Urinary proteome profiling for stratifying patients with familial Parkinson's disease

Sebastian Virreira Winter[1,a,b], Ozge Karayel[1,a], Maximilian T Strauss[1,b], Shalini Padmanabhan[2], Matthew Surface[3], Kalpana Merchant[4], Roy N. Alcalay[3], Matthias Mann[1,5,c]

[1] Department of Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, Martinsried, Germany
[2] The Michael J. Fox Foundation for Parkinson's Research, NY, USA
[3] Department of Neurology, Columbia University, NY, USA
[4] Northwestern University Feinberg School of Medicine, IL, USA
[5] Novo Nordisk Foundation Center for Protein Research, Faculty of Health Sciences, University of Copenhagen, Copenhagen, Denmark

[a] These authors contributed equally
[b] Current address: OmicEra Diagnostics GmbH, Am Klopferspitz 19, 82152 Planegg, Germany
[c] Corresponding author

Correspondence: mmann@biochem.mpg.de

Keywords: Parkinson's disease, urine, body fluid, mass spectrometry, LC-MS/MS, DIA, neurodegeneration, proteomics, biomarker

SUMMARY

The prevalence of Parkinson's disease (PD) is increasing but the development of novel treatment strategies and therapeutics altering the course of the disease would benefit from specific, sensitive and non-invasive biomarkers to detect PD early. Here, we describe a scalable and sensitive mass spectrometry (MS)-based proteomic workflow for urinary proteome profiling. Our workflow enabled the reproducible quantification of more than 2,000 proteins in more than 200 urine samples using minimal volumes from two independent patient cohorts. The urinary proteome was significantly different between PD patients and healthy controls, as well as between *LRRK2* G2019S carriers and non-carriers in both cohorts. Interestingly, our data revealed lysosomal dysregulation in individuals with the *LRRK2* G2019S mutation. When combined with machine learning, the urinary proteome data alone was sufficient to classify mutation status and disease manifestation in mutation carriers remarkably well, identifying VGF, ENPEP and other PD-associated proteins as the most discriminating features. Taken together, our results validate urinary proteomics as a valuable strategy for biomarker discovery and patient stratification in PD.

INTRODUCTION

With a population prevalence of 0.2%, Parkinson's disease (PD) is the second most common neurodegenerative disorder after Alzheimer's disease [1]. It is characterized by the progressive loss of dopaminergic neurons and accumulation of α-synuclein-containing protein aggregates called Lewy bodies in the cytoplasm of the remaining neurons. As a result of dopaminergic neuron loss, PD manifests with motor signs and symptoms including bradykinesia, tremor and rigidity, and these characteristics are used for diagnosing the disease [1-3].

PD is a genetically complex disorder. Most patients do not carry a single pathogenic variant linked to PD, but a subset of about 10% of patients carry an identifiable pathogenic variant in genes such as *SNCA*, *PRKN*, *LRRK2* or *GBA*. For these individuals, the risk of developing the disease increases to 2 to 5% [2]. Among these genes, *LRRK2* is relatively common and causes PD in an autosomal dominant with incomplete penetrance fashion. How *LRRK2* mutations cause PD is unknown, however, several studies have indicated that disease-linked *LRRK2* mutations elevate its kinase activity and contribute to PD pathogenesis [4]. We have previously identified multiple Rab GTPases as endogenous targets of mutant LRRK2 [5-7]. Furthermore, inhibitors of this kinase have emerged as promising therapeutics for PD and clinical trials have already passed phase 1[8]. Although idiopathic forms of PD presumably represent a heterogeneous collection of pathogenic mechanisms, *LRRK2*-associated PD and idiopathic PD (iPD) show a similar phenotype in terms of disease symptoms and response to levodopa. The interest in LRRK2 as a therapeutic target is also fueled by the association between common variants in *LRRK2* and sporadic PD [9] and the observation that LRRK2 activity is increased in autopsied brain tissue from iPD patients without a known pathogenic mutation [10]. Thus, it will be important to conduct studies on ante-mortem biospecimens to gain insights into *LRRK2* mutation-induced changes and thereby identify iPD patients who could benefit from LRRK2-targeted therapies.

Current treatments, including levodopa - the most effective PD medication, mainly alleviate the motor symptoms rather than slow disease progression or reverse the pathology. Given the growing number of PD patients worldwide, and escalating economic and societal implications, there is an urgent need for disease-modifying therapeutics. The development of new therapeutic strategies requires better insights into the pathophysiologic mechanisms of PD as well as biomarkers to detect the earliest stages of PD before severe motor impairment is evident and irreversible brain damage has already occurred. Although cerebrospinal fluid (CSF) has been frequently used for biomarker studies of brain disorders, recent studies indicate that urine offers another promising clinically viable matrix for PD since it can be frequently and non-invasively collected in large volumes [11]. Importantly, urine contains not only kidney and urinary tract proteins but also filtered plasma proteins originating from distal organs, including the

brain [12, 13]. Therefore, urine protein analysis may provide diagnostic and prognostic opportunities for both urogenital and non-renal diseases [14-26]. Recent technical advances in fast and high-throughput sample preparation methods in conjunction with improvements in high-accuracy mass spectrometry (MS)-based proteomics have enabled characterization of the urinary proteome [27-30]. However, to what extent neurodegenerative disorders including PD affect the urinary proteome remains unknown.

Our group has recently employed state-of-the-art MS-based proteomics to obtain proteome profiles of the two body fluids, plasma and CSF, in multiple disease conditions [31-35]. Here, we extend this technology to urinary proteome profiling and provide first evidence that this approach can be used for PD biomarker discovery. More specifically, we focused our analysis on two large patient cohorts, both including healthy control subject, non-manifesting carriers of the frequently found *LRRK2* G2019S mutation, manifesting patients with the same mutational signature, and PD patients without the *LRRK2* mutation. The composition of the cohorts, quality of the data and the depth of proteome coverage allowed us to identify pathogenic *LRRK2*-regulated lysosomal protein signatures that could serve as biomarkers to stratify subjects with pathogenic *LRRK2*. Taken together, our study offers evidence that quantitative MS-based proteomics represents a clinically useful strategy for non-invasive monitoring of disease progression and treatment response as well as patient stratification in PD.

RESULTS

Overview of PD cohorts for urinary proteomics

Here, we employ a recently described 'rectangular' biomarker discovery strategy in which as many proteins as possible are measured using shotgun MS-based proteomics for all the individuals in both discovery and validation cohorts [33, 36]. To map proteome changes between individuals with different mutation status and manifestation of disease, we analyzed 235 urine samples from two independent cross-sectional cohorts each comprised of four subject groups: (1) healthy controls (HC, *LRRK2*-/PD-); (2) non-manifesting carriers (NMC) harboring the *LRRK2* G2019S mutation (*LRRK2*+/PD-); (3) idiopathic PD patients (iPD, *LRRK2*-/PD+); and (4) manifesting PD patients with *LRRK2* G2019S (LRRK2 PD, *LRRK2*+/PD+) (**Figure 1A & Table 1**).

The first cohort was recruited at Columbia University Irving Medical Center (hereinafter referred to as 'Columbia cohort' and color-coded with orange) [11, 37]. Participants in the Columbia cohort included 35 HC, 16 NMC, 40 iPD, 28 LRRK2 PD individuals, and one PD patient with an unknown *LRRK2* status. The cohort included 52 female sex and 68 male sex individuals (**Figure 1A and Table 1)**. The *GBA* (gene that encodes for lysosomal acid glucosylceramidase (GCase)) mutation status

was also available for all individuals, with 22 of them harboring a *GBA* variant and 98 the wild type allele. PD+ and PD- subjects were frequency-matched by age with means of 67.0±9.3 and 64.1±12.0 (± SD) years, respectively (**Supplementary Figure 1A**). Their motor skills were assessed using the Unified Parkinson's Disease Rating Scale part III (UPDRS-III) and cognitive functioning with the Montreal Cognitive Assessment (MoCA) test (**Supplementary Figure 1B-C)**. Genotyping for *LRRK2* G2019S mutation was conducted as previously described [38].

To confirm findings from the Columbia cohort, we additionally analyzed a subset of biobanked urine samples from the Michael J. Fox Foundation for Parkinson's Research (MJFF)-funded *LRRK2* Cohort Consortium (LCC) (hereinafter referred to as 'LCC cohort' and color-coded with blue). We determined urine proteomes for 26 HC, 37 NMC, 29 iPD, and 23 LRRK2 PD individuals (53 female and 62 male) (**Figure 1A and Table 1)**. In the LCC cohort, individuals in the non-diseased group were somewhat younger (53.8±13.9) than PD patients (67±7.6) (means ± SD) years (**Supplementary Figure 1D)**. In addition, LCC sample collection protocols were less stringent than in the Columbia cohort and UPDRS-III and MoCA scores were not available, indicating that the Columbia cohort is more powerful for our analyses. Both studies were approved by local institutional review boards, and each participant signed an informed consent (See **Supplementary Table 1** for a detailed overview).

Proteomic characterization of urine samples

For the proteomic profiling of individual urine samples, we developed a high-throughput proteomics workflow building on the PVDF-based sample processing method MStern blotting by the Steen group [27] combined with data-independent acquisition (DIA) LC-MS/MS [39, 40] (**Figure 1A**). To maximize proteome depth, we generated two 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 neat urine samples; (2) a DDA library consisting of 8 fractions of extracellular vesicles isolated from pooled neat urine samples; and (3) a direct-DIA library generated from the DIA analysis of all analyzed samples (see Methods). In these hybrid libraries, we identified a total of 4,564 and 5,725 protein groups for the Columbia and LCC cohorts, respectively (**Supplementary Figure 1E**). Applying this robust workflow, we quantified on average 2,026 (Columbia) and 2,162 (LCC) protein groups per neat urine sample, in single runs of 45 minutes and using less than 100 µl of starting material (**Supplementary Table 2**). Three outlier samples were excluded from further analysis due to low proteome depth (**Fig 1B and 1C, Supplementary Table 1**). The quantified protein intensities spanned five orders of magnitude in both cohorts and the top ten most abundant proteins contributed about half to the total urinary proteome signal (**Fig 1D and 1E**). As observed before [17], the molecular weight distribution spanned a wide range with many proteins exceeding 100 kDa. More than 2,000

proteins were in common between the two cohorts. To best of our knowledge, this study presents the deepest urinary proteome coverage for single-run analysis to date, a promising basis for the discovery of biomarkers.

Data from repeated measurements of individual samples revealed a high reproducibility with more than 90% of proteins having an intra- and inter-plate coefficient of variation (CV) below 50% in both studies and about 60% of proteins with a CV below 20% (**Figure 1G, 1H, Supplementary Figure 2**). The intra- and inter-plate variability within each cohort was even lower (**Supplementary Figure 2C and 2I**), while the inter-individual variability was much larger with no protein having a CV below 20% (**Supplementary Figure 2F and 2L**). Thus, our proteomic quantification accuracy greatly exceeds the biological variability that we seek to measure.

