham and Women's Hospital, Boston, MA, USA

⁵Department of Neurology, Massachusetts General Hospital, Boston, MA, USA

⁶Northwestern University Feinberg School of Medicine, Chicago, IL, USA

⁷Novo Nordisk Foundation Center for Protein Research, Faculty of Health Sciences, University of Copenhagen, Copenhagen, Denmark

⁸Present address: Department of Physiological Chemistry, Genentech, South San Francisco, CA, USA

⁹Present address: OmicEra Diagnostics GmbH, Behringstr. 6, 82152 Planegg, Germany

¹⁰These authors contributed equally

¹¹Lead contact

*Correspondence: virreirawinter@omicera.com (S.V.W.), mmann@biochem.mpg.de (M.M.)

<https://doi.org/10.1016/j.xcrm.2022.100661>

SUMMARY

Parkinson's disease (PD) is a growing burden worldwide, and there is no reliable biomarker used in clinical routines to date. Cerebrospinal fluid (CSF) is routinely collected in patients with neurological symptoms and should closely reflect alterations in PD patients' brains. Here, we describe a scalable and sensitive mass spectrometry (MS)-based proteomics workflow for CSF proteome profiling. From two independent cohorts with over 200 individuals, our workflow reproducibly quantifies over 1,700 proteins from minimal CSF amounts. Machine learning determines OMD, CD44, VGF, PRL, and MAN2B1 to be altered in PD patients or to significantly correlate with clinical scores. We also uncover signatures of enhanced neuroinflammation in LRRK2 G2019S carriers, as indicated by increased levels of CTSS, PLD4, and HLA proteins. A comparison with our previously acquired urinary proteomes reveals a large overlap in PD-associated changes, including lysosomal proteins, opening up new avenues to improve our understanding of PD pathogenesis.

INTRODUCTION

Parkinson's disease (PD) is the second most common neurodegenerative disease, affecting millions of people worldwide and having a strikingly increased incidence with age.^{1,2} The hallmark pathology of PD is well characterized as α -synuclein aggregates and dopamine neuronal loss;² however, molecular events that trigger PD are not fully understood. As a result, all current treatments only target symptoms, without slowing or reversing disease progression.³ Thus, detecting protein level alterations in PD could provide insight into the underlying mechanism of disease and aid in the development of novel therapeutics.

The majority of PD cases are idiopathic, while for some a genetic linkage is apparent.^{3–5} Mutations in the LRRK2 gene, most commonly G2019S, are the most frequent genetic cause of autosomal dominant PD.^{6,7} Importantly, all PD-associated mutations activate leucine-rich repeat kinase 2 (LRRK2) kinase activity, offering a promising therapeutic target for PD by inhibiting this function.^{7–9} Better understanding of LRRK2-PD pathophysiology and discovery of related biomarkers would facilitate LRRK2-targeted therapies. It is also essential to determine whether pathogenic mechanisms associated with LRRK2-PD

may be present at the prodromal stage in non-manifesting LRRK2 mutation carriers and in a subset of idiopathic PD.

Cellular and animal model studies indicate that LRRK2 is involved in the regulation of several pathophysiologic processes, including autophagy, endolysosomal membrane/vesicular trafficking, and immune responses, with at least some of the effects mediated through phosphorylation of a subgroup of Rab GTPase.^{10–13} Whereas cell biology research on molecular pathways affected by LRRK2 is critical for insights into disease mechanisms, a complementary approach would be to examine molecular changes in biospecimens from deeply phenotyped LRRK2 and idiopathic PD cohorts compared with controls. We thus reasoned that an unbiased, mass spectrometry (MS)-based proteomics analysis of cerebrospinal fluid (CSF) from healthy controls, PD individuals with and without LRRK2 G2019S mutation and non-manifesting LRRK2 G2019S carriers could uncover much-needed biomarkers. Compared with other biofluids, CSF should reflect the disease-related pathology of the brain and spinal cord more accurately.^{14,15}

MS-based proteomics is a very powerful technology for detecting differences in protein abundance levels in healthy individuals and patients, and thus, in principle, an ideal tool for



Table 1. Demographics of all participants

LCC cohort	LRRK2-/PD- (n = 31)	LRRK2+/PD- (n = 35)	LRRK2-/PD+ (n = 34)	LRRK2+/PD+ (n = 21)
Sex (male/female)	13/18	19/16	23/11	10/11
Age at collection, mean (SD)	54.2 (13.9)	52.1 (15.2)	57.2 (11.2)	62.2 (10.8)
Age at onset, mean (SD)	n/a	n/a	51.1 (11.4)	53.9 (11.4)
UPDRS III, mean (SD)	n/a	n/a	25.9 (9.4)	23.1 (15.6)
MOCA score, mean (SD)	27.3 (2)	26.2 (2.5)	26.5 (3)	24.7 (3.9)
HBS cohort	PD- (n=47)	PD+ (n=47)		
Sex (male/female)	29/18	34/13		
Age at collection, mean (SD)	56.4 (11.3)	62.8 (8.9)		
Age at onset, mean (SD)	n/a	58.4 (8.4)		
GBA (mut/WT), mean (SD)	0/47	6/41		
LRRK2 (mut/WT), mean (SD)	0/47	1/46		
UPDRS total, mean (SD)	n/a	37.5 (14.3)		
UPDRS III, mean (SD)	n/a	22.1 (14.1)		
MMSE score, mean (SD)	28.4 (1.8)	28.3 (1.6)		

biomarker discovery. However, proteomic analysis of CSF has been challenging due to low protein concentration combined with the high dynamic range of protein abundances, resulting in low quantification precision, throughput, and limited proteome depth.^{16–18} Recent advances in the proteomics field, from automated sample preparation to more sensitive MS instrumentation, data acquisition methods such as data-independent acquisition (DIA) and processing software, have enabled substantial proteome coverage and precise quantitation in single LC-MS runs.^{19–23} Our group has combined these advances into a streamlined and highly reproducible workflow resulting in a large number of consistently measured and biologically meaningful proteome changes in various biofluid/tissue specimens in a variety of clinical cohorts.^{16,17,24–27}

In this study, we extended our pipeline of biomarker discovery to analyze more than 200 CSF samples from two independent cohorts in which we detected over 1,700 proteins from minimal CSF sample amounts. By employing co-variate (ANCOVA) analysis and machine learning, we identified unique protein signatures whose abundance was specifically changed in PD patients versus healthy controls and LRRK2 G2019S carriers versus non-carriers. Our study demonstrates that modern MS-based proteomics is a powerful technology for biomarker discovery in biofluids. It also provides potential biomarkers of PD as well as insights into biological pathways associated with PD and/or LRRK2 mutations.

RESULTS

Overview of PD cohorts for CSF proteome profiling

To investigate how PD affects the CSF of patients and to identify potential biomarkers, we employed the “rectangular” biomarker discovery strategy, which aims to discover discriminating proteome signatures using rather large sample sizes in both discovery and validation cohorts.^{16,17,24,28,29} Applying this approach, we analyzed CSF samples from 215 individuals from two independent cohorts including 113 healthy controls (HC) and 102 PD pa-

tients. The first cohort consisted of 94 CSF samples from the Harvard Biomarkers Study (hereinafter referred to as the HBS cohort).^{30–33} The second cohort was a subset of biobanked CSF samples from the Michael J. Fox Foundation for Parkinson’s Research (MJFF)-funded *LRRK2* Cohort Consortium (hereinafter referred to as the LCC cohort). More information about the cohorts is summarized in Table 1, Table S1 and Figure S1.

Proteomic characterization of CSF samples

We have developed a robust, automated, and high-throughput MS-based proteomics workflow using a data-independent acquisition (DIA) strategy to perform proteome profiling of minimal amounts of CSF^{21,24,34} (Figure 1A). This strategy has previously resulted in an unprecedented depth at high data completeness in CSF and revealed biologically meaningful proteome changes across multiple independent Alzheimer’s disease cohorts.²⁴ Here, we applied our workflow to discover proteome changes in the CSF of PD patients with or without the disease-associated G2019S mutation in *LRRK2*. To maximize proteome depth and coverage, we generated cohort-specific hybrid spectral libraries by merging three sub-libraries: (1) a library constructed by data-dependent acquisition (DDA) consisting of 24 fractions of pooled CSF samples; (2) another DDA library consisting of eight fractions of extracellular vesicles enriched from pooled CSF samples; and (3) a direct-DIA library generated from the DIA analysis of all samples (see STAR Methods). Matching to cohort-specific hybrid libraries of 5,418 and 3,167 proteins yielded 1,493 and 1,626 proteins in total for the HBS and LCC cohorts, respectively, with more than 1,300 in common (Figure 1B). On average, we quantified 1,357 (HBS) and 1,481 (LCC) protein groups per neat CSF sample in single runs with 1,290 (HBS) and 1,440 (LCC) protein groups quantified in more than 70% of the samples (Figures 1C, 1D, S1J, S1K, and Table S2). Our DIA workflow resulted in a much deeper proteome coverage compared with previous studies applying a similar single-run DIA strategy without compromising throughput.^{28,29,35} The quantified protein intensities spanned over four orders of magnitude (Figures 1E and 1F). The top 10

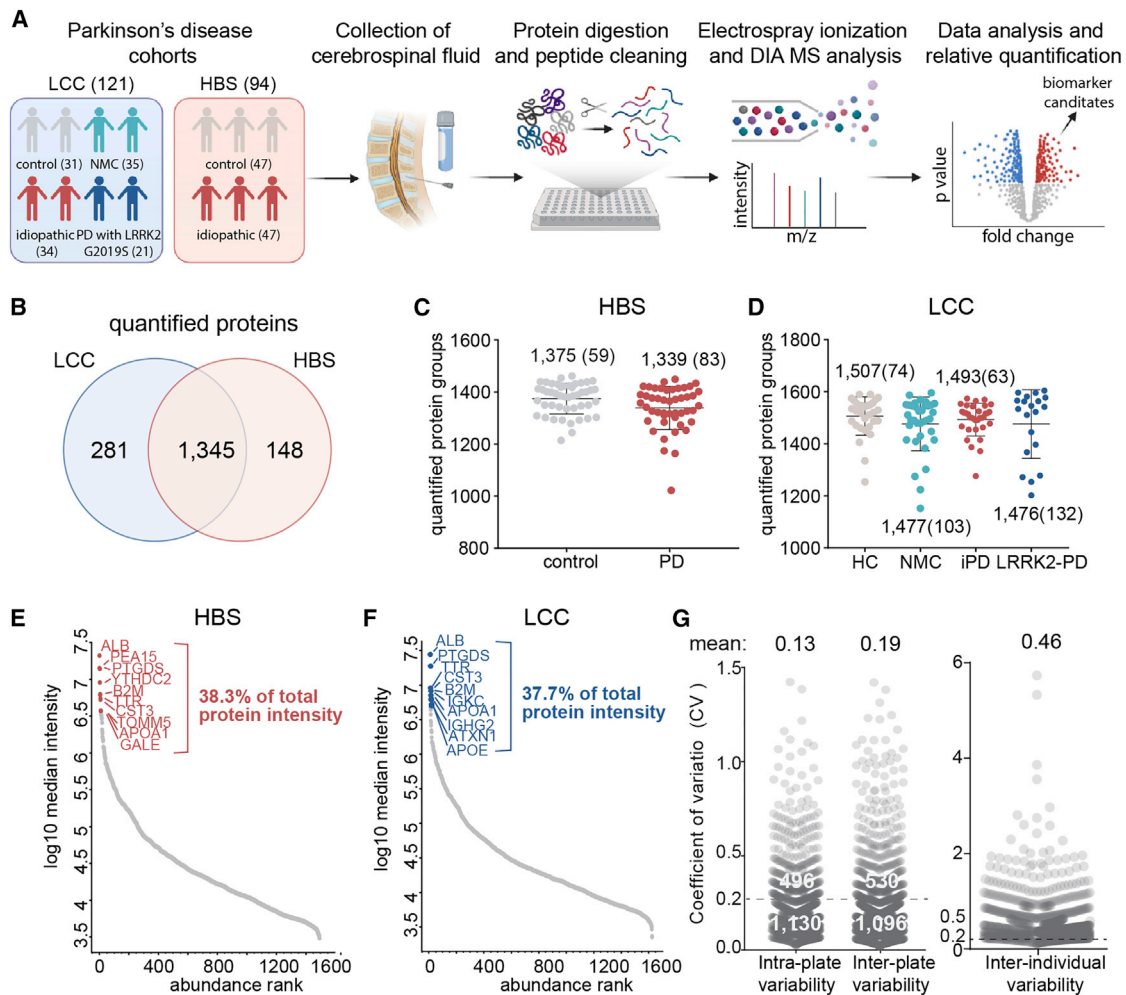


Figure 1. MS-based proteomic analysis of two independent CSF PD cohorts

(A) Overview of the CSF proteomic workflow. CSF samples were prepared in 96-well plates using an automated liquid-handling system and analyzed by LC-MS/MS using data-independent acquisition (DIA). The total number of subjects per cohort group is shown.

(B) A total of 1,345 proteins were consistently quantified in both cohorts.

(C and D) Number of proteins identified and quantified with a 1% false discovery rate (FDR) in each sample in the HBS (C) and LCC (D) cohorts. Numbers indicate mean and standard deviation (SD).

(E and F) Proteins identified in the HBS (E) and LCC (F) cohorts were ranked according to their MS signals. The top 10 most abundant proteins are labeled, and their relative contribution to the total protein intensity is indicated.

(G) Quantification precision assessed by calculating the intra- and inter-plate (between repeated measurements of the same sample) and inter-individual coefficients of variation (CVs) of all proteins. Number of proteins with a CV below and above 20% and mean CV values are shown.

most abundant proteins alone contributed around 38% to the total CSF proteome signals, illustrating the analytic challenges (Figures 1E and 1F).