Table 1. Demographics of all participants

Columbia cohort	HC (*LRRK2*-/PD-) (n = 35)	NMC (*LRRK2*+/PD-) (n = 16)	iPD (*LRRK2*-/PD+) (n = 40)	LLRK2 PD (*LRRK2*+/PD+) (n = 28)
Age at collection, mean (SD)	67.5 (10.3)	56.8 (12.7)	64.9 (9.2)	70.7 (8.7)
Age at onset, mean (SD)	n/a	n/a	57.8 (11)	57.9 (11.3)
Sex (female/male)	17/18	8/8	15/25	11/17
GBA (mut/WT)	7/28	1/15	11/29	3/25
MoCA	27.5 (2)	28.7 (1.1)	26.9 (1.6)	26.3 (4.5)
UPDRS-III	1.1 (1.5)	0.8 (1.1)	17.7 (10)	20.6 (8)
LCC cohort	HC (*LRRK2*-/PD-) (n = 26)	NMC (*LRRK2*+/PD-) (n = 37)	iPD (*LRRK2*-/PD+) (n = 29)	LLRK2 PD (*LRRK2*+/PD+) (n = 23)
Age at collection, mean (SD)	56.1 (16)	52.1 (12.1)	65 (9.1)	68.4 (5)
Age at onset, mean (SD)	n/a	n/a	58.4 (9.1)	57.7 (7.6)
Sex (female/male)	15/11	16/21	9/20	13/10

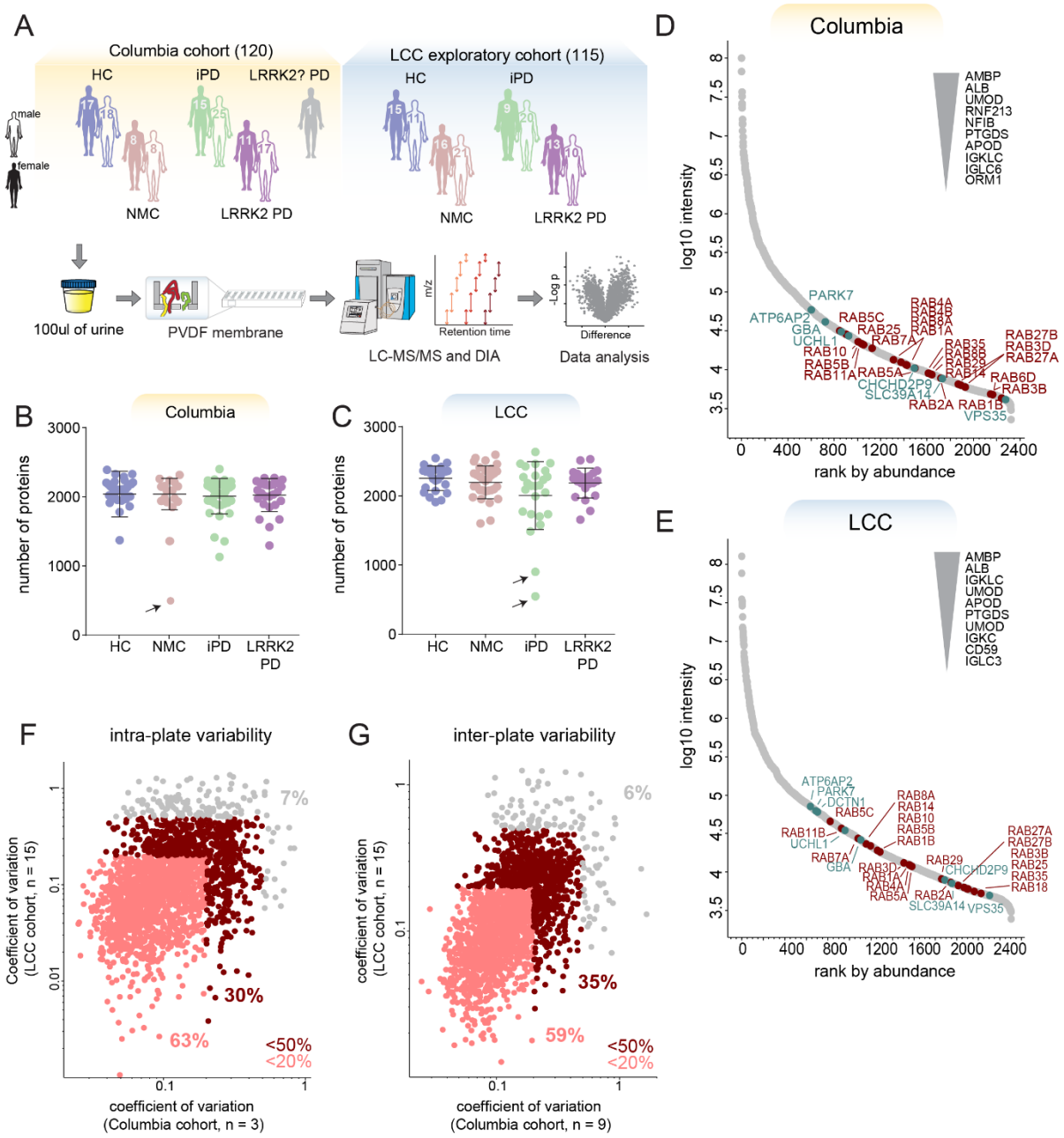


Figure 1. MS-based proteomic analysis of two independent urinary PD cohorts has high depth and precision

A) Overview of the two cohorts and the proteomic workflow. Urine samples comprised of four subject groups ((HC, iPD, NMC and LRRK2 PD) were prepared using MStern blotting and analyzed by LC-MS/MS using data-independent acquisition (DIA). The sex and total number of subjects per cohort group is shown.

B-C) Number of proteins identified and quantified with a 1% false discovery rate (FDR) in each sample in the Columbia (B) and LCC (C) cohorts. Bars indicate mean and standard deviation.

Arrows point at one subject from the Columbia and two subjects from the LCC cohort that were excluded from further analysis due to low proteome depth.

D-E) Proteins identified in the Columbia (D) and LCC (E) cohort were ranked according to their MS signals, which covered more than five orders of magnitude. The top ten most abundant Parkinson-related proteins (green) and Rab GTPases (red) are labeled.

F-G) Quantification precision assessed by calculating the intra-plate (F) and inter-plate (G) coefficients of variation (CVs) for the Columbia and LCC cohorts. Proteins with a CV below 20% and 50% in both cohorts are highlighted in light and dark red, respectively and the fractions of proteins above and below these CV thresholds are shown. A total of 2,051 proteins were consistently quantified in both cohorts.

Quality assessment of urine samples

Pre-analytical variation caused by inconsistent sample processing and contaminations during sample collection can have a strong impact on the results and may cause the reporting of incorrect biomarkers [34]. To ensure that the observed proteome changes are not caused by artifacts related to sample handling and processing, we assessed each sample for potential quality issues. To this end, we used a previously reported quality marker panel to determine the degree of contamination with erythrocytes [34] (**Figure 2A-Band Supplementary Table 3**). Insufficient removal of cells and cellular debris from urine leads to an increased detection of intracellular proteins with a high sample-to-sample variability compared to regularly secreted urinary proteins [41]. We therefore generated a second urine-specific quality marker panel to assess the degree of contamination with cells and cellular debris that could originate from aged, inflamed or damaged tissue of the kidneys, bladder or the urinary tract (see Methods). Although urine samples from both cohorts were cleared by centrifugation following collection to avoid this systematic bias, our procedure flagged four samples from the Columbia cohort for potential contamination with cellular components (**Figure 2A-B)**. Taken together, 6 samples from the Columbia cohort and 4 samples from the LCC cohort showed increased intensities of contamination markers and were thus excluded from further analyses. In addition, we further excluded one sample from the Columbia cohort, as it clustered far away from all other samples in a principal component analysis (PCA), likely indicating pre-analytical variation.

Next, we generated a global correlation map of the urinary proteome to identify clusters of functional co-regulation as previously reported for plasma proteome profiling [31]. The global correlation map contains pairwise relations of all urinary proteins across 112 samples from the Columbia cohort. Unsupervised hierarchical clustering of the pairwise Pearson correlation coefficients revealed four main and several small clusters of co-regulated proteins (**Figure 2C).** The largest of these clusters was chiefly enriched for proteins with the Gene ontology (GO)-term ‘extracellular exosome’ as well as other significant terms (**Supplementary Table 4**).

We also identified a cluster of highly correlated proteins that was enriched for the GO-terms 'immunoglobulin' and 'B-cell receptor', suggesting that these proteins originate from immune cells. The two further main clusters were enriched for proteins originating from sex-specific tissues such as the prostate and vagina (**Figure 2C**) [42]. This shows that sex-dependent anatomical differences strongly affect the urinary proteome and thus should be considered as confounding factors. Indeed, a principal component analysis indicated sex as the strongest contributor to the inter-individual variance of the urinary proteome (**Figure 2D and 2E**).

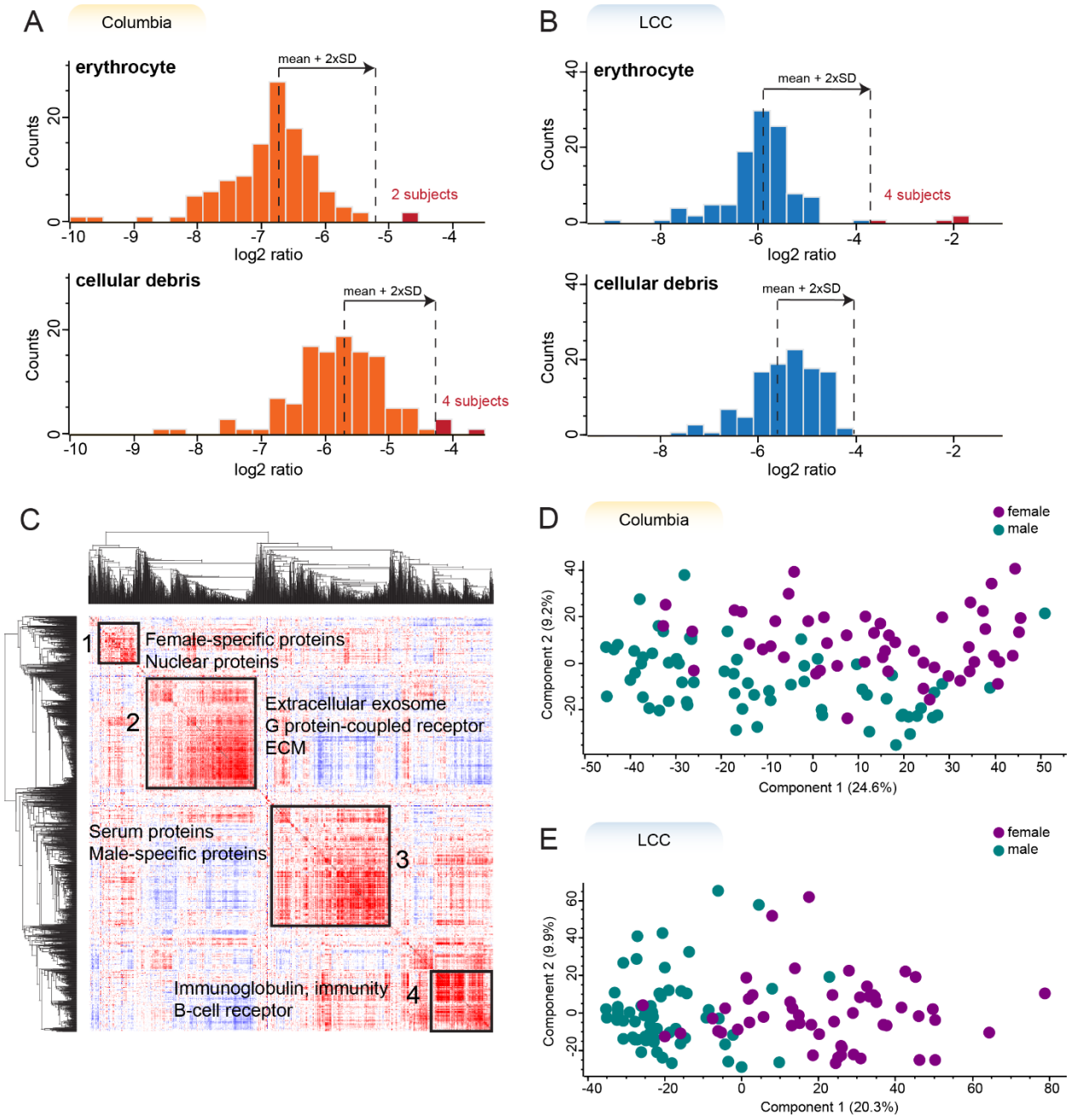


Figure 2. The large majority of urine samples has high quality and shows sex-specific protein expression

A-B) Histograms of log2 transformed ratios of the summed intensity of the proteins in the respective quality marker panel and the summed intensity of all proteins in Columbia (A) and LCC (B) cohorts. A sample was flagged for potential contamination and removed from further analysis if the ratio differed more than two standard deviations from the mean of all samples within the cohort. The proteins in each quality marker panel are listed in Supplementary Table 3.

C) Global correlation map of proteins generated by clustering the Pearson correlation coefficients of all pairwise protein comparisons for the Columbia cohort.

D-E) Principal component analysis (PCA) of all subjects based on their urinary proteome profiles. Female subjects are shown in purple and males in green.

Detection of PD-related proteome alterations in urine

Although PD primarily manifests in the central nervous system and is characterized by motor impairments, it is known to affect and potentially initiate in peripheral tissues and is associated with non-motor symptoms [43, 44]. Thus, we asked if the disease also causes changes of the urinary proteome, which reflects proteins from both central and peripheral organs. To establish PD-associated changes in urine proteome, we first determined which proteins are differentially present in the urine of PD patients compared to the controls, irrespective of their *LRRK2* status (HC and NMC). To control for confounders, we performed an analysis of covariance (ANCOVA) considering sex, age at sample collection, *LRRK2* status and *GBA* status (only available for the Columbia cohort) as confounding factors. Applying a 5% false discovery rate (FDR) cut off, we identified 361 proteins that displayed significantly different levels in PD patients when compared to controls (HC and NMC) (298 in Columbia cohort and 73 in LCC cohort) (**Supplementary Table 5**). The smaller number of significantly different proteins in the LCC cohort as well as the relatively small overlap between the cohorts could be explained by a less stringent sample collection protocol and worse age-matching in the LCC cohort. The log2 fold-changes between PD patients and non-diseased individuals show a good correlation between the two cohorts (Pearson r = 0.65) (**Figure 3A**), reflecting both reproducibility of the applied proteomic workflow and pathobiological consistency. The mean fold-changes of the 330 PD-associated proteins that were quantified in both cohorts were larger for the Columbia cohort (Columbia: 1.43 (up) & 0.49 (down) vs. LCC: 1.27 (up) & 0.75 (down)). Furthermore, 90% of the PD-associated proteins were detected with at least two peptides and quantified with CVs below 50% (**Supplementary Figure 3A**).

Protein misfolding is known to be involved in many neurodegenerative conditions including PD [45]. Interestingly, some of the proteins exhibiting the largest differential levels between the urine of controls vs. PD

patients include proteins assisting other proteins in folding, such as peptidyl-prolyl cis-trans isomerase B (PPIB) and T-complex protein 1 subunit gamma (CCT3) (**Figure 3A**). We also identified two of the eight human canonical ribonucleases (RNASE1 and RNASE2) to be PD-associated in both cohorts (**Figure 3A**). The levels of the four apolipoproteins APOA1, APOA2, APOA4 and APOC1 were also elevated in PD patients (**Figure 3A**). While they show a similar trend in both cohorts, they reached statistical significance only in the Columbia cohort, corroborating that this cohort has greater power to detect PD-associated changes.

Next, we analyzed if any GO-terms assigned to the 361 PD-associated proteins were significantly enriched compared to the urinary proteome (**Figure 3B**). This analysis examines if PD affects individual cellular compartments and particular biological signaling networks in urine. The term 'bone development' was significantly enriched, in line with previous findings that PD patients are at increased risk for osteoporosis and osteopenia [46]. In summary, we observed disease-associated protein signatures with a high correlation between the two independent cohorts and identified promising candidates that could serve as biomarkers for PD and provide mechanistic insights into disease pathogenesis.

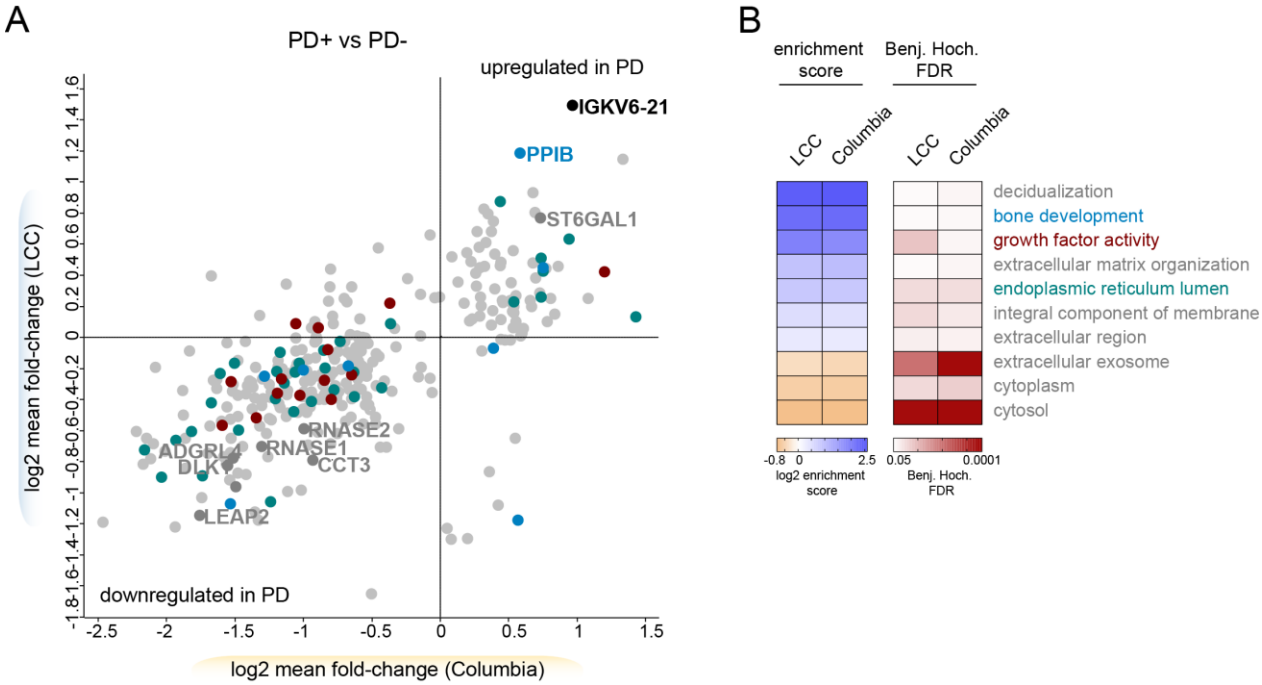


Figure 3. PD affects the urinary proteome

A) Correlations of mean fold-changes of the proteins changing PD-dependently in the Columbia and LCC cohorts. Only proteins quantified in both cohorts are shown (n=330). The colors match

to the GO terms shown in (B). Proteins overlapping between the two cohorts are labeled with their name.
B) Fisher exact test to identify significantly enriched GO-terms in the PD-associated proteins in urine. All GO-terms that were significant in both cohorts are displayed (FDR < 5%).

Pathogenic *LRRK2*-dependent changes are linked to lysosomes and glycosphingolipid metabolism

Encouraged by the observation of disease-dependent proteome changes in urine, we next asked if the urinary proteome is altered by the presence of the *LRRK2* G2019S mutation. We again applied an ANCOVA analysis with sex, age at sample collection, PD status and *GBA* status (only available for the Columbia cohort) as confounding factors and compared the proteomes between G2019S and wild type allele carriers. Applying an FDR of 5%, the mutation altered the abundance of 237 proteins (FDR < 5%, Columbia: 166, LCC: 104) (**Figure 4A and Supplementary Table 5**). A subset of 33 proteins differed significantly in G2019S carriers in both cohorts and all these proteins were upregulated in pathogenic *LRRK2* carriers. A pairwise comparison of the four subject groups (HC, NMC, iPD and LRRK2 PD) using a student's t-test confirmed that the abundance of the overlapping proteins changed in a G2019S-dependent manner but was unaffected by the PD status (**Figure 4B)**.

In total, 227 *LRRK2*-status associated proteins were quantified in both cohorts and the fold-changes of these were similar between the two cohorts (**Figure 4C**), although the effect sizes were slightly larger for the Columbia cohort (Columbia: 1.43 (up) & 0.76 (down) vs. LCC: 1.39 (up) & 0.89 (down)). Interestingly, one of the proteins exhibiting the largest increase in *LRRK2* G2019S carriers in both cohorts was a phosphatase, the Intestinal-type Alkaline Phosphatase (ALPI). As for the proteins that changed dependent on PD disease status, most *LRRK2*-status associated proteins were detected with at least two peptides and quantified with CVs below 50% (**Supplementary Figure 3B**).