We further classified the identified proteins based on their human protein atlas (HPA) annotation³⁶ as secreted from the cell, intra-cellular or in the cellular membranes (note that many proteins are assigned to several compartments). In total, 94% of identified proteins carried at least one annotation, of which 36% were secreted proteins, while 70% were intra-cellular and 36% membrane-spanning proteins (Figure S2A). Furthermore, the majority of quantified proteins were annotated to be enriched in brain and liver, in line with the fact that CSF is the extracellular fluid of the CNS and derived from blood plasma,

which contains many proteins synthesized in the liver (Figure S2B). In summary, using small CSF volumes, we obtained a very high CSF proteome coverage for single-run analysis, thus providing a promising basis for the discovery of biomarkers in PD.

Assessment of quantification precision and sample quality

To assess quantification precision in our study, we investigated intra- and inter-assay variabilities of our automated workflow by repeated measurements of a pooled CSF sample (Figure S2C). This analysis revealed high technical reproducibility with 13% intra-plate and 19% inter-plate coefficients

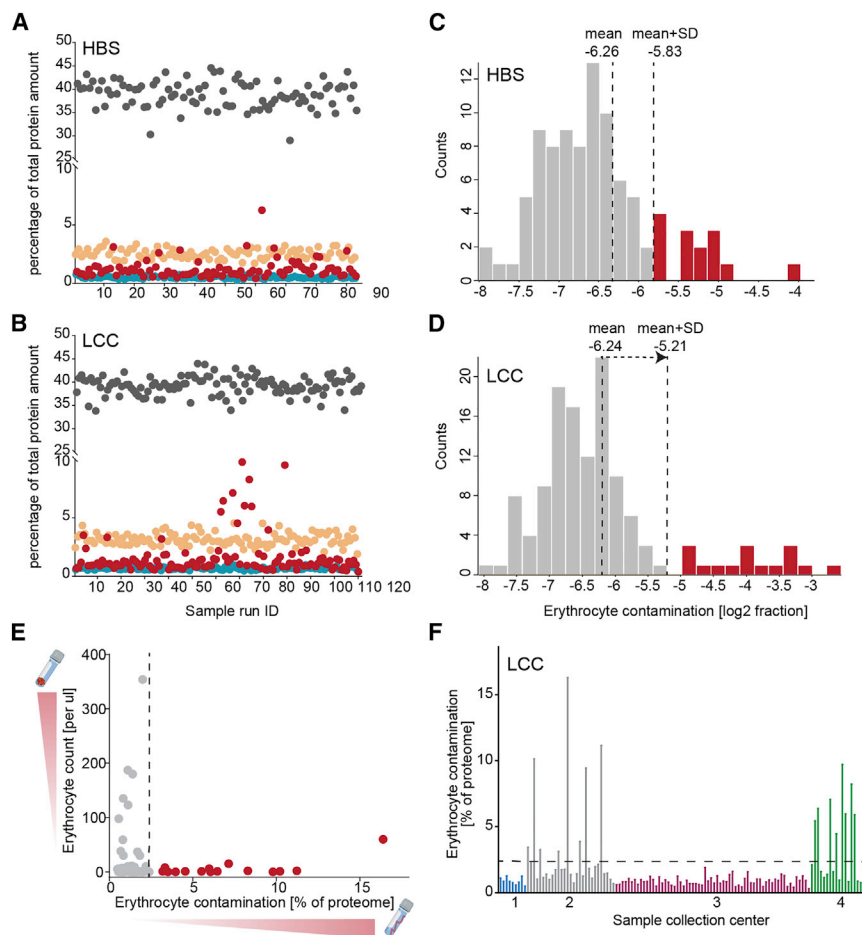


Figure 2. Quality assessment of two independent CSF PD cohorts

(A and B) Assessment of study quality by determining the percentage of the summed intensity of the proteins in the respective quality marker panel and the summed intensity of all proteins in the HBS (A) and LCC (B) cohorts. Erythrocyte-specific protein panel (red), platelet marker panel (turquoise), coagulation marker panel (orange), and the top 10 most abundant protein panel (dark gray) are included in these analyses. The proteins in each quality marker panel are listed in [Table S3](#).

(C and D) Histograms of log₂ ratios of the summed intensity of the erythrocyte-specific proteins and the summed intensity of all proteins in the HBS (C) and LCC (D) cohorts. A sample was flagged for potential contamination and removed from further analysis if the ratio differed more than one SD from the mean of all samples within the cohort.

(E) Comparison of erythrocyte counts in CSF following sample collection and degree of erythrocyte contamination as determined by MS-based proteomics of all LCC cohort samples. Samples colored in red were excluded from further analysis based on the distribution shown in (D).

(F) Grouping samples in the LCC cohort for four sample collection centers demonstrated a high degree of contamination with erythroid-specific proteins for study centers 2 and 4, whereas there was no indication of this in study centers 1 and 3.

of variation (CVs). Around 1,100 had inter- and intra-assay CVs below 20% and 1,500 below 50% (Figures 1G, S2C and S2D). The inter-individual biological variability between subjects is much larger, with only 10% of all proteins having a CV below 20% (Figure 1G). This demonstrates that the technical variability of our assay is much smaller than the biological variability—an important pre-condition for the successful discovery of PD-specific proteome signatures and potential biomarkers in CSF.

Inconsistent sample collection and handling may result in systematic bias and hamper the discovery of true biomarkers. To ensure that the quantified changes in CSF proteins are due to disease-related pathological alterations, we assessed the quality of all samples according to previously established quality marker panels for coagulation-related proteins, platelets, and erythrocytes to identify samples with potential issues in pre-analytical processing^{16,17,37} (Figures 2A and 2B). This revealed high sample-to-sample variability for the degree of contamination with our erythroid-specific marker panel, presumably due to puncture-related contamination with blood (Figures 2A and 2B). We flagged 14 samples (8 PD, 6 HC) from the HBS cohort and 17 samples (4 PD, 13 HC) from the LCC cohort for erythrocyte contamination and removed them from further analysis (Figures 2C and 2D). In total, pro-

teomes of 182 subjects fulfilled the quality criteria and were used for statistical analysis. We found no correlation between erythrocyte counts and the number of erythrocyte-specific proteins in the CSF (Figure 2E). Our samples were centrifuged following collection to remove cells and debris, explaining why initially high erythrocyte counts did not affect the proteome measurements. Hemolysis occurring during collection, however, would be visible only in the proteome and not in erythrocyte counts. Sorting the samples by study center revealed a systematic bias in the sample taking and processing procedure and identified two centers with high pre-analytical variation and a corresponding high degree of proteome contamination with erythroid-specific proteins. This further emphasizes the importance of standardized sample collection and processing procedures to minimize or avoid such biases and the advantage of unbiased proteomics to flag and remove such problematic cases.

We additionally further excluded a single sample from the HBS cohort that clustered far away from all other samples in a principal-component analysis (PCA) and a single sample from the LCC cohort that was an outlier in proteome depth. Our subsequent bioinformatics analysis was based on 79 samples from the HBS cohort and 103 subjects from the LCC cohort to determine the impact of disease manifestation and LRRK2 mutation status on the CSF proteome (Figures S2E and S2F).

PD-related proteome alterations in CSF and machine-learning-based classification of PD patients and controls

To investigate alterations in the CSF proteome of PD patients compared with controls, we performed an ANCOVA, considering age, sex, and *LRRK2* mutation status as confounding factors. We also included the different study centers as a confounding factor for the LCC cohort. For the HBS cohort, we further included the *GBA* mutation status. At 5% false discovery rate (FDR), we identified three significantly regulated proteins (CPM, OMD, and RFNG) in the CSF of PD patients compared with controls in the HBS and one (PRCP) in the LCC cohort (Figure 3A and Table S4). Osteomodulin (gene name: OMD) and the cell surface marker CD44 almost reached statistical significance in both cohorts (OMD: *q* values of 0.063 in LCC and 0.039 in HBS; CD44: *q* values of 0.08 in LCC and 0.12 in HBS; Figure 3A). One reason for the small overlap between the cohorts could be a less stringent and inconsistent sample collection protocol in the multi-center LCC study. Yet, the osteomodulin and CD44 proteins were robustly quantified in both cohorts and detected with three (CD44) and seven (OMD) peptides in all samples (Table S4). The levels of both proteins were significantly elevated in PD patients with mean fold changes of 1.22 (OMD) and 1.12 (CD44) in the HBS and 1.17 (OMD) and 1.07 (CD44) in the LCC cohort (Figures 3B–3E). These results demonstrate that the rectangular strategy was able to distinguish PD-related alterations comprising a few proteins differentially present in PD patients compared with controls from cohort-specific effects in the quantified CSF proteome, even in cohorts constrained by biases such as less stringent and inconsistent sample collection.

To examine whether PD affects particular cellular compartments and biological networks in CSF, we performed a gene ontology (GO) annotation enrichment analysis using the mean fold changes of PD versus HC.³⁸ The proteins elevated in the PD samples compared with the controls were enriched for lysosomal-related terms, further supporting the emerging role of lysosomes and mounting evidence for lysosomal dysregulation and associated α -synuclein aggregation in PD^{13,39,40} (Figure 3F).

Motivated by the presence—but small number—of commonly altered proteins in both cohorts, we tested how well machine learning (ML) could discriminate PD patients from controls by using the recently introduced open-source tool OmicLearn.⁴¹ ML approaches strongly benefit from a large number of samples, which prompted us to combine the HBS and LCC cohorts to identify reliable signatures. Using the Extra-Trees package, the 35 most discriminating proteins for training the model were selected in each training iteration, and they were ranked by the classifier according to their feature importance (Figure 3G). Interestingly, among them, prolactin (gene name: PRL), was the most important feature based on 50 training iterations (5 splits, 10 repeats). Although prolactin was not significantly regulated in either cohort, its release is known to be suppressed in PD patients taking dopaminergic medications.⁴² PRL did not correlate with levodopa equivalent daily doses (LEDD) provided to all PD subjects in the HBS cohorts (Figure S3E). Further promising candidates to classify PD versus controls were CD44, VGF, and—in agreement with the lysosomal pathway enrichment—lysosomal proteins cathepsin K (CTSK) and MAN2B1. When these features

to train XGBoost were used, an ensemble-tree-based model, with our cross-validation scheme, the mean area under the curve (AUC) of the receiver operating characteristic (ROC) curve was 0.72 ± 0.08 . The AUC is a frequently used measure of performance of the classification model and ranges from 0 (all predictions are wrong) to 1 (all predictions are correct). The sensitivity, which is the rate of correctly classified PD patients, and specificity, which is the rate of correctly classified PD-negative individuals, were 67% and 66%, respectively (Figure 3H). Taken together, our CSF proteome data, when combined with ML algorithms, classified disease status and, more importantly, identified several promising PD-associated proteins, opening up interesting leads for future studies.

Impact of the pathogenic *LRRK2* G2019S mutation on the CSF proteome

Given the substantial number of subjects carrying the *LRRK2* G2019S mutation in the LCC cohort, we explored whether the CSF proteome is altered by *LRRK2* mutation status. We again applied an ANCOVA with sex, age at sample collection, study center, and PD status as confounding factors. At a FDR of 5%, the mutation significantly altered the abundance of the proteins HLA-DRA, HLA-DRB1, HLA-DPA1, CTSS, PLD4, TKT, ITGB2, PRDX3, ITIH5, CNDP1, and FAH (Figure 4A). On the basis of the peptides identified in the samples, we could distinguish two forms of the protein HLA-DRB1, which were both significantly enriched in *LRRK2* G2019S carriers. Furthermore, Student's *t* test with a FDR of 5% visually confirmed the upregulation of HLA-DRA, HLA-DRB1, and HLA-DPA1 in the mutation carriers (Figures 4A and 4B). These HLA proteins as well as PLD4 and CTSS were robustly quantified with at least four peptides in the LCC cohort, and their levels were significantly elevated in PD patients (>1.7-fold; Figures 4B–4D and Table S4). Using this approach globally, we observed that proteins elevated in the CSF proteomes of *LRRK2* G2019S carriers compared with wild-type (WT) allele carriers were enriched for categories related to immune and inflammatory responses, further supporting the substantial evidence of a close association between enhanced inflammatory response and PD^{43,44} (Figure 4E).

Moreover, the analysis of non-manifesting *LRRK2* carriers (NMCs) to healthy individuals also revealed several HLA molecules, CTSS as well as ADAM10 and FAM234A (ITFG3) to be upregulated in pathogenic *LRRK2* carriers, with HLA-DRB1 being the only protein reaching the significance level at a FDR of 5% (Figure S2G). Pathogenic *LRRK2* carriers have a higher likelihood of developing PD compared with healthy individuals; thus, proteins identified in this analysis may include prognostic factors.

Cathepsins are proteases mediating protein degradation and turnover in endolysosomal compartments, and several of them have been implicated in inflammatory diseases, lysosomal storage, and neurodegenerative disorders such as PD.^{45,46} We have recently shown that the levels of multiple members of the cathepsin family, including cathepsins A, B, C, D, H, L, O, S, and Z, significantly increased in the urine of *LRRK2* G2019S carriers in two independent cohorts.²⁵ We also detected these cathepsins in CSF but—in contrast to urine—only cathepsin S (CTSS) was significantly affected by the mutation status (Figure 4F).