A GO-term analysis revealed strong enrichment of proteins associated with lysosome-related terms such as 'autolysosome', 'lysosome', 'lysosomal lumen', 'azurophil granule lumen' and 'lysosomal membrane' as well as 'glycosphingolipid metabolic processes' in *LRRK2* G2019S carriers in both cohorts (**Figure 4D**). Among the proteins associated with the lysosome-related GO-terms were multiple members of the cathepsin family including cathepsins A, B, C, D, H, L, O, S, and Z. The widely used lysosomal marker proteins, LAMP1 and LAMP2, were also significantly altered in *LRRK2* carriers in the LCC cohort, while LAMP3 was significantly changed in the Columbia cohort. In total, 13 proteins were associated with the GO term 'sphingolipid metabolic process', most of them upregulated in *LRRK2* G2019S carriers. Among them were multiple lysosomal enzymes including GCase (encoded by *GBA*), galactocerebrosidase (GALC),

sphingomyelin phosphodiesterase (SMPD1), and the beta-hexosaminidase subunits alpha and beta (HEXA and HEXB).

Heterozygous pathogenic mutations in *GBA* are one of the most common PD risk factors while homozygous loss of function mutations of the same protein cause the lysosomal storage disorder (LSD) Gaucher's disease. Past studies have described increased GCase activity in *LRRK2* deficient mice, and decreased GBA activity in *LRRK2* G2019S carrier neurons [47, 48]. Additionally, it has been reported that PD patients with mutations in both proteins develop symptoms at a younger age compared to patients with only one affected gene [49-51]. However, despite these reports, it remained unclear whether mutations in *GBA* and *LRRK2* contribute to the pathogenesis of PD via common pathways. Participants in the Columbia cohort were sequenced for mutations in *GBA* [38] and 22 individuals were found to carry a pathogenic mutation in this locus. To determine which proteins were changed specifically in carriers of *GBA* variants, we performed an ANCOVA analysis with sex, age at sample collection, PD status and *LRRK2* status as confounding factors. Using an FDR of 5%, we found that levels of 74 proteins were affected by *GBA* (**Figure 4E and Supplementary Table 5**). Interestingly, only Intercellular adhesion molecule 1 (ICAM1), Adenosylhomocysteinase (AHCY) and Stomatin (STOM) were affected by pathogenic mutations in both *LRRK2* and *GBA*, suggesting that the two mutations largely affect distinct pathways. Furthermore, the *GBA*- and *LRRK2*-dependent protein fold-changes were poorly correlated (Pearson r = 0.21) (**Figure 4F**) but future well-powered studies on *GBA* cohorts are needed to firm up the data. Of note, most proteins associated with the GO term 'glycosphingolipid metabolic process' were increased in *LRRK2* G2019S carriers but decreased in pathogenic *GBA* carriers, most notably GM2 activator (GM2A).

Together, we identified pathogenic mutant *LRRK2*-dependent protein signatures with a high correlation between the two independent cohorts. The *LRRK2* mutational status-dependent changes of the urinary proteome include lysosomal proteins that could serve as biomarkers to stratify subjects with pathogenic *LRRK2*.

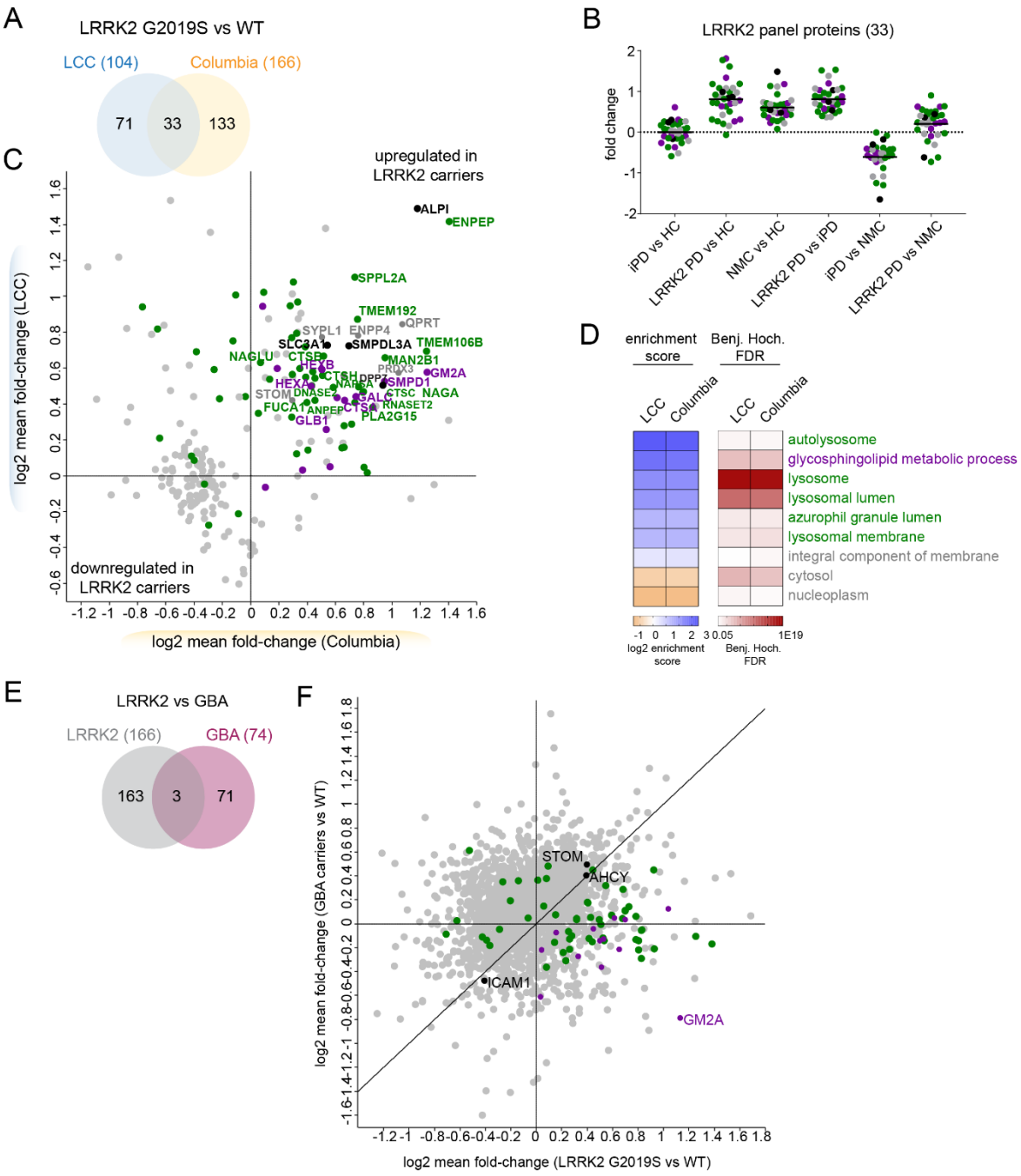


Figure 4. Pathogenic *LRRK2*-dependent lysosomal dysregulation is reflected in the urinary proteome and distinct from pathogenic *GBA*-induced alterations

A) Proteins that differ significantly differ between pathogenic *LRRK2* carriers and controls using an ANCOVA analysis with sex, age, PD status and *GBA* status as confounders and an FDR of 5%. **B**) Mean fold-changes for each of the 33 proteins that were *LRRK2*-dependently regulated in both cohorts using a pairwise t-test comparing the four subgroups (HC, NMC, iPD and *LRRK2* PD). **C**) Correlation of mean fold-changes of the proteins changing *LRRK2*-dependently in the Columbia and LCC cohorts. Only proteins identified in both cohorts are shown (n=227). The

colors match to the GO terms shown in (D). Proteins overlapping between the two cohorts are labeled with their name.

D) Fisher exact test was performed to identify significantly enriched GO-terms in the *LRRK2*-dependently regulated proteins in urine. All GO-terms that were significant in either cohort are displayed (FDR < 5%).

E) Proteins that differ significantly between pathogenic *GBA* carriers and controls or pathogenic *LRRK2* carriers and controls using an ANCOVA analysis with sex, age, PD status and *LRRK2*/*GBA* status as confounders and an FDR of 5%. In total, 237 proteins were differentially expressed in these two comparisons with 166 and 74 regulated proteins in the *LRRK2* carriers and *GBA* carriers, respectively, only three of which were common between both mutations.

F) Correlation of mean fold-changes of the proteins changing *LRRK2*-dependently and *GBA*-dependently (n=237) in the Columbia cohort. Carriers of pathogenic variants in both *GBA* and *LRRK2* were excluded from the analysis. The colors match to the GO terms shown in (D).

Correlation of proteome profiles with clinical parameters

Given that clinical parameters, including disease severity scores, were available for the Columbia cohort, we were interested in exploring whether any of these clinical parameters correlate with proteomic changes we detected. We were especially interested in the cognitive capabilities of the participants as evaluated using the Montreal Cognitive Assessment (MoCA) test, and the motor performance as assessed using the Unified Parkinson's Disease Rating Scale part III (UPDRS-III). Within the Columbia cohort, MoCA scores ranged from 8 to 30, on a scale from 0, for severe cognitive impairment, to 30, for no measurable cognitive impairment. We observed that two proteins, Tenascin-R (TNR) and Furin (FURIN), showed a strong negative correlation with the MoCA score in PD patients (TNR Pearson r: -0.66; FURIN r: -0.65; $p < 10^{-7}$ for both), mainly driven by *LRRK2* G2019S carriers (TNR r: -0.77; FURIN r: -0.78; $p < 10^{-5}$ for both) (**Figure 5A**). When similar type of analysis was done with UPDRS-III scores, which ranged from 0 to 38 in the Columbia cohort (on a scale from 0 assigned for normal to 56 for severely affected motor function), we observed that immunoglobulin kappa variable 6-21 (IGKV6-21), was the highest correlated protein in PD patients (r: 0.54, $p < 10^{-5}$) (**Figure 5C**). This protein also exhibited one the highest fold-change in abundance when comparing urine of PD patients with non-diseased individuals (**Figure 3B**). Of note, the correlation between UPDRS-III scores and levels of IGKV6-12 was mainly driven by iPD patients (r: 0.68; $p < 10^{-5}$) and much weaker in *LRRK2* G2019S PD patients (r: 0.36; not significant) (**Figure 5C**). Collectively, this analysis suggests that iPD and *LRRK2* G2019S patients could be stratified based on the differences between MoCA and UPDRS-III score correlations with different urine proteins.

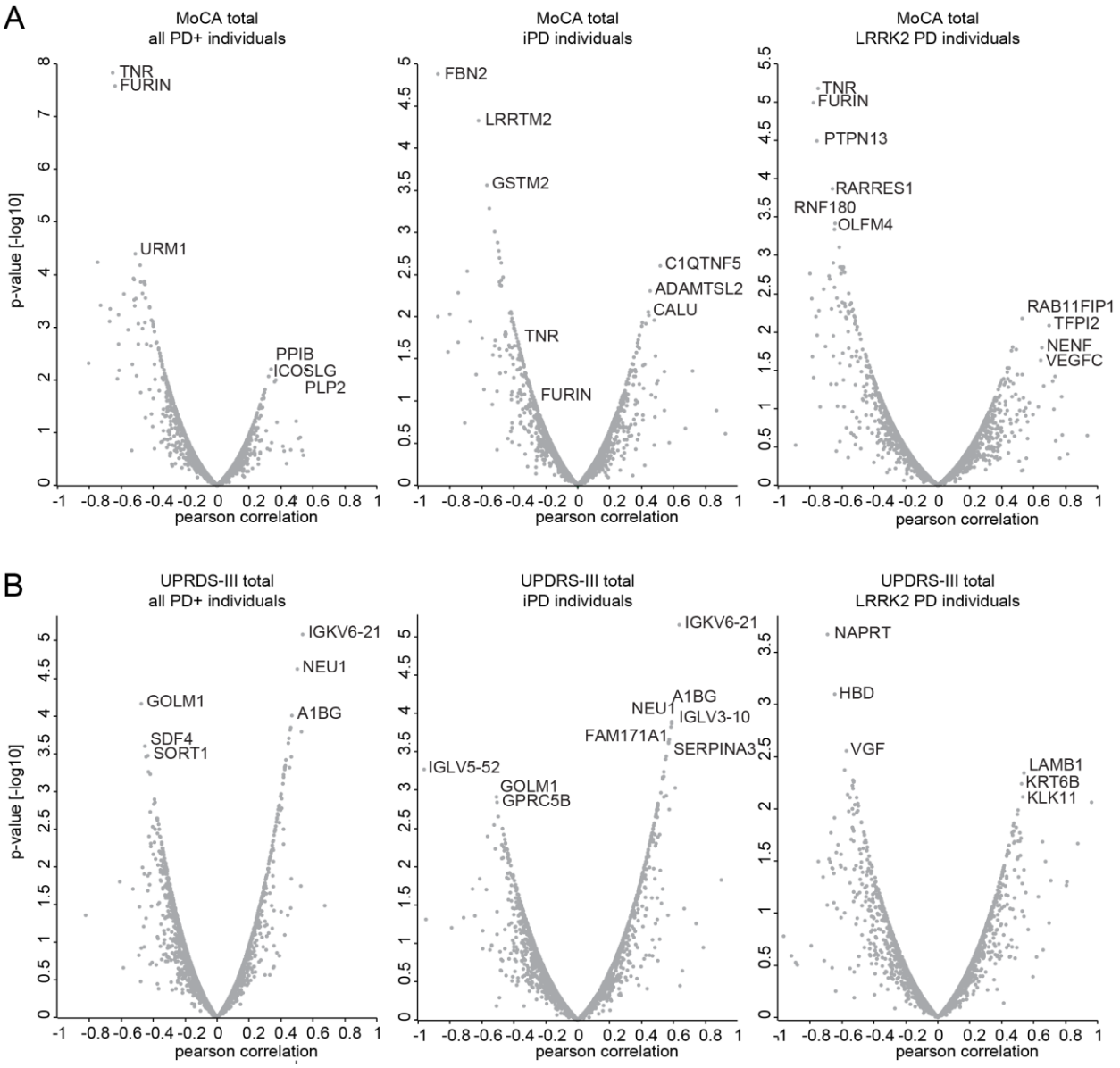


Figure 5. Correlations with clinical parameters

A) Pearson correlation scores and associated p-values [-log10] of all protein intensities with the MoCA total score. Either all PD patients (left), iPD patients (middle) or LRRK2 PD patients (right) were included in the analysis.

B) Pearson correlation scores and associated p-values [-log10] of all protein intensities with the UPDRS-III score. Either all PD patients (left), iPD patients (middle) or LRRK2 PD patients (right) were included.

Machine learning-based classification of urinary proteomes

Finally, we assessed how well machine learning models can discriminate between PD patients and non-diseased individuals, between *LRRK2* G2019S and wildtype allele carriers, and between NMC individuals and *LRRK2* G2019S patients based on the acquired urinary proteome profiles. Since the accuracy of the model largely depends on the number of samples, we combined all samples from the Columbia and LCC cohorts for these analyses. We first selected and ranked which protein features to use in the machine learning model by employing a decision tree. To classify individuals as having PD or not, the decision tree selected the 15 most important features of the PD^- vs. PD^+ urinary proteomes, with the intensity of PPIB, one of the proteins that displayed the largest difference in abundance when PD samples were compared to the controls (**Figure 3A**), being on top of the list (**Supplementary Figure 5A**). Using these proteins, we trained an XGBoost model, a commonly applied algorithm for gradient boosting, a machine learning technique that is used to build robust predictive models based on ensembles of weaker predictions, such as decision trees. Samples were cross-validated by applying a stratified 4-fold split. This was repeated (n=15) with shuffling the dataset to have a total of 60 train/test-splits to achieve a robust estimate of model performance. Each time, we determined a receiver operating characteristic (ROC) curve and found the mean area under the curve (AUC), which is often used to assess the performance of a model, to be 0.84±0.05 (**Figure 6A**). On average, we correctly classified 91 out of 117 PD patients and 77 out of 106 controls in the test sets (**Figure 6B**). Accordingly, the machine learning model reached a sensitivity of 78% and a specificity of 73%. When we trained the model on the one cohort and tested it on the other cohort, we obtained AUCs of 0.86 or 0.72, further demonstrating the robustness of the model (**Figure 6C**).

We next used the same machine learning methods to classify G2019S and wildtype *LRRK2* carriers using the same strategy as described above. The decision tree selected the 15 most important features, with the intensity of ENPEP being the most important one (**Figure 4A**; **Supplementary Figure 5B**). Using these proteins and the XGBoost algorithm, we obtained a mean AUC of the ROC curves of 0.87±0.04 (**Figure 6D**). For the test sets, we could correctly classify 73 out of the 99 *LRRK2* G2019S carriers and 103 out of the 123 wildtype allele carriers, corresponding to a 74% sensitivity and 84% specificity (**Figure 6E**). When we trained the model on one of the cohorts and tested it on the other, we obtained AUCs of 0.76 or 0.80 (**Figure 6F**). Additionally, we trained the model on all individuals with a known *LRRK2* status and classified the sample from the Columbia cohort with an unknown *LRRK2* status with an 87% probability to be wild type *LRRK2*. After we had finished this machine learning modelling, the mutational status of this individual was determined as wild type *LRRK2*, further verifying the machine learning model.

Encouraged by these results, we wanted to see how well machine learning can discriminate *LRRK2*+ PD patients from

NMCs that also carry a *LRRK2* mutation and are at increased risk of developing the disease. Using a decision tree, we selected seven proteins for training the model (**Supplementary Figure 5C**). Interestingly, VGF, a neurotrophic factor, was identified as the most important feature. When using these features to train a classifier with our cross-validation scheme, the obtained mean AUC of the ROC curve was 0.94±0.05 and the obtained sensitivity and specificity were both 88% (**Figure 6G/H**). Using samples from only one cohort as a training set and applying the model to the other cohort resulted in AUCs of 0.93 and 0.74 (**Figure 6I**). Taken together, machine learning allowed us to classify the PD and *LRRK2* states with high specificities and sensitivities.

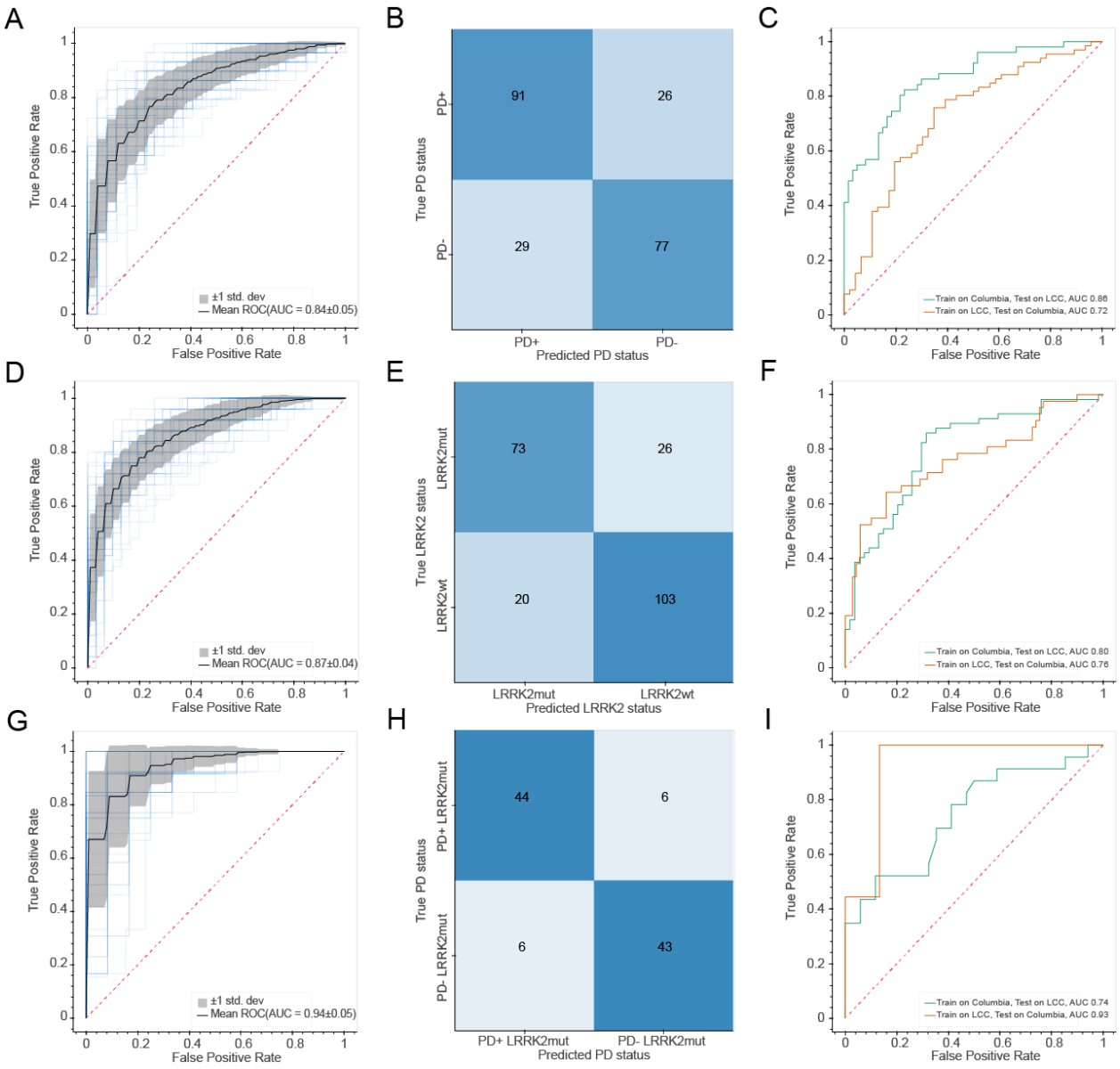


Figure 6. Machine learning-based classification of PD and *LRRK2* status

A) Receiver operating characteristic (ROC) curve for the XGBoost-based model to classify PD+ vs. PD- individuals. Random performance is indicated by the dotted diagonal line. The gray area represents the standard deviation from the mean ROC curve. The blue lines show the values for a total of 15 repeats with four stratified train-test splits.