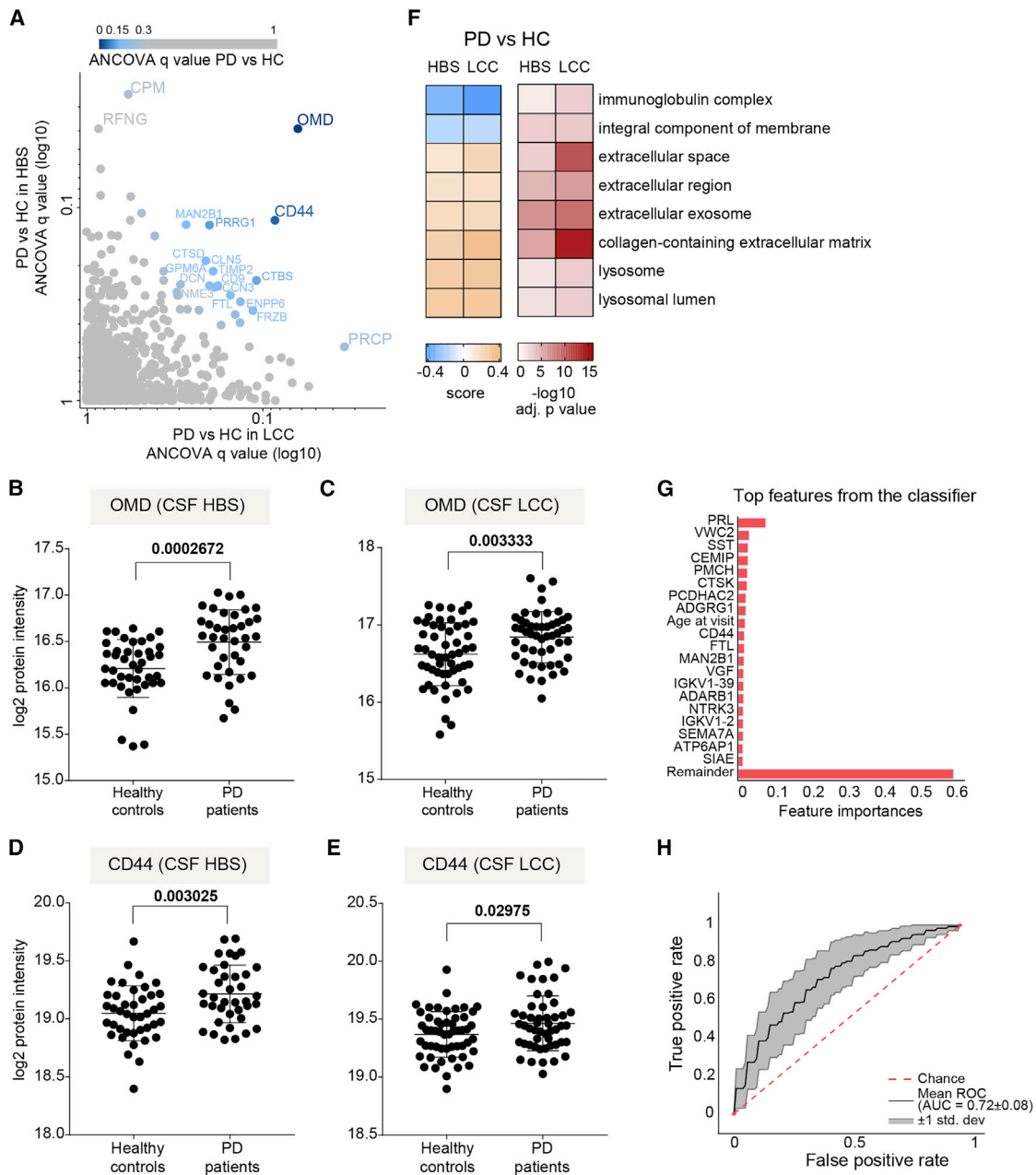


Figure 3. PD-related alterations in CSF proteome and ML-based classification of PD status

(A) Correlation of ANCOVA q values of all proteins quantified in the CSF of PD patients compared with controls in the HBS and LCC cohorts. Color gradient is based on the mean of ANCOVA q values (PD versus HC) obtained in the LCC and HBS cohorts.

(B and C) Osteomodulin (OMD) protein intensity (log₂) distribution in controls and PD patients of the HBS (B) and LCC (C) cohorts. We applied an unpaired t test and the resulting p value is shown.

(D and E) CD44 protein intensity (log₂) distribution in controls and PD patients of the HBS (D) and LCC (E) cohorts; the corresponding p values from unpaired t test are shown. In panels B–E, lines indicate mean and SD.

(F) Annotation enrichment of GO terms using the PD versus HC fold changes (5% FDR). All significantly enriched GO terms that were common in both cohorts are displayed. Terms with positive enrichment scores are enriched in PD over HC and vice versa.

(G) Feature importance of the top 20 most important features used to distinguish PD+ versus PD– individuals.

(H) ROC curve and corresponding AUC statistics in 5-fold cross-validation repeated 10 times using the XGBoost-based model to classify PD versus HC based on protein panel in (G). Random performance is indicated by the dotted diagonal red line for comparison. Gray area represents the SD from the mean ROC curve. Blue lines show the values for a total of repeats with five stratified train-test splits.

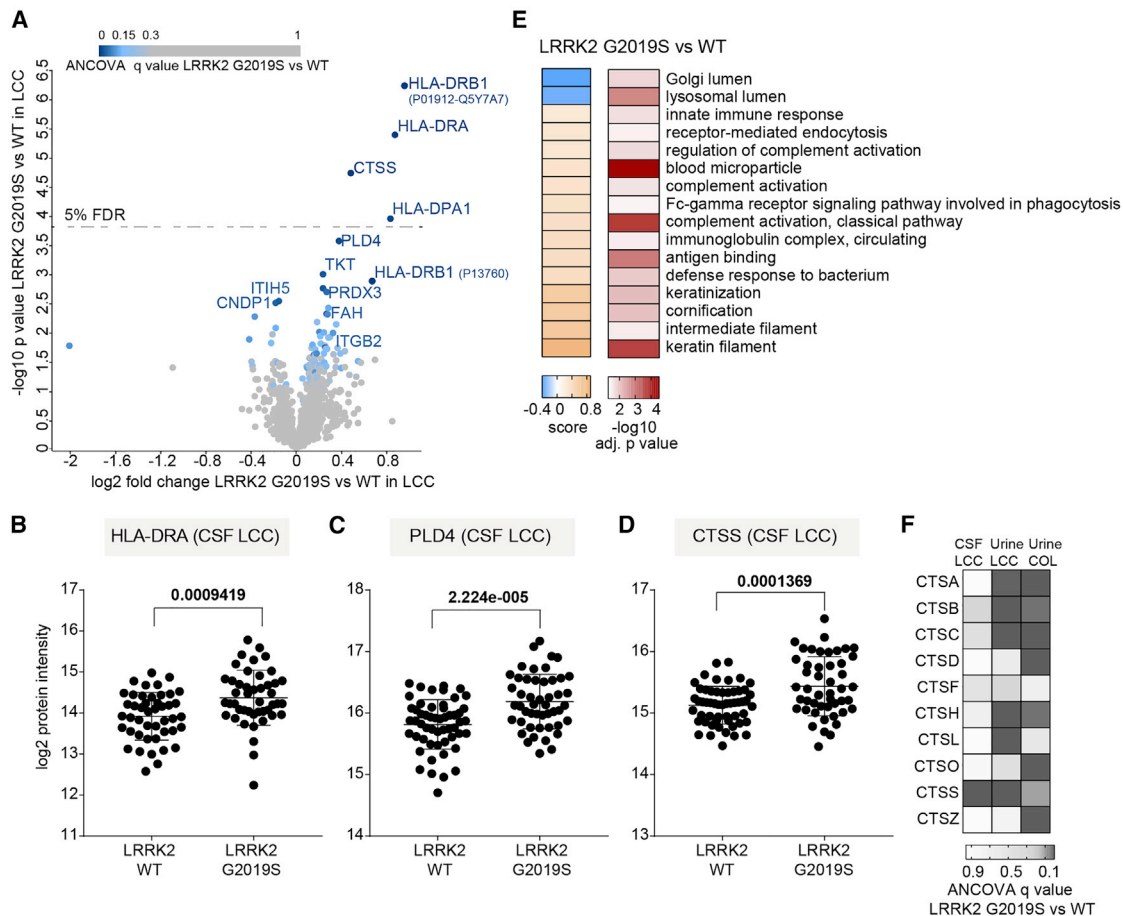


Figure 4. Effect of the pathogenic LRRK2 G2019S mutation on the CSF proteome

(A) Volcano plot comparing the CSF proteomes of LRRK2 G2019S versus WT carriers. The fold change in protein levels is depicted on the x axis and the $-\log_{10}$ t test p value on the y axis. Color scale is based on ANCOVA q values of the proteins differentially present in the CSF of LRRK2 G2019S carriers compared with the WT controls in the LCC cohort. Proteins with ANCOVA q values <5% are labeled.

(B–D) HLA-DRA (B), PLD4 (C), and CTSS (D) protein intensity (\log_2) distributions in LRRK2 G2019S and WT carriers of the LCC cohort; p values of an unpaired t test are shown. Lines indicate mean and SD.

(E) Annotation enrichment of GO terms using the LRRK2 G2019S versus WT fold changes (5% FDR). Terms with positive enrichment scores are enriched in the G2019S mutation over the WT and vice versa.

(F) Heatmap of ANCOVA q values of all cathepsin proteins, which were quantified in both CSF and urine of LRRK2 G2019S carriers compared with the WT controls.

Integration of CSF and urinary proteome profiles

We previously analyzed 235 urine samples from two independent cross-sectional cohorts (Columbia and LCC), including two types of controls, healthy individuals, and LRRK2 G2019S carriers not manifesting the disease, and PD patients with and without the LRRK2 G2019S mutation, quantifying 2,365 urinary proteins in total.²⁵ Encouraged by the great depth of the acquired urine and CSF proteomes in our previous and the present study, we decided to integrate both proteomes to determine co-regulated proteins. Although we analyzed both urine and CSF from subsets of the LCC cohort, there were no matching urine and CSF samples from the same individuals. More than 1,000 proteins overlapped, corresponding to 36% of all identified proteins in both biofluids (Figure 5A). Matching the common protein abundances revealed a clear correlation between the two biofluids (Pearson $r = 0.49$; Figure 5B). The most abundant proteins

present in both included ALB, PTGDS, ORM1, SERPINA1, B2M, and several apolipoproteins and immunoglobulins (Figure 5B, labeled proteins at the top right). Moreover, Fischer’s exact test on the GO terms associated with the proteins specifically found in either urine or CSF or their common overlap revealed terms significantly enriched compared with all proteins identified in the two body fluids (Figure 5C). As expected, the terms related to nervous system including “postsynaptic membrane,” “memory,” “neuropeptide signaling pathway,” and “nervous system development” were enriched among proteins present exclusively in CSF. In contrast, terms related to the endosome-lysosome pathway as well as Rab protein signal transduction were enriched among urine-specific proteins, indicating that proteins of the Rab-LRRK2-pathway are highly abundant in urine, in line with our previous finding that the LRRK2 G2019S mutation strongly affects the urinary proteome.²⁵

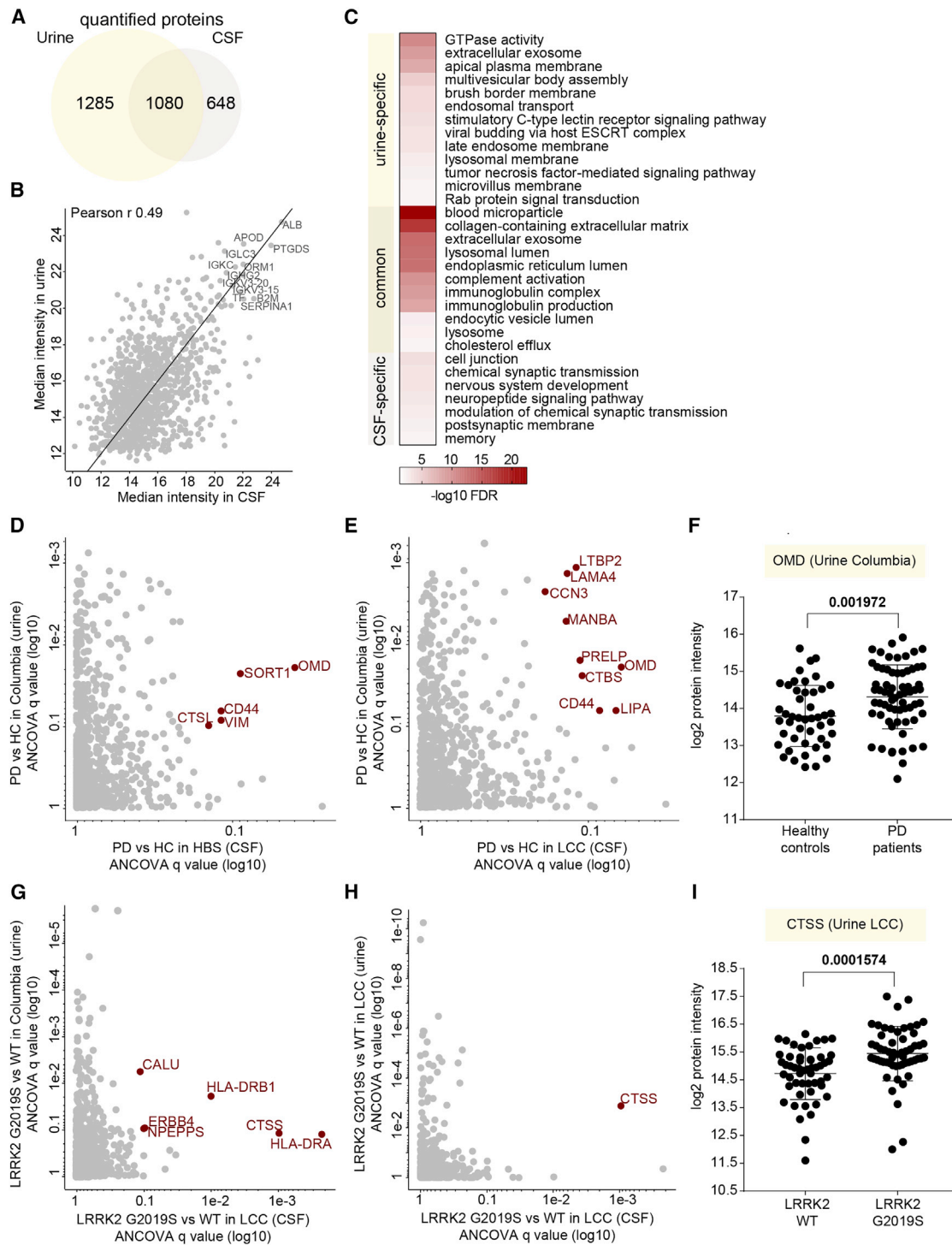


Figure 5. Integration of CSF and urine proteome profiles

(A) Overlapping proteins between the CSF and urinary proteomes.