B) Confusion matrix showing the model performance for classifying PD+ vs. PD- individuals. Numbers represent the mean number from 15 repeats of cross-validation with four stratified train-test splits.

C) ROC curve for the XGBoost-based model when trained on one cohort and tested on the other cohort. Random performance is indicated by the dotted diagonal line.

D) Same as A) but for classification of *LRRK2* G2019S vs. *LRRK2* WT carriers.

E) Same as B but for classification of *LRRK2* G2019S vs. *LRRK2* WT carriers.

F) Same as C) but for classification of *LRRK2* G2019S vs. *LRRK2* WT carriers.

G) Same as A) but for classification of PD+ vs. PD- in *LRRK2* G2019S carriers.

H) Same as B) but for classification of PD+ vs. PD- in *LRRK2* G2019S carriers.

I) Same as C) but for classification of PD+ vs. PD- in *LRRK2* G2019S carriers.

DISCUSSION

The pathophysiology of PD leads to progressive decline of motor function and results in numerous quality of life issues for patients and their families, and inevitably leads to death within 7 to 14 years from the initial diagnosis. The majority of previous PD biomarker discovery and validation efforts have focused on CSF, serum and blood [52]. Additional strategies included targeted monitoring of α-synuclein levels, given the known relationship between α-synuclein accumulation and PD progression [53]. To address this problem, we developed a shotgun proteomics workflow for urinary proteome profiling. We chose to focus on urine given the non-invasive nature of obtaining clinical samples, which is a major advantage when developing a strategy that can be used not only for diagnostic and prognostic purposes, but for long-term disease progression and treatment response monitoring. Additionally, instead of focusing on a single biomarker and/or a subset of molecular entities, our shotgun proteomic approach provides a multiparameter global map of the disease state. We previously showed that this strategy can yield powerful, data-driven descriptors of a disease [33, 35, 54], and we now confirm, for the first time, that this also works for urinary proteome analysis in the context of a complex neurodegenerative disease, such as PD.

Our quantitative shotgun proteomic workflow represents a sensitive and scalable approach for rapid analysis of a large number of samples. Applying this workflow to more than 200 urine samples from two independent cohorts allowed us to precisely quantify on average more than 2,000 proteins per sample while using minimal sample amounts of less than 100 µl. Our approach successfully

determined proteins with abundances that varied over more than five orders of magnitude, and quantified more than 1,200 proteins with a CV below 20% across the two cohorts, highlighting the high depth and precision of our study. Moreover, the observed variability between samples was much smaller than the biological variability between subjects, further illustrating the quantitative robustness of our workflow.

Another factor contributing to the quality of the urinary proteome dataset reported here, is the composition of the cohorts we analyzed. The cohorts included two types of controls, the healthy controls as well as asymptomatic individuals that are carriers of PD-associated mutation G2019S *LRRK2*. The cohorts also included PD patients with and without the mutation, and patients of both sexes, thus allowing for different types of comparisons. For example, the global correlation map and PCA analysis showed that the sex of an individual has a dominant effect on the urinary proteome. This is in line with basic physiology and previous reports [55] but highlights the importance of incorporating sex as a confounding factor for statistical analyses. This is further illustrated by the fact that 42 and 12% of the proteins with differential abundance in PD patients vs. controls, as well as 35 and 14% of the proteins that exhibit different abundance in a *LRRK2* mutational status dependent manner also significantly differed between the sexes in the Columbia and LCC cohorts, respectively.

Applying our 'rectangular' strategy for biomarker discovery [36], we discovered 361 and 237 significantly altered proteins in PD patients and pathogenic *LRRK2* carriers, respectively. The observed overlap of proteins exhibiting significantly perturbed levels in the two independent cohorts confirms that valuable information can be inferred from the urinary proteome for neurodegenerative diseases. We note the scalability of our workflow, which will allow its application to larger cohorts with more comprehensive genetic and clinical information. This extension of our work will be important to further validate our results and to discover additional biomarker candidates with improved statistical power.

Our data analysis led to several interesting observations that might suggest opportunities for follow up. Here, we will briefly discuss only a small number of such examples. For those interested in more in depth data mining, we made our datasets available via publicly accessible depository (see Materials and Methods for accession numbers). An interesting insight that emerged from the GO-term analysis of PD patient vs. control proteomes identified the GO-term 'bone development' as significantly enriched. The enzyme PPIB was significantly upregulated in PD patients in both cohorts. This cyclophilin assists the folding of type I collagen and can protect cells against MPP^+-induced cell death in a PD cell culture model [56]. Inhibitors of the closely related family member cylophilin D (CypD) are considered as therapeutic agents against several neurodegenerative diseases including PD [57]. Most other proteins associated with the GO-term 'bone development' were downregulated in PD patients, in line with

recent findings that PD patients frequently suffer from osteoporosis and osteopenia [46, 58]. Going forward, it would be important to examine the relationship between PD profession and bone health more closely, as this connection is currently underexplored.

Another enriched term was 'growth factor activity', although none of the proteins associated with this term was significantly regulated in both cohorts. Growth factors and particularly neurotrophic factors have gained strong interest as therapeutic agents in Parkinson's disease but so far have not produced convincing clinical benefits [59]. The neurosecretory protein VGF was strongly decreased in PD patients in both cohorts (Columbia: 0.24, LCC: 0.54) but only reached statistical significance in the better-controlled Columbia cohort. VGF is synthesized as a prohormone and proteolytically processed to various biologically active peptides. In this study, we identified peptides covering most of the VGF sequence, including sequences contained in the neuroendocrine regulatory peptide-1. However, the applied tryptic digestion complicates a direct link to the endogenous hormone peptides. VGF is exclusively synthesized and secreted by neuronal and neuroendocrine tissues. In the CNS, VGF promotes neurite growth and exhibits neuroprotective activity, while it also regulates energy homeostasis in peripheral tissues. Gene expression of VGF in the cortex [60] and peptides derived from this gene are reduced in post-mortem parietal brain cortex and plasma from PD patients [61, 62]. Furthermore, VGF has been suggested as a biomarker in CSF for Alzheimer's Disease (AD) and Amyotrophic lateral sclerosis (ALS) and its expression was reduced in the CSF of AD and ALS patients compared to controls [63, 64].

We also identified several apolipoproteins - the major proteinaceous constituent of lipoproteins - to be significantly upregulated in PD patients in the Columbia cohort. They have been linked to neurodegenerative disorders including Alzheimer disease, including in our recent proteomic study of CSF [54]. APOE variants were shown to exhibit neuroprotective activity (reviewed in [65]). APOA1 is the major protein component of plasma high-density lipoprotein and its low levels in CSF and plasma have been reported as a potential PD biomarker [66-68]. While APOE and ApoA1 are the most abundant apolipoproteins in the CSF and highly enriched in the brain [69, 70], APOC1 - a less-abundant brain apolipoprotein- was implicated in Alzheimer disease although its regulation and possible role is poorly understood [71].

In another illustrative example, we analyzed proteomic differences between patients with and without a major inherited mutation associated with familial PD, *LRRK2* G2019S. Lysosomal dysregulation and associated α-synuclein aggregation appear to be a central event in the pathogenesis of PD [72] and LRRK2, through its regulation of the endolysosomal pathways, is a key player in this mechanism [73, 74]. Interestingly, the *LRRK2*-dependent signature in the urinary proteome seemed to be more consistent than the PD-dependent signature, as indicated by the larger overlap of 33 vs. 10 proteins

between the two cohorts. This suggests that the genetic mutation of *LRRK2* not only manifests in the central nervous systems but also dysregulates multiple pathways in distal organs such as the bladder and kidney, where LRRK2 is actually highly expressed [75]. Our results demonstrate that urine of pathogenic *LRRK2* carriers strongly reflects lysosomal dysregulation associated with increase in LRRK2 activity [72] [72, 73]. This suggests that the genetic mutation of *LRRK2* not only manifests in the central nervous systems but also dysregulates multiple pathways in distal organs such as the bladder and kidney, where LRRK2 is highly expressed [75]. Additionally, one of the strongest upregulated proteins in *LRRK2* G2019S carriers was the alkaline phosphatase ALPI. We suggest that this enzyme may counteract the hyperactive kinase by as yet undiscovered feedback mechanisms. Interestingly, knockdown of ALPI has been shown to decrease both LRRK2 levels and activity in cells [76]. We also found several lysosomal proteins including α- and β-subunit of β-hexosaminidase A (HEXA and HEXB), GM2A and GCase, whose genes are associated with LSDs, to be upregulated in *LRRK2* G2019S carriers in both cohorts. Mutations in many LSD genes have been associated with PD, suggesting common pathogenic mechanisms underly both diseases. The GO-term 'glycosphingolipid metabolic process' was enriched among *LRRK2*-regulated proteins, in agreement with increased interest in understanding how sphingolipids contribute to PD that stems from the fact that several PD-associated genes including *GBA* are linked to their metabolism [77, 78]. Ceramide levels are increased in *LRRK2*-deficient mouse brains and this decrease is regulated by *GBA* [48]. It still remains unclear how the disruption of sphingolipid metabolism may result in PD-associated neurodegeneration or if LRRK2 directly or indirectly regulates this process. Our data suggest that pathogenic mutations in *GBA* and *LRRK2* mainly affect distinct regulatory networks, as only three proteins were significantly altered in common by mutations in both genes. However, further studies on larger *GBA* cohorts are needed to confirm and extend our findings.

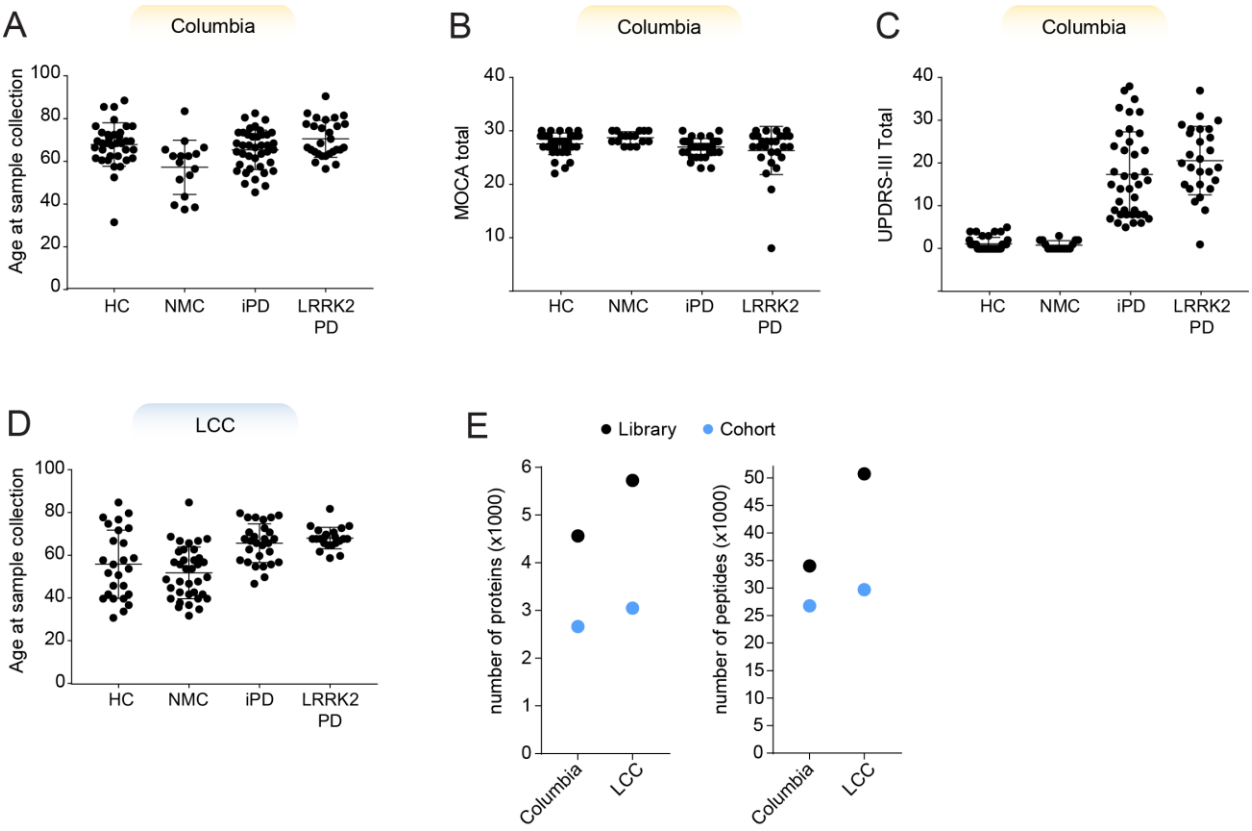
One of the cohorts we analyzed (Columbia) included information on clinical scores of cognitive and motor performances. This allowed us to correlate proteomic changes to clinical score, thus revealing that TNR and FURIN levels were strongly correlated with higher cognitive impairment. FURIN is a protease and is involved in NMDA-induced neuronal injury [79]. Furthermore, its homologue in the fruit fly, Furin1, has been reported to be a translational target of pathogenic LRRK2 and to be involved in neurotoxicity [80]. TNR is a neural extracellular matrix protein exclusively expressed in the brain. It is involved in neurogenesis [81] and extracellular matrix aggregates in the brain called perineural nets [82]. Of note, rare TNR variants have also been associated with familial PD [83]. Interestingly, IGKV6-21 was highly upregulated in PD patients and also strongly correlated with the UPDRS-III score. Although the underlying biology is unclear, the association with both PD risk and severity

make this V region a promising biomarker candidate to pursue in future studies.

To extend utility of our datasets, we developed a machine learning model for stratifying PD patients and *LRRK2* G2019S carriers with high sensitivities and specificities. Importantly, the machine learning model excelled in classifying the PD status in *LRRK2* G2019S carriers. This is of high interest, because although these carriers are at an increased risk of developing PD, there is no predictive marker to determine whether or not and when a mutation carrier develops the disease. Given the performance of the machine learning model, VGF, LTF, CELA3A, TUBB4B, and SOD2 are promising candidates as predictive markers to early indicate disease development.

In summary, we have demonstrated that a distal body fluid like urine contains brain-specific proteins and can inform about the disease and mutation status in a neurodegenerative disease. Our urinary proteomics workflow is relatively straightforward, readily scalable and thus easily applicable to larger and more powerful cohorts. It would be important to also apply it to longitudinal data to confirm increased levels of PPIB and IGKV6-21 in PD patients and VGF as a potential indicator for disease manifestation in *LRRK2* G2019S carriers but also identify new biomarkers for PD risk and disease progression in idiopathic and genetic forms of PD. Our results demonstrate that urinary proteome profiling enables the discovery of better biomarkers, which could have a major impact on important aspects of disease management: (i) a diagnostic biomarker will enable early and objective diagnosis of PD, (ii) a prognostic biomarker will provide information about the progression of the disease, and (iii) predictive and treatment response biomarkers will allow to monitor whether and how the patients respond to a therapy. Reliable biomarkers assessing LRRK2 activity can also aid with monitoring compliance of LRRK2 kinase inhibitors and treatment efficacy, early detection of non-manifesting carriers to prevent disease onset and stratify idiopathic PD patients who could benefit from LRRK2-based therapies.

SUPPLEMENTARY FIGURES AND LEGENDS



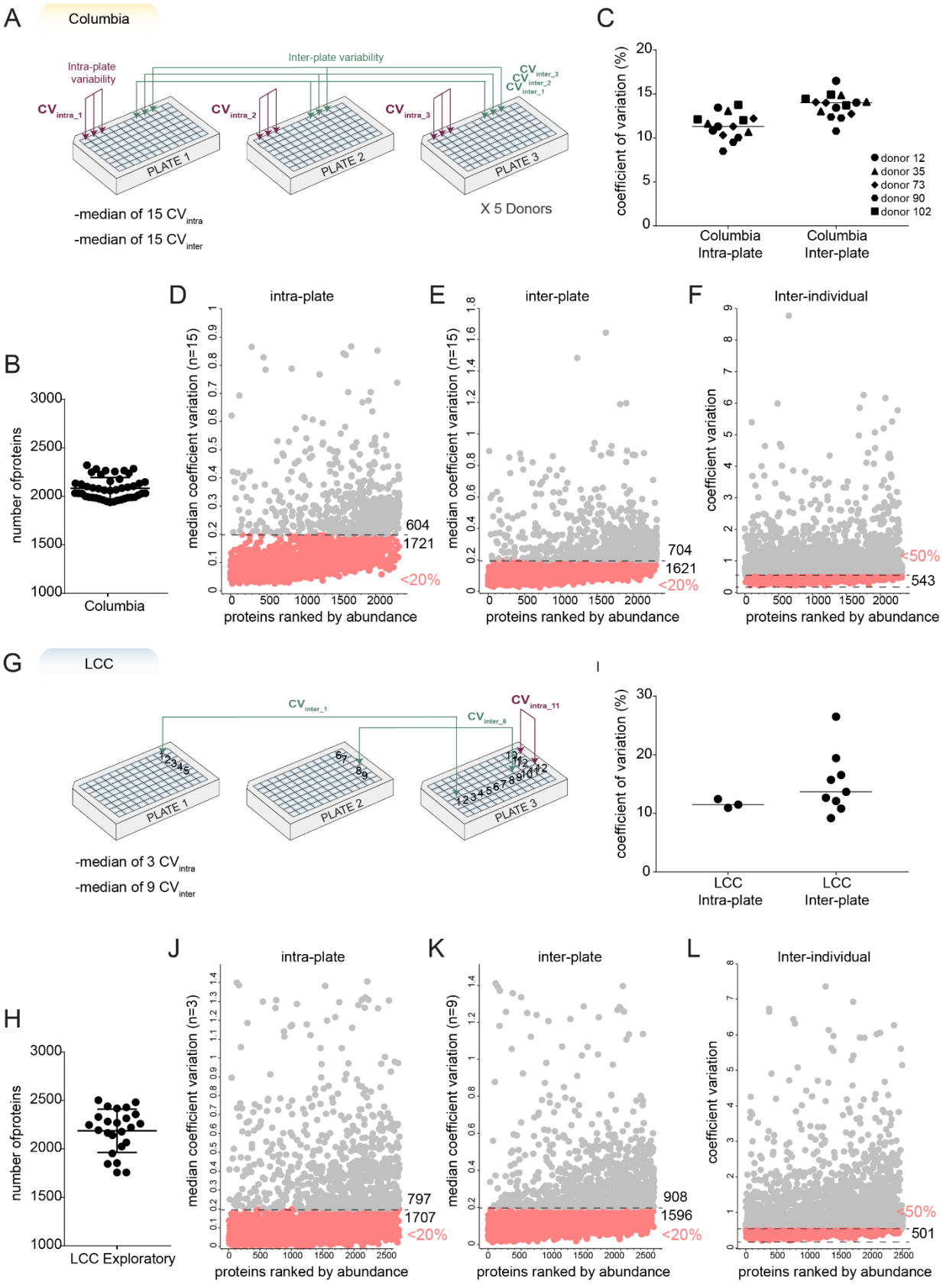
Supplementary Figure 1. Clinical information and library depth

A) Age of subjects at the time of sample collection for all participants of the Columbia cohort. Bars represent mean and standard deviation in all panels.

B-C) Clinical scores from MoCA (B) and UPDRS-III (C) for all individuals of the Columbia cohort.

D) Age of subjects at the time of sample collection for all participants of the LCC cohort.

E) Number of proteins and peptides identified in each cohort-specific hybrid library.



Supplementary Figure 2. Validation of quantification precision

A) Graphical overview of the experiment to determine coefficients of variation (CVs) for the analysis of the Columbia cohort.

B) Number of proteins identified in samples of CV determination experiment for the Columbia cohort.

C) Median intra- and inter-plate CV values for the Columbia cohort.

D-**E**) Median intra-plate (D) and inter-plate (E) CVs for each protein quantified in the Columbia cohort. Proteins were ranked according to their abundance and proteins with a median CV below 20% are highlighted in pink. Numbers of proteins above and below this CV threshold are given.

F) Inter-individual CVs were calculated from the analysis of every individual in the Columbia cohort. Proteins with a CV below 50% are highlighted in pink.