(B) CSF-urine proteome abundance map based on median MS intensities of common proteins. Highly abundant proteins in both datasets are labeled as examples.

(C) Fisher exact test to identify significantly enriched GO terms among the common and urine- and CSF-specific protein groups. Significant and non-redundant GO terms are displayed (FDR <5%).

(D) Correlation of ANCOVA q values of the proteins differentially present in the CSF (HBS) and urine (Columbia) of PD patients compared with controls.

(legend continued on next page)

Correlating the ANCOVA *q* values of all common proteins quantified in CSF (HBS and LCC) and urine (Columbia cohort) of PD patients compared with the controls revealed several proteins regulated in both fluids (Figures 5D and 5E). Among those was osteomodulin (OMD), the level of which was also significantly elevated in the urine of PD patients, with a mean fold change of 1.47 (Figure 5F). Next, we integrated ANCOVA *q* values of LRRK2 G2019S versus LRRK2 WT of all common proteins quantified in the CSF samples of the LCC cohort and the urine samples of either the Columbia or the LCC cohorts (Figures 5G and 5H). This identified cathepsin S to be regulated in a LRRK2 status-dependent manner in both matrices (Figures 5G and 5H). The levels of cathepsin S were higher in the CSF of LRRK2 G2019S carriers compared with the LRRK2 WT carriers, with mean fold changes of 1.31 (*p* value of 0.0619) and 1.65 (*p* value of 0.0001574) in the Columbia and LCC cohorts, respectively (Figures 5I and S2H).

Correlation of CSF proteome profiles with clinical scores indicating disease severity

We next investigated whether any protein level changes in the HBS cohort correlated with the severity of PD pathology as assessed by the Unified Parkinson's Disease Rating Scale (UPDRS) (Figures 6, S3, S4, and S5). The UPDRS scores and protein intensities are listed in Tables S1 and S2. We found 27 proteins to be significantly correlated with the UPDRS scores in idiopathic PD patients (*p* < 0.001; Figure 6A). Proteins showing the highest positive correlation in PD patients included CHST6 (*p* = 1.6E-5 and Pearson correlation *r* = 0.72), MIF (*p* = 1.8E-5 and *r* = 0.67), LYVE1 (*p* = 4.8E-5 and *r* = 0.64), EFNA1 (*p* = 5E-5 and *r* = 0.64), and ADM (*p* = 1.3E-4 and *r* = 0.61; Figure 6A). Proteins showing the highest negative correlation in PD patients included POMGNT1 (*p* = 4.6E-5 and *r* = -0.64), TMEM132A (*p* = 5.7E-5 and *r* = -0.63), ADAM22 (*p* = 8.8E-5 and *r* = -0.62), PAM (*p* = 1.9E-4 and *r* = -0.6), and ST6GAL2 (*p* = 2.7E-5 and *r* = -0.59; Figure 6A). Interestingly, none of these proteins was significantly altered in PD cases compared with controls. However, ADM was the strongest negatively correlated protein with the Mini-Mental State Examination (MMSE) scores (*p* = 9.3E-5 and *r* = -0.43; Figure S3D, a quantitative measure of cognitive impairment. Of note, no individual suffered from significant dementia or cognitive impairment at the time of sample collection (Figure S1D: average scores 28.3 in PD versus 28.4 in controls).

Furthermore, we compared the correlation scores of CSF proteins with those obtained in our previous study in urine. The comparison of Pearson coefficients for the UPDRS III scores (the only score available for urine) revealed 10 proteins (APOF, MAN2B2,

LYVE1, APLP1, ALDOA, IGF2, SUSD5, CD99, S100A7, and VGF) to have coefficients either higher than 0.4 or lower than -0.4 in both matrices (Figure 6B). Among these, VGF exhibited a significantly negative correlation with the UPDRS III scores in both datasets (*p* = 2.8E-3 and 7.6E-3 in urine and CSF, respectively; Figure 6C). Overall, this analysis suggests that disease progression in PD patients, assessed by motor function, affects a similar set of proteins in CSF and urine.

Within the LCC cohort, Montreal Cognitive Assessment (MoCA) scores for all subjects were available (Figure S1I). Interestingly, we found CHIT1 to significantly negatively correlate with MoCA scores in all patients (*p* = 1.6E-5 and *r* = -0.48 in all subjects; Figure S5D), and this correlation was even stronger in PD patients (*p* = 1.9E-9 and *r* = -0.81 in PD patients; Figure S5E). Two proteins strongly positively correlated with MoCA scores, RELN and PENK, were negatively correlated with age in CSF (Figures S5D and S5E). Thus, multiple proteins in the CSF proteome correlate well with disease severity and different clinically important aspects of the disease.

DISCUSSION

Here, we applied a scalable and highly reproducible MS-based proteomics workflow to CSF samples from two independent PD cohorts. To the best of our knowledge, our approach resulted in the deepest single-run CSF proteome acquired by MS to date, with about 1,400 proteins quantified per sample. In addition, our workflow is sensitive, using only 40 μ L of CSF for sample preparation. It does not require depletion of highly abundant proteins such as albumin or any biochemical enrichment, and the amount of purified peptides is sufficient for several MS runs, thus allowing the re-measurement of individual samples if required. Despite the unmatched depth of our dataset, we have not yet identified some of the known and well-studied PD-relevant proteins such as α -synuclein and neurofilament light chain, although both proteins were present in the LCC cohort library, and neurofilament light chain was also present in the HBS cohort library. The high dynamic range of protein abundances in CSF limits the sensitivity of MS-based proteomics compared with antibody-based assays that are frequently used to study these low abundant PD markers. However, the analytical variation of our assay, with a median CV of 19%, was much better than the biological variation, with a CV of median 46%. Although single-analyte antibody-based assays like ELISAs often have even lower technical CVs, we found our workflow well suited to study disease-related biological differences on a proteome-scale.

To reduce systemic biases in the analyzed samples and minimize the effect of pre-analytical variation, we performed a

(E) Correlation of ANCOVA *q* values of the proteins differentially present in the CSF (LCC) and urine (Columbia) of PD patients compared with controls.

(F) OMD protein intensity distribution in the urine of controls and PD patients of the Columbia cohort. Results of unpaired *t* test are shown. The lines indicate mean and SD.

(G) Correlation of ANCOVA *q* values of the proteins differentially present in the CSF (LCC) and urine (Columbia) of LRRK2 G2019S carriers compared with the LRRK2 WT controls.

(H) Correlation of ANCOVA *q* values of the proteins differentially present in the CSF (LCC) and urine (LCC) of LRRK2 G2019S carriers compared with the LRRK2 WT controls.

(I) CTSS protein intensity (log₂) distribution in the urine of LRRK2 G2019S and WT carriers of the LCC cohort. Results of an unpaired *t* test are shown. The lines indicate mean and SD.

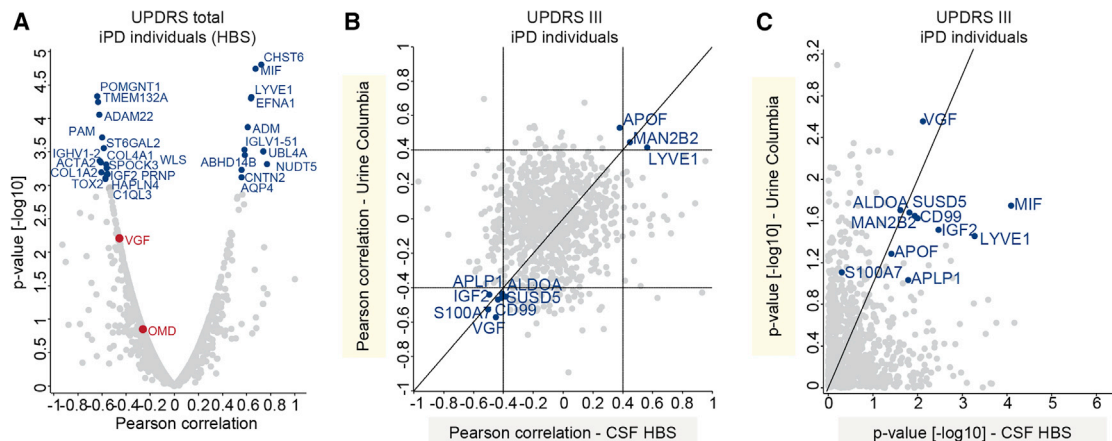


Figure 6. Correlation of CSF and urinary proteomes with UPDRS scores in idiopathic PD (iPD) patients

(A) Correlation analysis of protein intensities in CSF with the UPDRS total scores in iPD patients. Pearson correlation coefficients and $-\log_{10}$ p values are displayed on the x and y axes, respectively. Proteins significantly correlating with UPDRS score (positively or negatively with a p value < 0.001) are labeled. (B) Correlation between Pearson correlation coefficients for correlation of urinary and CSF proteomes with UDPRS III scores. (C) Correlation between $-\log_{10}$ p values for correlation of urinary and CSF proteomes with UDPRS III scores. The proteins with Pearson correlation coefficients > 0.4 and < -0.4 in both datasets are labeled in (B) and (C).

thorough quality assessment of every sample using our previously reported quality marker panels.³⁷ This analysis flagged several samples of both cohorts as contaminated with erythrocyte-specific proteins, and these were excluded from further analysis. Samples with a high degree of contamination were restricted to two of the four study centers, and erythrocyte counts, which are often determined following CSF sample collection, did not correlate with the degree of proteome contamination by erythrocyte-specific proteins. This is presumably because intact erythrocytes, which are typically determined in clinical laboratories, are frequently removed by centrifugation before samples are stored. Contaminations in the proteome are, however, not caused by intact erythrocytes but by hemolysis. We also cannot exclude the possibility that different CSF aliquots were collected from each patient and that standard laboratory tests like erythrocyte counts were performed on only one of these aliquots. For the future, we recommend collecting CSF in bulk first and aliquoting only following thorough mixing. Our findings using untargeted proteomics clearly underline the importance of stringent quality control and study protocols to avoid systemic biases, which may result in seemingly significant regulations in some quality markers that would then be reported as potential biomarker candidates.

ANCOVA analysis considers confounding factors such as age, sex, or LRRK2 status and thus is well suited to stringently assess which protein changes are truly associated with PD status. This is important when analyzing cohorts with slightly imbalanced demographics between the compared groups like increased age of patients compared with controls, as frequently found in PD cohorts. When we employed ANCOVA to identify proteins that were disease-status-dependently regulated in both cohorts, OMD and CD44 stood out in both cohorts. OMD belongs to the small leucine-rich proteoglycans (SLRPs) and is involved in the organization and homeostasis of the extracellular matrix. CD44 is a cell surface glycoprotein involved in cell-adhesion

and cell-cell interactions.^{47,48} Our results are in line with previous studies that identified OMD as a potential blood biomarker for PD⁴⁹ and upregulated in a SNCA transgenic mouse model.⁵⁰ A recent study has also shown the elevated expression of CD44 in the substantia nigra of human PD brains and CD44-mediated anti-inflammatory effects in primary mouse astrocytes.⁵¹ Induced expression of CD44 by α -synuclein in microglia, likely affecting PD pathogenesis by recruiting reactive microglia into the pathological region of the PD brain, was also reported.⁵² The low number of significantly regulated proteins in our study may be explained by the limited number of samples per disease group in each cohort combined with many potential confounders taken into account. Furthermore, patient heterogeneity due to inconsistent criteria for PD diagnosis and patient recruitment likely diminishes the overlap between independent cohorts. A recent meta-analysis also revealed that the overlap between independent studies is rather small and that future studies with larger-sample cohorts are required to identify protein changes with small effect sizes.⁵³

A GO term enrichment analysis revealed that proteins upregulated in PD patients were associated with lysosome-related terms, which agrees with previous data that lysosomal dysregulation is evident in PD patients and involved in disease pathogenesis. In fact, lysosomal enzymes have long been investigated for their potential as biomarkers in the diagnosis of neurodegenerative disorders including PD.^{14,54–58} Changes in enzyme activities or abundances of lysosomal enzymes in CSF have been suggested to mirror the neuropathological changes linked to PD, although the basis of these alterations is not well understood. Levels of several lysosomal proteases including cathepsin D and cathepsin B are increased in CSF and *post mortem* brain tissue of Alzheimer's disease patients.^{59,60} Furthermore, activities of the lysosomal β -galactosidase and β -hexosaminidase as well as cathepsin L in CSF or *post mortem* brain tissue of PD patients are elevated.^{55,57,61–63} Recently, we have shown that the pathogenic

LRRK2-dependent changes of the urinary proteome included dozens of lysosomal proteins that could serve as biomarkers to stratify individuals with pathogenic LRRK2. In line with all these findings, we here identified two lysosomal proteins, cathepsin K and MAN2B1, among the most important features to classify an individual's PD status based on their CSF proteome. CTSS, another cathepsin, was one of the proteins with the highest upregulation in CSF of LRRK2 G2019S carriers compared with the LRRK2 WT controls in the LCC cohort. While cathepsin K has the potential to ameliorate α -synuclein pathology by degrading α -synuclein amyloids,⁶⁴ increased expression of the CTSS gene—together with other genes involved in the antigen processing and presentation pathway and related immune pathways such as HLA-DQA1, HLA-DRA, HLA-DPA1, and HLA-DMB—was reported in idiopathic PD patients in a study in which brain transcriptomic profiling was performed in idiopathic and LRRK2-associated PD.⁶⁵ Moreover, cathepsin S has been shown to regulate the MHC class II antigen presentation process.^{66,67}

Strikingly, several HLA proteins (HLA-DRA, HLA-DRB1, and HLA-DPA) were significantly increased in LRRK2 G2019S carriers compared with controls. The HLA locus is one of the key loci associated with susceptibility for PD.⁶⁸ HLA-DR and HLA-DP are commonly expressed MHC-II molecules on antigen-presenting cells, including microglia in the CNS.⁶⁹ Moreover, increased expression of HLA-DR is a hallmark of activated microglia, which are present in multiple neurodegenerative diseases including PD.^{69–72} In addition, specific HLA-DRB1 variants can bind α -synuclein with high affinity, and genome-wide association studies identified HLA-DRB1 and HLA-DRA alleles to be associated with PD in different populations.^{73–78} A recent study has revealed a positive correlation between LRRK2 and MHC-II levels in PD patients and a negative correlation in healthy controls.⁷⁹ Our data support these findings and suggest a contribution of immunity and MHC-II molecules to the pathogenesis of familial PD.

Comparing the analyzed CSF proteomes with our urinary proteome profiles in PD patients revealed 1,080 common proteins. This large overlap and the clear correlation of the corresponding protein intensities (Pearson r 0.49) are remarkable and presumably due to the shared blood plasma origin of both CSF and urine. Interestingly, OMD was upregulated in PD patients in both CSF and urine, suggesting that the pathophysiology of PD affecting the OMD pathway is not restricted to the brain. Furthermore, CTSS was upregulated in LRRK2 G2019S carriers in both body fluids. LRRK2 is known to be ubiquitously expressed in many organs, and the observed dysregulation of lysosomal enzymes including CTSS may be due to the hyperactivity of the mutated kinase. Moreover, we found VGF to be important in classifying PD patients from healthy controls by ML. Strikingly, we also reported this in our previous urinary-proteome-profiling study where VGF was the most important feature for classifying LRRK2 PD patients from NMCs, and its levels were strongly decreased in PD patients.²⁵ Consistently, VGF was also one of the handful of proteins found to be statistically significantly under-expressed in CSF of two independent PD cohorts.²⁸ Its negative correlation with the UPDRS-III scores of PD patients in both CSF and urine could indicate a protective role of this growth factor for

motor function. Despite these interesting observations, the number of samples analyzed in our cohorts is still low for ML approaches, and larger cohorts are needed to further improve the accuracy and generalizability of the extracted models. Nevertheless, our data show that CSF and urine have large overlaps in their proteomes and that similar PD- and LRRK2-associated proteome changes can be identified in both body fluids. Presumably, processes in both the CNS and distal organs contribute to the commonly observed regulations in the biofluid proteomes.

We also identified several proteins including CHI3L1, FCGR3A, NCR3LG1, and ZP2 ($p < 0.01$ in all analyses and both cohorts) that correlate well with the subjects' age at sample collection (Figures S3A–S3C and S5A–B). CHI3L1 is a secreted glycoprotein that serves as a migration factor for astrocytes and a marker of glial inflammation.^{80,81} Interestingly, CHI3L was reported to be upregulated in the CSF of Alzheimer's disease patients and was therefore suggested as a biomarker for this disease. Its expression levels in various regions of the brain were shown to be correlated with age.^{24,82–84} Our data demonstrate that age correlates well with multiple proteins in the CSF and that it is thus important to consider these confounders in statistical analyses to avoid a biased interpretation.

In line with our finding of CHIT1 being negatively correlated with the MoCA scores, chitinases including CHIT1 have emerged as biomarkers in neurological disorders including amyotrophic lateral sclerosis (ALS), as their levels correlate with disease activity and progression, likely reflecting microglia/macrophage activation.^{85–87} In addition, levels of proteins such as CHST6 correlate well with disease progression, as measured by the UPDRS, especially UPDRS III (Figures 6A and S4). The sulfotransferase CHST6 plays an important role in keratan sulfonation, and mutations in the corresponding gene cause macular corneal dystrophy.⁸⁸ Keratan sulfate, which is a type of sulfated glycosaminoglycan (GAG), is abundant in the brain, where it fulfills a multitude of functions.⁸⁹ CHST6 expression is increased in the brains of AD patients, and its deficiency in mouse models mitigates AD pathology.⁹⁰ Interestingly, the sulfation state of GAGs affects α -synuclein aggregation by regulating lysosomal degradation.⁹¹ Together, our data corroborate findings that sulfated GAGs like keratan sulfate can affect the pathophysiology of neurodegenerative diseases like PD. We also identified a putative correlation between UPDRS II scores and reduced LRPPRC expression, whose expression is also reduced in the blood of PD patients.⁹² Mutations in *LRPPRC* cause the early-onset progressive mitochondrial neurodegenerative disorder French-Canadian-type Leigh syndrome, characterized by defects in oxidative phosphorylation reminiscent to those found in prodromal PD.⁹³

In conclusion, we have applied a highly reproducible and scalable MS-based proteomics workflow to perform proteome profiling of CSF in PD patients. We observed interesting proteome changes in PD patients and identified biomarker signatures that are specific to LRRK2 G2019S carriers. Further studies analyzing larger cohorts of patients will be required to confirm our findings and extend the panels of potential biomarkers. In a next step, clinical and targeted assays need to be developed to

validate the biomarkers,¹² followed by a test that can be used in clinical routine to enable early disease detection and patient stratification.

Limitations of the study

The number of samples per disease group in each cohort in combination with patient heterogeneity and biological variation of the CSF proteome limits the power of the statistical analyses performed. Additional studies are needed to validate the identified biomarker candidates from this study using an independent cohort and potentially an orthogonal technology. Furthermore, clinical assays to measure these markers reliably in a routine fashion are still to be developed.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
 - Lead contact
 - Materials availability
 - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
 - Sample preparation
 - LC-MS/MS analysis
 - Mass spectrometry data processing
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.xcrm.2022.100661>.

ACKNOWLEDGMENTS

We thank all members of the Proteomics and Signal Transduction Group at the Max Planck Institute of Biochemistry and the Clinical Proteomics Group at the NNF Center for Protein Research for help and discussions and in particular Jakob Bader, Philipp Geyer, Igor Paron, Johannes Mueller-Reif and Niels Skotte. We further thank Dario Alessi and Suzanne Pfeffer and their group members and employees of the Michael J. Fox Foundation for Parkinson's research for helpful discussions. Biospecimens used in the analyses presented in this article were obtained from the MJFF-sponsored *LRRK2* Cohort Consortium (LCC). For up-to-date information on the study: <https://www.michaeljfox.org/news/lrrk2-cohort-consortium>. The Harvard Biomarkers Study ("HBS"); <https://www.bwhparkinsoncenter.org> is a collaborative initiative of Brigham and Women's Hospital and Massachusetts General Hospital, co-directed by Dr. Clemens Scherzer and Dr. Bradley T. Hyman. The HBS Investigators have not participated in reviewing the current manuscript. Figure 1A and graphical abstract were created with <http://biorender.com>. The work carried out in this project was supported by the Max Planck Society for the Advancement of Science and the Michael J. Fox Foundation (MJFF) with the grant ID 15101. C.R.S.'s work was supported by NIH grants NINDS/NIA R01NS115144, U01NS095736, U01NS100603, the MJFF and the American Parkinson Disease Association Center for Advanced Parkinson Research. HBS was made possible by generous support from the Harvard NeuroDiscovery Center, with additional contributions from the MJFF, NINDS U01NS082157, U01NS100603, and the Massachusetts Alzheimer's Disease Research Center NIA P50AG005134.

AUTHOR CONTRIBUTIONS

O.K. and S.V.W. designed the experiments, performed, analyzed and interpreted all data; D.T.V. helped with analyzing the data; Y.K., I.T., A.M.W., and C.R.S. were responsible for sample collection and their distribution and reviewed and edited the manuscript; S.P., K.M., and C.R.S. helped with interpretation of the results and reviewed and edited the manuscript; M.M. supervised the project and interpreted results; O.K., S.V.W., and M.M. wrote the manuscript.

DECLARATION OF INTERESTS

The authors declare no conflicts of interest related to this work.

Received: July 2, 2021

Revised: October 29, 2021

Accepted: May 23, 2022

Published: June 21, 2022

REFERENCES

1. Reeve, A., Simcox, E., and Turnbull, D. (2014). Ageing and Parkinson's disease: why is advancing age the biggest risk factor? *Ageing Res. Rev.* *14*, 19–30. <https://doi.org/10.1016/j.arr.2014.01.004>.
2. Tysnes, O.B., and Storstein, A. (2017). Epidemiology of Parkinson's disease. *J. Neural. Transm.* *124*, 901–905. <https://doi.org/10.1007/s00702-017-1686-y>.
3. de Lau, L.M., and Breteler, M.M. (2006). Epidemiology of Parkinson's disease. *Lancet Neurol.* *5*, 525–535. [https://doi.org/10.1016/s1474-4422\(06\)70471-9](https://doi.org/10.1016/s1474-4422(06)70471-9).
4. Klein, C., and Westenberger, A. (2012). Genetics of Parkinson's disease. *Cold Spring Harb. Perspect. Med.* *2*, a008888. <https://doi.org/10.1101/cshperspect.a008888>.
5. Schneider, S.A., and Alcalay, R.N. (2020). Precision medicine in Parkinson's disease: emerging treatments for genetic Parkinson's disease. *J. Neurol.* *267*, 860–869. <https://doi.org/10.1007/s00415-020-09705-7>.
6. West, A.B., Moore, D.J., Biskup, S., Bugayenko, A., Smith, W.W., Ross, C.A., Dawson, V.L., and Dawson, T.M. (2005). Parkinson's disease-associated mutations in leucine-rich repeat kinase 2 augment kinase activity. *Proc. Natl. Acad. Sci. U S A* *102*, 16842–16847. <https://doi.org/10.1073/pnas.0507360102>.
7. West, A.B. (2015). Ten years and counting: moving leucine-rich repeat kinase 2 inhibitors to the clinic. *Mov. Disord.* *30*, 180–189. <https://doi.org/10.1002/mds.26075>.
8. Taymans, J.M., and Greggio, E. (2016). LRRK2 kinase inhibition as a therapeutic strategy for Parkinson's disease, where do we stand? *Curr. Neuropharmacol.* *14*, 214–225. <https://doi.org/10.2174/1570159x13666151030102847>.
9. Tolosa, E., Vila, M., Klein, C., and Rascol, O. (2020). LRRK2 in Parkinson disease: challenges of clinical trials. *Nat. Rev. Neurol.* *16*, 97–107. <https://doi.org/10.1038/s41582-019-0301-2>.
10. Steger, M., Tonelli, F., Ito, G., Davies, P., Trost, M., Vetter, M., Wachter, S., Lorentzen, E., Duddy, G., Wilson, S., et al. (2016). Phosphoproteomics reveals that Parkinson's disease kinase LRRK2 regulates a subset of Rab GTPases. *Elife* *5*, e12813. <https://doi.org/10.7554/eLife.12813>.
11. Steger, M., Diez, F., Dhekne, H.S., Lis, P., Nirujogi, R.S., Karayel, O., Tonelli, F., Martinez, T.N., Lorentzen, E., Pfeffer, S.R., et al. (2017). Systematic proteomic analysis of LRRK2-mediated Rab GTPase phosphorylation establishes a connection to ciliogenesis. *Elife* *6*, e31012. <https://doi.org/10.7554/eLife.31012>.
12. Karayel, Ö., Tonelli, F., Virreira Winter, S., Geyer, P.E., Fan, Y., Sammler, E.M., Alessi, D.R., Steger, M., and Mann, M. (2020). Accurate MS-based Rab10 phosphorylation stoichiometry determination as readout for

- LRRK2 activity in Parkinson's disease. *Mol. Cell. Proteomics* 19, 1546–1560. <https://doi.org/10.1074/mcp.RA120.002055>.
13. Alessi, D.R., and Sammler, E. (2018). LRRK2 kinase in Parkinson's disease. *Science* 360, 36–37. <https://doi.org/10.1126/science.aar5683>.
 14. Gaetani, L., Paolini Paoletti, F., Bellomo, G., Mancini, A., Simoni, S., Di Filippo, M., and Parnetti, L. (2020). CSF and blood biomarkers in neuroinflammatory and neurodegenerative diseases: implications for treatment. *Trends Pharmacol. Sci.* 41, 1023–1037. <https://doi.org/10.1016/j.tips.2020.09.011>.
 15. Maass, F., Schulz, I., Lingor, P., Mollenhauer, B., and Bähr, M. (2019). Cerebrospinal fluid biomarker for Parkinson's disease: an overview. *Mol. Cell. Neurosci.* 97, 60–66. <https://doi.org/10.1016/j.mcn.2018.12.005>.
 16. Geyer, P.E., Kulak, N., Pichler, G., Holdt, L., Teupser, D., and Mann, M. (2016). Plasma proteome profiling to assess human health and disease. *Cell Syst.* 2, 185–195. <https://doi.org/10.1016/j.cels.2016.02.015>.
 17. Geyer, P.E., Holdt, L.M., Teupser, D., and Mann, M. (2017). Revisiting biomarker discovery by plasma proteomics. *Mol. Syst. Biol.* 13, 942. <https://doi.org/10.15252/msb.20156297>.
 18. Holewinski, R.J., Jin, Z., Powell, M.J., Maust, M.D., and Van Eyk, J.E. (2013). A fast and reproducible method for albumin isolation and depletion from serum and cerebrospinal fluid. *Proteomics* 13, 743–750. <https://doi.org/10.1002/pmic.201200192>.
 19. Bruderer, R., Bernhardt, O.M., Gandhi, T., Xuan, Y., Sondermann, J., Schmidt, M., Gomez-Varela, D., and Reiter, L. (2017). Optimization of experimental parameters in data-independent mass spectrometry significantly increases depth and reproducibility of results. *Mol. Cell. Proteomics* 16, 2296–2309. <https://doi.org/10.1074/mcp.RA117.000314>.
 20. Bruderer, R., Sondermann, J., Tsou, C.C., Barrantes-Freer, A., Stadelmann, C., Nesvizhskii, A.I., Schmidt, M., Reiter, L., and Gomez-Varela, D. (2017). New targeted approaches for the quantification of data-independent acquisition mass spectrometry. *Proteomics* 17, 1700021. <https://doi.org/10.1002/pmic.201700021>.
 21. Gillet, L.C., Navarro, P., Tate, S., Rost, H., Selevsek, N., Reiter, L., Bonner, R., and Aebersold, R. (2012). Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Mol. Cell. Proteomics* 11. <https://doi.org/10.1074/mcp.o111.016717>.
 22. Kelstrup, C.D., Bekker-Jensen, D.B., Arrey, T.N., Hogrebe, A., Harder, A., and Olsen, J.V. (2018). Performance evaluation of the Q exactive HF-X for shotgun proteomics. *J. Proteome Res.* 17, 727–738. <https://doi.org/10.1021/acs.jproteome.7b00602>.
 23. Demichev, V., Messner, C.B., Vernardis, S.I., Lilley, K.S., and Ralsler, M. (2020). DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. *Nat. Methods* 17, 41–44. <https://doi.org/10.1038/s41592-019-0638-x>.
 24. Bader, J.M., Geyer, P.E., Muller, J.B., Strauss, M.T., Koch, M., Leyboldt, F., Koertvelyessy, P., Bittner, D., Schipke, C.G., Incesoy, E.I., et al. (2020). Proteome profiling in cerebrospinal fluid reveals novel biomarkers of Alzheimer's disease. *Mol. Syst. Biol.* 16, e9356. <https://doi.org/10.15252/msb.20199356>.
 25. Virreira Winter, S., Karayel, O., Strauss, M.T., Padmanabhan, S., Surface, M., Merchant, K., Alcalay, R.N., and Mann, M. (2021). Urinary proteome profiling for stratifying patients with familial Parkinson's disease. *EMBO Mol. Med.* 13, e13257. <https://doi.org/10.15252/emmm.202013257>.
 26. Wewer Albrechtsen, N.J., Geyer, P.E., Doll, S., Treit, P.V., Bojsen-Moller, K.N., Martinussen, C., Jorgensen, N.B., Torekov, S.S., Meier, F., Niu, L., et al. (2018). Plasma proteome profiling reveals dynamics of inflammatory and lipid homeostasis markers after roux-en-Y gastric bypass surgery. *Cell Syst.* 7, 601–612.e3. <https://doi.org/10.1016/j.cels.2018.10.012>.
 27. Niu, L., Geyer, P.E., Wewer Albrechtsen, N.J., Gluud, L.L., Santos, A., Doll, S., Treit, P.V., Holst, J.J., Knop, F.K., Vilsboll, T., et al. (2019). Plasma proteome profiling discovers novel proteins associated with non-alcoholic fatty liver disease. *Mol. Syst. Biol.* 15, e8793. <https://doi.org/10.15252/msb.20188793>.
 28. Rotunno, M.S., Lane, M., Zhang, W., Wolf, P., Oliva, P., Viel, C., Wills, A.M., Alcalay, R.N., Scherzer, C.R., Shihabuddin, L.S., et al. (2020). Cerebrospinal fluid proteomics implicates the granin family in Parkinson's disease. *Sci. Rep.* 10, 2479. <https://doi.org/10.1038/s41598-020-59414-4>.
 29. Higginbotham, L., Ping, L., Dammer, E.B., Duong, D.M., Zhou, M., Gearing, M., Hurst, C., Glass, J.D., Factor, S.A., Johnson, E.C.B., et al. (2020). Integrated proteomics reveals brain-based cerebrospinal fluid biomarkers in asymptomatic and symptomatic Alzheimer's disease. *Sci. Adv.* 6, eaaz9360. <https://doi.org/10.1126/sciadv.aaz9360>.
 30. Ding, H., Sarokhan, A.K., Roderick, S.S., Bakshi, R., Maher, N.E., Ashourian, P., Kan, C.G., Chang, S., Santarlasci, A., Swords, K.E., et al. (2011). Association of SNCA with Parkinson: replication in the Harvard NeuroDiscovery center biomarker study. *Mov. Disord.* 26, 2283–2286. <https://doi.org/10.1002/mds.23934>.
 31. Mohammadi, D. (2013). The Harvard Biomarker Study's big plan. *Lancet Neurol.* 12, 739–740. [https://doi.org/10.1016/s1474-4422\(13\)70155-8](https://doi.org/10.1016/s1474-4422(13)70155-8).
 32. Liu, G., Locascio, J.J., Corvol, J.C., Boot, B., Liao, Z., Page, K., Franco, D., Burke, K., Jansen, I.E., Trisini-Lipsanopoulos, A., et al. (2017). Prediction of cognition in Parkinson's disease with a clinical-genetic score: a longitudinal analysis of nine cohorts. *Lancet Neurol.* 16, 620–629. [https://doi.org/10.1016/s1474-4422\(17\)30122-9](https://doi.org/10.1016/s1474-4422(17)30122-9).
 33. Liu, G., Boot, B., Locascio, J.J., Jansen, I.E., Winder-Rhodes, S., Eberly, S., Elbaz, A., Brice, A., Ravina, B., van Hilten, J.J., et al. (2016). Specifically neuropathic Gaucher's mutations accelerate cognitive decline in Parkinson's. *Ann. Neurol.* 80, 674–685. <https://doi.org/10.1002/ana.24781>.
 34. Ludwig, C., Gillet, L., Rosenberger, G., Amon, S., Collins, B.C., and Aebersold, R. (2018). Data-independent acquisition-based SWATH-MS for quantitative proteomics: a tutorial. *Mol. Syst. Biol.* 14, e8126. <https://doi.org/10.15252/msb.20178126>.
 35. Barkovits, K., Linden, A., Galozzi, S., Schilde, L., Pacharra, S., Mollenhauer, B., Stoepel, N., Steinbach, S., May, C., Uszkoreit, J., et al. (2018). Characterization of cerebrospinal fluid via data-independent acquisition mass spectrometry. *J. Proteome Res.* 17, 3418–3430. <https://doi.org/10.1021/acs.jproteome.8b00308>.
 36. Thul, P.J., Akesson, L., Wiking, M., Mahdessian, D., Geladaki, A., Ait Blal, H., Alm, T., Asplund, A., Bjork, L., Breckels, L.M., et al. (2017). A subcellular map of the human proteome. *Science* 356, eaal3321. <https://doi.org/10.1126/science.aal3321>.
 37. Geyer, P.E., Voytik, E., Treit, P.V., Doll, S., Kleinhempel, A., Niu, L., Muller, J.B., Buchholtz, M., Bader, J.M., Teupser, D., et al. (2019). Plasma Proteome Profiling to detect and avoid sample-related biases in biomarker studies. *EMBO Mol. Med.* 11, e10427. <https://doi.org/10.15252/emmm.201910427>.
 38. Cox, J., and Mann, M. (2012). 1D and 2D annotation enrichment: a statistical method integrating quantitative proteomics with complementary high-throughput data. *BMC Bioinf.* 13, S12. <https://doi.org/10.1186/1471-2105-13-s16-s12>.
 39. Navarro-Romero, A., Montpeyó, M., and Martínez-Vicente, M. (2020). The emerging role of the lysosome in Parkinson's disease. *Cells* 9, 2399. <https://doi.org/10.3390/cells9112399>.
 40. Youn, J., Lee, S.B., Lee, H.S., Yang, H.O., Park, J., Kim, J.S., Oh, E., Park, S., and Jang, W. (2018). Cerebrospinal fluid levels of autophagy-related proteins represent potentially novel biomarkers of early-stage Parkinson's disease. *Sci. Rep.* 8, 16866. <https://doi.org/10.1038/s41598-018-35376-6>.
 41. Torun, F.M., Winter, S.V., Doll, S., Riese, F.M., Vorobyev, A., Mueller-Reif, J.B., Geyer, P.E., and Strauss, M.T. (2021). Transparent exploration of machine learning for biomarker discovery from proteomics and omics data. Preprint at bioRxiv. <https://doi.org/10.1101/2021.03.05.434053>.

42. Lawton, N.F., and MacDermot, J. (1980). Abnormal regulation of prolactin release in idiopathic Parkinson's disease. *J. Neurol. Neurosurg. Psychiatry* 43, 1012–1015. <https://doi.org/10.1136/jnnp.43.11.1012>.
43. Stojkowska, I., Wagner, B.M., and Morrison, B.E. (2015). Parkinson's disease and enhanced inflammatory response. *Exp. Biol. Med.* 240, 1387–1395. <https://doi.org/10.1177/1535370215576313>.
44. Caggiu, E., Arru, G., Hosseini, S., Niegowska, M., Sechi, G., Zarbo, I.R., and Sechi, L.A. (2019). Inflammation, infectious triggers, and Parkinson's disease. *Front. Neurol.* 10, 122. <https://doi.org/10.3389/fneur.2019.00122>.
45. De Pasquale, V., Moles, A., and Pavone, L.M. (2020). Cathepsins in the pathophysiology of mucopolysaccharidoses: new perspectives for therapy. *Cells*, 979. <https://doi.org/10.3390/cells9040979>.
46. Pellegrini, L., Hauser, D.N., Li, Y., Mamais, A., Beilina, A., Kumaran, R., Wetzel, A., Nixon-Abell, J., Heaton, G., Rudenko, I., et al. (2018). Proteomic analysis reveals co-ordinated alterations in protein synthesis and degradation pathways in LRRK2 knockout mice. *Hum. Mol. Genet.* 27, 3257–3271. <https://doi.org/10.1093/hmg/ddy232>.
47. Senbanjo, L.T., and Chellaiah, M.A. (2017). CD44: a multifunctional cell surface adhesion receptor is a regulator of progression and metastasis of cancer cells. *Front. Cell Dev. Biol.* 5, 18. <https://doi.org/10.3389/fcell.2017.00018>.
48. Iczkowski, K.A. (2010). Cell adhesion molecule CD44: its functional roles in prostate cancer. *Am. J. Tourism Res.* 3, 1–7.
49. Posavi, M., Diaz-Ortiz, M., Liu, B., Swanson, C.R., Skrinak, R.T., Hernandez-Con, P., Amado, D.A., Fullard, M., Rick, J., Siderowf, A., et al. (2019). Characterization of Parkinson's disease using blood-based biomarkers: a multicohort proteomic analysis. *PLoS Med.* 16, e1002931. <https://doi.org/10.1371/journal.pmed.1002931>.
50. Hendrickx, D.M., Garcia, P., Ashrafi, A., Sciortino, A., Schmit, K.J., Kollmus, H., Nicot, N., Kaoma, T., Vallar, L., Buttini, M., and Glaab, E. (2021). A new synuclein-transgenic mouse model for early Parkinson's reveals molecular features of preclinical disease. *Mol. Neurobiol.* 58, 576–602. <https://doi.org/10.1007/s12035-020-02085-z>.
51. Neal, M.L., Boyle, A.M., Budge, K.M., Safadi, F.F., and Richardson, J.R. (2018). The glycoprotein GPNMB attenuates astrocyte inflammatory responses through the CD44 receptor. *J. Neuroinflammation* 15, 73. <https://doi.org/10.1186/s12974-018-1100-1>.
52. Kim, S., Cho, S.H., Kim, K.Y., Shin, K.Y., Kim, H.S., Park, C.H., Chang, K.A., Lee, S.H., Cho, D., and Suh, Y.H. (2009). α -Synuclein induces migration of BV-2 microglial cells by up-regulation of CD44 and MT1-MMP. *J. Neurochem.* 109, 1483–1496. <https://doi.org/10.1111/j.1471-4159.2009.06075.x>.
53. Halbgebauer, S., Öckl, P., Wirth, K., Steinacker, P., and Otto, M. (2016). Protein biomarkers in Parkinson's disease: focus on cerebrospinal fluid markers and synaptic proteins. *Mov. Disord.* 31, 848–860. <https://doi.org/10.1002/mds.26635>.
54. Parnetti, L., Chiasserini, D., Persichetti, E., Eusebi, P., Varghese, S., Qureshi, M.M., Dardis, A., Deganuto, M., Carlo, C., Castrioto, A., et al. (2014). Cerebrospinal fluid lysosomal enzymes and alpha-synuclein in Parkinson's disease. *Mov. Disord.* 29, 1019–1027. <https://doi.org/10.1002/mds.25772>.
55. van Dijk, K.D., Persichetti, E., Chiasserini, D., Eusebi, P., Beccari, T., Calabresi, P., Berendse, H.W., Parnetti, L., and van de Berg, W.D.J. (2013). Changes in endolysosomal enzyme activities in cerebrospinal fluid of patients with Parkinson's disease. *Mov. Disord.* 28, 747–754. <https://doi.org/10.1002/mds.25495>.
56. Parnetti, L., Balducci, C., Pierguidi, L., De Carlo, C., Peducci, M., D'Amore, C., Padiglioni, C., Mastrocola, S., Persichetti, E., Paciotti, S., et al. (2009). Cerebrospinal fluid β -glucocerebrosidase activity is reduced in Dementia with Lewy Bodies. *Neurobiol. Dis.* 34, 484–486. <https://doi.org/10.1016/j.nbd.2009.03.002>.
57. Persichetti, E., Chiasserini, D., Parnetti, L., Eusebi, P., Paciotti, S., De Carlo, C., Codini, M., Tambasco, N., Rossi, A., Agnaf, O.M.E., et al. (2014). Factors influencing the measurement of lysosomal enzymes activity in human cerebrospinal fluid. *PLoS One* 9, e101453. <https://doi.org/10.1371/journal.pone.0101453>.
58. Tasegian, A., Paciotti, S., Ceccarini, M.R., Codini, M., Moors, T., Chiasserini, D., Albi, E., Winchester, B., van de Berg, W.D., Parnetti, L., and Beccari, T. (2017). Origin of α -mannosidase activity in CSF. *Int. J. Biochem. Cell Biol.* 87, 34–37. <https://doi.org/10.1016/j.biocel.2017.03.016>.
59. Schwagerl, A.L., Mohan, P.S., Cataldo, A.M., Vonsattel, J.P., Kowall, N.W., and Nixon, R.A. (2002). Elevated levels of the endosomal-lysosomal proteinase cathepsin D in cerebrospinal fluid in Alzheimer disease. *J. Neurochem.* 64, 443–446. <https://doi.org/10.1046/j.1471-4159.1995.64010443.x>.
60. Cataldo, A.M., Barnett, J.L., Pieroni, C., and Nixon, R.A. (1997). Increased neuronal endocytosis and protease delivery to early endosomes in sporadic Alzheimer's disease: neuropathologic evidence for a mechanism of increased beta-amyloidogenesis. *J. Neurosci.* 17, 6142–6151. <https://doi.org/10.1523/jneurosci.17-16-06142.1997>.
61. McNeill, A., Magalhaes, J., Shen, C., Chau, K.Y., Hughes, D., Mehta, A., Foltynie, T., Cooper, J.M., Abramov, A.Y., Gegg, M., and Schapira, A.H. (2014). Amroxol improves lysosomal biochemistry in glucocerebrosidase mutation-linked Parkinson disease cells. *Brain* 137, 1481–1495. <https://doi.org/10.1093/brain/awu020>.
62. Li, L., Wang, X., Fei, X., Xia, L., Qin, Z., and Liang, Z. (2011). Parkinson's disease involves autophagy and abnormal distribution of cathepsin L. *Neurosci. Lett.* 489, 62–67. <https://doi.org/10.1016/j.neulet.2010.11.068>.
63. Chiasserini, D., Paciotti, S., Eusebi, P., Persichetti, E., Tasegian, A., Kurzawa-Akanbi, M., Chinnery, P.F., Morris, C.M., Calabresi, P., Parnetti, L., and Beccari, T. (2015). Selective loss of glucocerebrosidase activity in sporadic Parkinson's disease and dementia with Lewy bodies. *Mol. Neurodegener.* 10, 15. <https://doi.org/10.1186/s13024-015-0010-2>.
64. McGlinchey, R.P., Lacy, S.M., Walker, R.L., 3rd, and Lee, J.C. (2020). Cathepsin K is a potent disaggregase of α -synuclein fibrils. *Biochem. Biophys. Res. Commun.* 529, 1106–1111. <https://doi.org/10.1016/j.bbrc.2020.06.155>.
65. Botta-Orfila, T., Sanchez-Pla, A., Fernandez, M., Carmona, F., Ezquerro, M., and Tolosa, E. (2012). Brain transcriptomic profiling in idiopathic and LRRK2-associated Parkinson's disease. *Brain Res.* 1466, 152–157. <https://doi.org/10.1016/j.brainres.2012.05.036>.
66. Riese, R.J., Wolf, P.R., Bromme, D., Natkin, L.R., Villadangos, J.A., Ploegh, H.L., and Chapman, H.A. (1996). Essential role for cathepsin S in MHC class II-associated invariant chain processing and peptide loading. *Immunity* 4, 357–366. [https://doi.org/10.1016/s1074-7613\(00\)80249-6](https://doi.org/10.1016/s1074-7613(00)80249-6).
67. Riese, R.J., Mitchell, R.N., Villadangos, J.A., Shi, G.P., Palmer, J.T., Karp, E.R., De Sanctis, G.T., Ploegh, H.L., and Chapman, H.A. (1998). Cathepsin S activity regulates antigen presentation and immunity. *J. Clin. Invest.* 101, 2351–2363. <https://doi.org/10.1172/jci1158>.
68. Nalls, M.A., Blauwendraat, C., Vallergera, C.L., Heilbron, K., Bandres-Ciga, S., Chang, D., Tan, M., Kia, D.A., Noyce, A.J., Xue, A., et al. (2019). Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet Neurol.* 18, 1091–1102. [https://doi.org/10.1016/s1474-4422\(19\)30320-5](https://doi.org/10.1016/s1474-4422(19)30320-5).
69. McGeer, P.L., Itagaki, S., Boyes, B.E., and McGeer, E.G. (1988). Reactive microglia are positive for HLA-DR in the substantia nigra of Parkinson's and Alzheimer's disease brains. *Neurology* 38, 1285–1291. <https://doi.org/10.1212/wnl.38.8.1285>.
70. Walker, D.G., and Lue, L.F. (2015). Immune phenotypes of microglia in human neurodegenerative disease: challenges to detecting microglial polarization in human brains. *Alzheimers Res. Ther.* 7, 56. <https://doi.org/10.1186/s13195-015-0139-9>.

71. McGeer, P.L., Itagaki, S., and McGeer, E.G. (1988). Expression of the histocompatibility glycoprotein HLA-DR in neurological disease. *Acta Neuropathol.* 76, 550–557. <https://doi.org/10.1007/bf00689592>.
72. McGeer, E.G., and McGeer, P.L. (1997). The role of the immune system in neurodegenerative disorders. *Mov. Disord.* 12, 855–858. <https://doi.org/10.1002/mds.870120604>.
73. Wissemann, W.T., Hill-Burns, E., Zabetian, C., Factor, S., Patsopoulos, N., Hoglund, B., Holcomb, C., Donahue, R., Thomson, G., Erlich, H., and Payami, H. (2013). Association of Parkinson disease with structural and regulatory variants in the HLA region. *Am. J. Hum. Genet.* 93, 984–993. <https://doi.org/10.1016/j.ajhg.2013.10.009>.
74. Ran, C., Willows, T., Sydow, O., Johansson, A., Soderkvist, P., Dizdar, N., Ahmadi, A., Olson, L., and Belin, A.C. (2013). The HLA-DRA variation rs3129882 is not associated with Parkinson's disease in Sweden. *Park. Relat. Disord.* 19, 701–702. <https://doi.org/10.1016/j.parkreldis.2013.03.001>.
75. Hamza, T.H., Zabetian, C.P., Tenesa, A., Laederach, A., Montimurro, J., Yearout, D., Kay, D.M., Doheny, K.F., Paschall, J., Pugh, E., et al. (2010). Common genetic variation in the HLA region is associated with late-onset sporadic Parkinson's disease. *Nat. Genet.* 42, 781–785. <https://doi.org/10.1038/ng.642>.
76. Guo, Y., Deng, X., Zheng, W., Xu, H., Song, Z., Liang, H., Lei, J., Jiang, X., Luo, Z., and Deng, H. (2011). HLA rs3129882 variant in Chinese Han patients with late-onset sporadic Parkinson disease. *Neurosci. Lett.* 507, 185–187. <https://doi.org/10.1016/j.neulet.2011.05.245>.
77. Chiang, H.L., Lee-Chen, G.J., Chen, C.M., Chen, Y.C., Lee, C.M., Liao, M.H., and Wu, Y.R. (2012). Genetic analysis of HLA-DRA region variation in Taiwanese Parkinson's disease. *Park. Relat. Disord.* 18, 391–393. <https://doi.org/10.1016/j.parkreldis.2011.12.014>.
78. Kannarkat, G.T., Cook, D.A., Lee, J.K., Chang, J., Chung, J., Sandy, E., Paul, K.C., Ritz, B., Bronstein, J., Factor, S.A., et al. (2015). Common genetic variant association with altered HLA expression, synergy with pyrethroid exposure, and risk for Parkinson's disease: an observational and case-control study. *NPJ Parkinsons Dis.* 1, 15002. <https://doi.org/10.1038/nnpjparkd.2015.2>.
79. Cook, D.A., Kannarkat, G.T., Cintron, A.F., Butkovich, L.M., Fraser, K.B., Chang, J., Grigoryan, N., Factor, S.A., West, A.B., Boss, J.M., and Tansy, M.G. (2017). LRRK2 levels in immune cells are increased in Parkinson's disease. *NPJ Parkinsons Dis.* 3, 11. <https://doi.org/10.1038/s41531-017-0010-8>.
80. Bonneh-Barkay, D., Bissel, S.J., Wang, G., Fish, K.N., Nicholl, G.C., Darko, S.W., Medina-Flores, R., Murphey-Corb, M., Rajakumar, P.A., Nyaundi, J., et al. (2008). YKL-40, a marker of simian immunodeficiency virus encephalitis, modulates the biological activity of basic fibroblast growth factor. *Am. J. Pathol.* 173, 130–143. <https://doi.org/10.2353/ajpath.2008.080045>.
81. Singh, S.K., Bhardwaj, R., Wilczynska, K.M., Dumur, C.I., and Kordula, T. (2011). A complex of nuclear factor I-X3 and STAT3 regulates astrocyte and glioma migration through the secreted glycoprotein YKL-40. *J. Biol. Chem.* 286, 39893–39903. <https://doi.org/10.1074/jbc.M111.257451>.
82. Muszyński, P., Groblewska, M., Kulczyńska-Przybyk, A., Kutakowska, A., and Mroczko, B. (2017). YKL-40 as a potential biomarker and a possible target in therapeutic strategies of Alzheimer's disease. *Curr. Neuropharmacol.* 15, 906–917. <https://doi.org/10.2174/1570159x15666170208124324>.
83. Sanfilippo, C., Castrogiovanni, P., Imbesi, R., Kazakowa, M., Musumeci, G., Blennow, K., Zetterberg, H., and Di Rosa, M. (2019). Sex difference in CHI3L1 expression levels in human brain aging and in Alzheimer's disease. *Brain Res.* 1720, 146305. <https://doi.org/10.1016/j.brainres.2019.146305>.
84. Craig-Schapiro, R., Perrin, R.J., Roe, C.M., Xiong, C., Carter, D., Cairns, N.J., Mintun, M.A., Peskind, E.R., Li, G., Galasko, D.R., et al. (2010). YKL-40: a novel prognostic fluid biomarker for preclinical Alzheimer's disease. *Biol. Psychiatr.* 68, 903–912. <https://doi.org/10.1016/j.biopsych.2010.08.025>.
85. Pinteac, R., Montalban, X., and Comabella, M. (2021). Chitinases and chitinase-like proteins as biomarkers in neurologic disorders. *Neurol. Neuroimmunol. Neuroinflamm.* 8, e921. <https://doi.org/10.1212/nxi.0000000000000921>.
86. Steinacker, P., Verde, F., Fang, L., Feneberg, E., Oeckl, P., Roeber, S., Anderl-Straub, S., Danek, A., Diehl-Schmid, J., Fassbender, K., et al. (2018). Chitotriosidase (CHIT1) is increased in microglia and macrophages in spinal cord of amyotrophic lateral sclerosis and cerebrospinal fluid levels correlate with disease severity and progression. *J. Neurol. Neurosurg. Psychiatry* 89, 239–247. <https://doi.org/10.1136/jnnp-2017-317138>.
87. Thompson, A.G., Gray, E., Thezenas, M.L., Charles, P.D., Evetts, S., Hu, M.T., Talbot, K., Fischer, R., Kessler, B.M., and Turner, M.R. (2018). Cerebrospinal fluid macrophage biomarkers in amyotrophic lateral sclerosis. *Ann. Neurol.* 83, 258–268. <https://doi.org/10.1002/ana.25143>.
88. Akama, T.O., Nishida, K., Nakayama, J., Watanabe, H., Ozaki, K., Nakamura, T., Dota, A., Kawasaki, S., Inoue, Y., Maeda, N., et al. (2000). Macular corneal dystrophy type I and type II are caused by distinct mutations in a new sulphotransferase gene. *Nat. Genet.* 26, 237–241. <https://doi.org/10.1038/79987>.
89. Caterson, B., and Melrose, J. (2018). Keratan sulfate, a complex glycosaminoglycan with unique functional capability. *Glycobiology* 28, 182–206. <https://doi.org/10.1093/glycob/cwy003>.
90. Zhang, Z., Takeda-Uchimura, Y., Foyez, T., Ohtake-Niimi, S., Narentuya, Akatsu, H., Nishitsuji, K., Michikawa, M., Wyss-Coray, T., Kadomatsu, K., and Uchimura, K. (2017). Deficiency of a sulfotransferase for sialic acid-modified glycans mitigates Alzheimer's pathology. *Proc. Natl. Acad. Sci. U S A* 114, E2947–E2954. <https://doi.org/10.1073/pnas.1615036114>.
91. Lehri-Boufala, S., Ouidja, M.O., Barbier-Chassefiere, V., Henault, E., Raisman-Vozari, R., Garrigue-Antar, L., Papy-Garcia, D., and Morin, C. (2015). New roles of glycosaminoglycans in α -synuclein aggregation in a cellular model of Parkinson disease. *PLoS One* 10, e0116641. <https://doi.org/10.1371/journal.pone.0116641>.
92. Scherzer, C.R., Eklund, A.C., Morse, L.J., Liao, Z., Locascio, J.J., Fefer, D., Schwarzschild, M.A., Schlossmacher, M.G., Hauser, M.A., Vance, J.M., et al. (2007). Molecular markers of early Parkinson's disease based on gene expression in blood. *Proc. Natl. Acad. Sci. U S A* 104, 955–960. <https://doi.org/10.1073/pnas.0610204104>.
93. Zheng, B., Liao, Z., Locascio, J.J., Lesniak, K.A., Roderick, S.S., Watt, M.L., Eklund, A.C., Zhang-James, Y., Kim, P.D., Hauser, M.A., et al. (2010). PGC-1 α , a potential therapeutic target for early intervention in Parkinson's disease. *Sci. Transl. Med.* 2, 52ra73. <https://doi.org/10.1126/scitranslmed.3001059>.
94. Kulak, N.A., Pichler, G., Paron, I., Nagaraj, N., and Mann, M. (2014). Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nat. Methods* 11, 319–324. <https://doi.org/10.1038/nmeth.2834>.
95. Kulak, N.A., Geyer, P.E., and Mann, M. (2017). Loss-less nano-fractionator for high sensitivity, high coverage proteomics. *Mol. Cell. Proteomics* 16, 694–705. <https://doi.org/10.1074/mcp.o116.065136>.
96. Fraser, K.B., Rawlins, A.B., Clark, R.G., Alcalay, R.N., Standaert, D.G., Liu, N., and West, A.B. (2016). Ser(P)-1292 LRRK2 in urinary exosomes is elevated in idiopathic Parkinson's disease. *Mov. Disord.* 31, 1543–1550. <https://doi.org/10.1002/mds.26686>.
97. Grinfeld, D., Aizikov, K., Kreutzmann, A., Damoc, E., and Makarov, A. (2017). Phase-constrained spectrum deconvolution for fourier transform mass spectrometry. *Anal. Chem.* 89, 1202–1211. <https://doi.org/10.1021/acs.analchem.6b03636>.
98. Wichmann, C., Meier, F., Virreira Winter, S., Brunner, A.D., Cox, J., and Mann, M. (2019). MaxQuant.Live enables global targeting of more than 25,000 peptides. *Mol. Cell. Proteomics* 18, 982–994. <https://doi.org/10.1074/mcp.TIR118.001131>.

99. Meier, F., Geyer, P.E., Virreira Winter, S., Cox, J., and Mann, M. (2018). BoxCar acquisition method enables single-shot proteomics at a depth of 10,000 proteins in 100 minutes. *Nat. Methods* *15*, 440–448. <https://doi.org/10.1038/s41592-018-0003-5>.
100. Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M.Y., Geiger, T., Mann, M., and Cox, J. (2016). The Perseus computational platform for comprehensive analysis of (prote) omics data. *Nat. Methods* *13*, 731–740. <https://doi.org/10.1038/nmeth.3901>.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Human Cerebrospinal Fluid (CSF)	Harvard Biomarker Study	Healthy adult donors and Parkinson's disease patients
Human Cerebrospinal Fluid (CSF)	Michael J. Fox Foundation for Parkinson's Research (MJFF) funded LRRK2 Cohort Consortium (LCC)	Healthy adult donors and Parkinson's disease patients, including LRRK2 G2019S carriers
Chemicals, peptides, and recombinant proteins		
Proteomics-grade modified trypsin	Sigma-Aldrich	Cat. no. T6567
Endopeptidase LysC	Wako Chemicals	Cat. no. 129-02541
Solid-phase extraction disks for SDB-RPS StageTips	Empore SDB-RPS	Cat. no.66886-U
Critical commercial assays		
PreOmics Lysis buffer	PreOmics GmbH	N/A
Deposited data		
Raw mass spectrometry data and Spectronaut output tables (identifier PXD02649)	This paper	https://www.ebi.ac.uk/pride/archive/
Software and algorithms		
Spectronaut (version 14.8.201029.47784)	Biognosys AG	https://biognosys.com/software/spectronaut/
Perseus (versions 1.6.0.7 and 1.6.1.3)	Tyanova et al., 2016 www.maxquant.org	https://maxquant.net/perseus/
GraphPad Prism (version 7.03)	N/A	https://www.graphpad.com/
MaxQuant Live software	Wichmann et al., 2019 www.maxquant.live	https://www.maxquant.org/mqlive/
OmicLearn (v1.0.0)	https://github.com/OmicEra/OmicLearn	https://share.streamlit.io/omicera/omiclearn/omic_learn.py
Python (version 3.7.6) using the pandas (version 1.0.1), numpy (version 1.18.1) and pingouin (version 0.3.4) packages	N/A	N/A

RESOURCE AVAILABILITY

Lead contact

Further information and request for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Dr. Matthias Mann (mmann@biochem.mpg.de).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD026491.
- This study did not generate custom computer code.
- Any additional information required to reanalyze the data reported in this work paper is available from the [lead contact](#) upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

In this study, CSF samples from two independent cross-sectional cohorts were analyzed. The first cohort consisted of 94 CSF samples from the Harvard Biomarkers Study (HBS) biobank and the second cohort was a subset of biobanked CSF samples from the Michael J. Fox Foundation for Parkinson's Research (MJFF)-funded *LRRK2* Cohort Consortium (LCC). HBS and the proteomics analysis of HBS samples conducted in the current study were approved by the Institutional Review Board of Brigham and Women's Hospital. The LCC study was established in 2009, when the MJFF LCC brought together investigators from North America, Europe, North Africa, and Asia to study individuals with mutations in the *LRRK2* gene. To be eligible to join the consortium, sites had to agree to share a core set of clinical data. Case report forms and standard operating procedures can be found at <https://www.michaeljfox.org/news/lrrk2-cohort-consortium>. Ethical review and approval was not required for the de-identified sample analysis in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study. Disease severity was assessed by the Unified Parkinson's Disease Rating Scale (UPDRS), which ranges from 0 for no impairment to a theoretical maximum of 199 for most severely affected individuals. UPDRS scores can be divided into four sub-scales for evaluating mentation, behavior, and mood (Part I, 0-16), activities of daily living (Part II, 0-52), motor examination (Part III, 0-108) and complications of therapy (Part IV, 0-23). Cognitive functioning was assessed using the Montreal Cognitive Assessment (MoCA), which ranges from 30 for no impairments to a theoretical minimum of 0 for most severely affected individuals. In addition, for the HBS cohort, Mini-Mental State Examination (MMSE) scores, which range from 30 to 0 (25-30 for normal cognition, 21-24 for mild dementia, 10-20 moderate dementia and 9 or lower for severe dementia) were also available.

METHOD DETAILS

Sample preparation

40 μ L of CSF samples were aliquoted in 96-well plates and processed with an automated set-up on an Agilent Bravo liquid handling platform.^{16,24} CSF samples were mixed with the equal amount of PreOmics lysis buffer (PreOmics GmbH) for reduction of disulfide bridges, cysteine alkylation, and protein denaturation at 95°C for 10 min. Upon 10 min cooling, 0.2 μ g of each protease trypsin (Sigma-Aldrich) and LysC (Wako) was added to each sample (well) and digestion was performed at 37°C overnight. Peptides were then purified on two 14-gauge StageTip plugs packed with styrenedivinylbenzene-reverse phase sulfonate (SDB-RPS).⁹⁴ Samples were first diluted with 1% trifluoroacetic acid (TFA) in isopropanol and loaded onto StageTips and subsequently washed with 200 μ L of 1% TFA in isopropanol twice and 200 μ L of 0.2% TFA/2% ACN (Acetonitrile). Peptides were eluted with 80 μ L of 1.25% Ammonium hydroxide (NH₄OH)/80% ACN, dried using a SpeedVac centrifuge at 45°C (Eppendorf, Concentrator plus), resuspended in 10 μ L buffer A* (2% v/v ACN, 0.2% v/v TFA, and stored at -20°C. Upon thawing before mass spectrometric analysis, samples were shaken for 5 min at 2,000 rpm (Thermomixer C, Eppendorf). Peptide concentrations were measured optically at 280nm (Nanodrop 2000, Thermo Scientific) and subsequently equalized using buffer A*. 500 ng peptide was subjected to LC-MS/MS analysis.

Cohort-specific libraries were generated by pooling of 24 randomly selected samples of each cohort and separating the peptides of this pooled sample into 24 fractions each by high pH (pH 10) reversed-phase chromatography on the "spider fractionator".⁹⁵ Fractions were concatenated automatically by shifting the collection tube every 120 seconds. Upon collection, fractions were dried and resuspended in buffer A* for LC-MS/MS analysis. To increase the depth of each library, we isolated extracellular vesicles (EV) from pooled CSF samples of each cohort by ultra-centrifugation.⁹⁶ Isolated EVs were resuspended in 100 μ L of a sodium deoxycholate-based lysis buffer containing chloroacetamide (PreOmics GmbH), heated to 95°C for 10 min for reduction and alkylation and then digestion using trypsin and LysC enzymes at 37°C overnight. Peptides were desalted with SDB-RPS StageTips as described above. Peptides were eluted 80% ACN/5% NH₄OH and the eluate was completely dried and resuspended in 0.1% formic acid (FA) for separation into eight fractions by high pH reversed-phase chromatography.⁹⁵

To determine coefficients of variation, five aliquots of a pooled CSF sample on one plate were subjected to sample preparation (intra-plate) and three aliquots of the same pool were subjected to sample preparation on two different plates (inter-plate).

LC-MS/MS analysis

LC-MS/MS analysis was performed on a Q Exactive HF-X Orbitrap mass spectrometer with a nano-electrospray ion source coupled to an EASY-nLC 1200 HPLC (all Thermo Fisher Scientific). Peptides were separated at 60°C on 50 cm columns with an inner diameter of 75 μ m packed in-house with ReproSil-Pur C18-AQ 1.9 μ m resin (Dr. Maisch GmbH). Mobile phases A and B were 99.9/0.1% water/FA (v/v) and 80/20/0.1% ACN/water/FA (v/v/v). MS data for single-shot CSF samples were acquired using the MaxQuant Live software and a data-independent acquisition (DIA) mode with phase-constrained spectrum deconvolution.^{97,98} Full MS scans were acquired in the range of m/z 300–1,650 at a resolution of 60,000 at m/z 200 and the automatic gain control (AGC) set to 3e6, followed by two BoxCar scans with 12 isolation windows each and a resolution of 60,000 at m/z 200 were acquired.⁹⁹ Full MS events were followed by 58 MS/MS windows per cycle in the range of m/z 300–1,650 at a resolution of 15,000 at m/z 200 and ions were accumulated to reach an AGC target value of 3e6 or for a maximum of 22 ms.

All fractionated samples including EV fractions were acquired using a top 12 data-dependent acquisition (DDA) mode. Full MS scans were acquired in the range of m/z 300–1,650 at a resolution of 60,000 at m/z 200. The automatic gain control (AGC) target was set to 3e6. MS/MS scans were acquired at a resolution of 15,000 at m/z 200.

Mass spectrometry data processing

The MS data of the fractionated pools and the single-shot CSF samples were combined into two cohort-specific hybrid libraries using Spectronaut version 14.8.201029.47784 (Biognosys AG). For all experiments except the machine learning with OmicLearn, the two cohorts were quantified separately. All searches were performed against the human SwissProt reference proteome of canonical and isoform sequences with 42,431 entries downloaded in July 2019. Searches used carbamidomethylation as fixed modification and acetylation of the protein N-terminus and oxidation of methionines as variable modifications. The Trypsin/P proteolytic cleavage rule was used, permitting a maximum of 2 missed cleavages and a minimum peptide length of 7 amino acids. The Q-value thresholds for library generation and DIA analyses were both set to 0.01. For individual protein correlations with clinical parameters and the machine learning, the Q-value data filtering setting in Spectronaut was set to "Qvalue". For all other analyses, the setting was set to "Qvalue percentile" with a cutoff of 25%, to use only those peptides for the protein quantification that passed the Q-value threshold in at least 25% of all analyzed samples. The various runs were normalized to each other by using median intensities of common peptides.

QUANTIFICATION AND STATISTICAL ANALYSIS

The Perseus software package versions 1.6.0.7 and 1.6.1.3¹⁰⁰ and GraphPad Prism version 7.03 were used for the data analysis. Protein intensities were log₂-transformed for further analysis apart from correlation and coefficient of variation analysis. Coefficients of variation (CVs) were calculated in Perseus for all inter-plate and intra-plate pairwise combinations of samples, the median values were reported as overall coefficient of variation. The protein CVs of the main study were calculated likewise within cohorts individually. For generation of the abundance curves, median protein abundances across all samples within a proteome were used. ANCOVA analysis was performed in python (version 3.7.6) using the pandas (version 1.0.1), numpy (version 1.18.1) and pingouin (version 0.3.4) packages. For the ANCOVA analysis, age at sample collection, *LRRK2* status (only in PD+ vs. PD-), and PD status (only LRRK2+ vs. LRRK2-) were set as confounding factors. The FDR (q values) was calculated using Benjamini-Hochberg correction. GO annotations were matched to the proteome data based on Uniprot identifiers. Annotation term enrichment was performed with Fisher exact test in Perseus separately for each cohort. Annotation terms were filtered for terms with an FDR of 5% after Benjamini-Hochberg correction in each cohort. Calculation of Pearson correlation scores and associated p values of protein intensities to UPDRS scores and other clinical parameters was performed in Perseus.

For machine learning, OmicLearn (v1.0.0) was utilized for performing the data analysis, model execution, and generating the plots and charts.⁴¹ Spectronaut output tables from the quantification analysis of both cohorts were used as the input for OmicLearn. No additional normalization on the data was performed. To impute missing values, a Zero-imputation strategy was used. Features were selected using ExtraTrees (n_trees = 500) strategy with the maximum number of 35 features. Normalization and feature selection was individually performed using the training data of each split. For classification, we used XGBoost-Classifer (random_state = 23 learning_rate = 0.3 min_split_loss = 0 max_depth = 15 min_child_weight = 1). We used (RepeatedStratifiedKFold) a repeated (n_repeats = 10), stratified cross-validation (n_splits = 5) approach to classify PD vs. HC, resulting in total 50 iterations for training the model, each time with 35 features. The average feature importance scores assigned by the classifier for each of the top20 features are shown in Figure 3G.