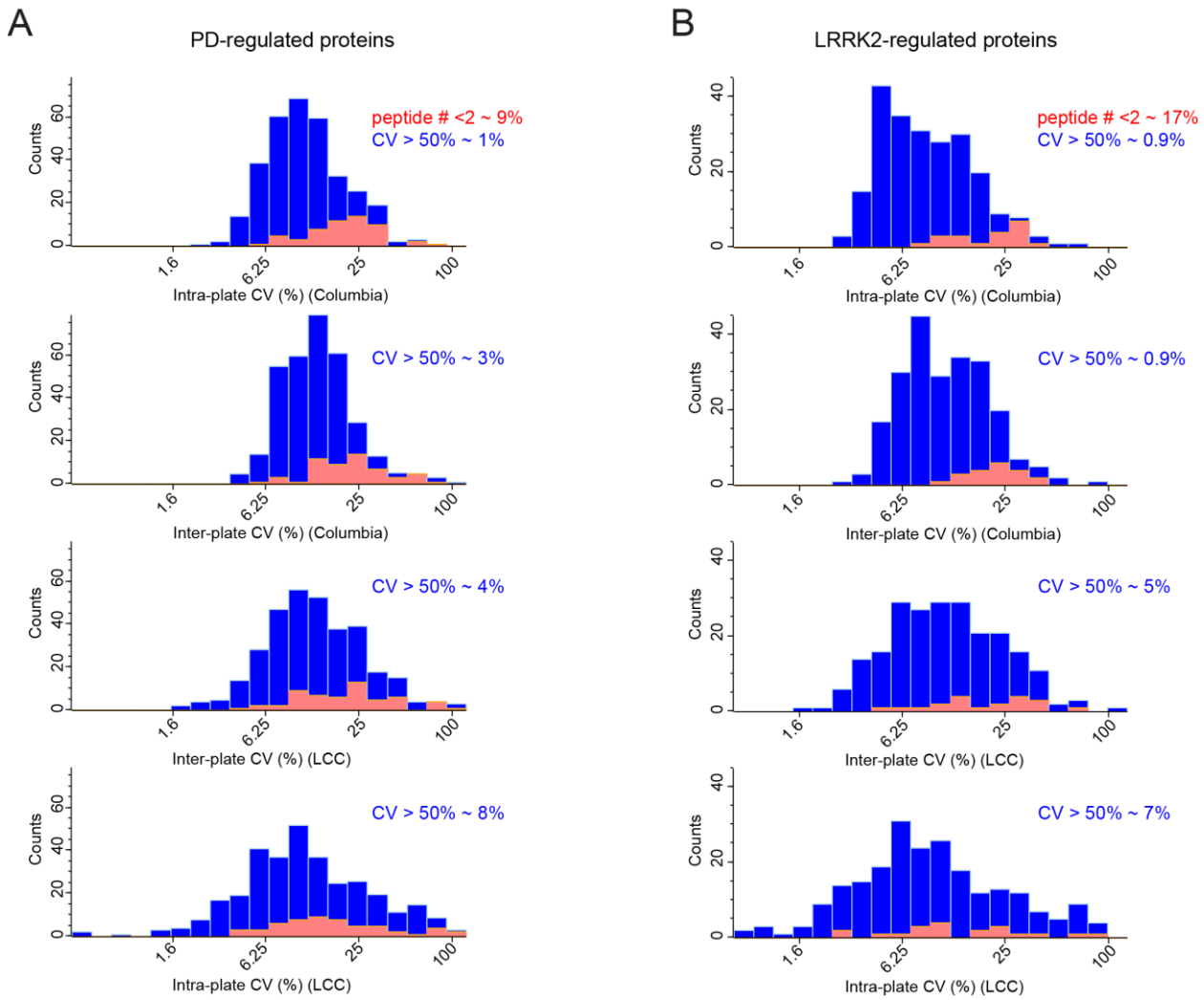
G) Graphical overview of the experiment to determine coefficients of variation (CVs) for the analysis of the LCC cohort.

H) Number of proteins identified in samples of CV determination experiment for the LCC cohort.

I) Median intra- and inter-plate CV values for the LCC cohort.

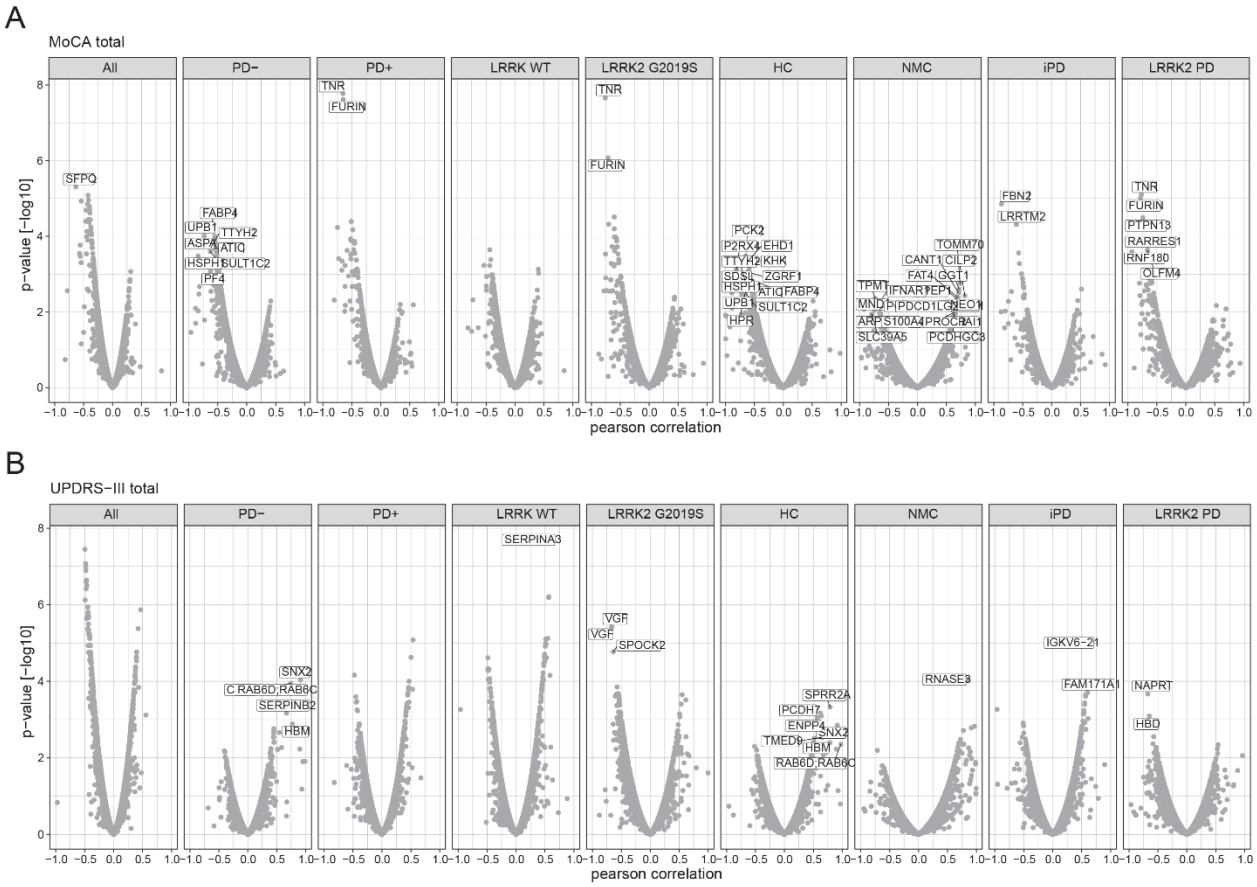
J-**K**) Median intra-plate (D) and inter-plate (E) CVs for each protein quantified in the LCC cohort. Proteins were ranked according to their abundance and proteins with a median CV below 20% are highlighted in pink. Numbers of proteins above and below this CV threshold are given.

L) Inter-individual CVs were calculated from the analysis of every individual in the LCC cohort. Proteins with a CV below 50% are highlighted in pink.



Supplementary Figure 3. Quality assessment of urine samples

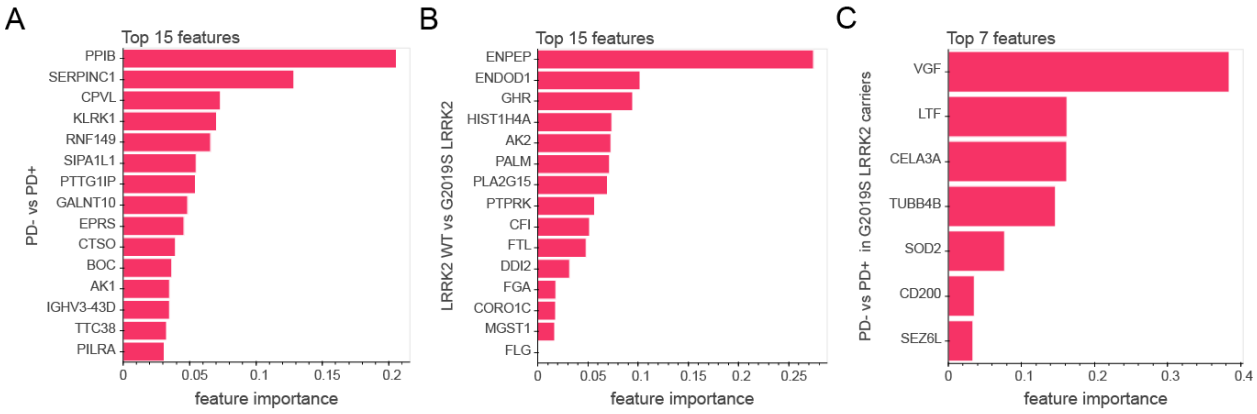
A-B) The quantification precision shown as intra- or inter-plate CVs for the Columbia and LCC cohorts for (A) PD-regulated proteins and (B) *LRRK2*-regulated proteins.



Supplementary Figure 4. Correlation analysis

A) Pearson correlation scores and associated p-values [-log10] of all protein intensities with the MoCA total score. The subset of individuals included in the analyses is shown on top.

B) Pearson correlation scores and associated p-values [-log10] of all protein intensities with the UPDRS-III score. The subset of individuals included in the analyses is shown on top.



Supplementary Figure 6. Decision-tree-based feature selection for machine learning
A) Top15 most important features according to a decision tree-based feature selection to classify PD+ vs. PD- individuals.
B) Top15 most important features according to a decision tree-based feature selection to classify *LRRK2* G2019S vs. *LRRK2* WT carriers.
C) Top7 most important features according to a decision tree-based feature selection to classify PD+ vs. PD- in *LRRK2* G2019S carriers.

SUPPLEMENTARY TABLES

Supplementary Table 1. Clinical information on cohort subjects
Supplementary Table 2. All proteins quantified in two cohorts
Supplementary Table 3. Contamination markers
Supplementary Table 4. GO-terms and associated proteins
Supplementary Table 5. Significantly regulated proteins

MATERIALS and METHODS

Study cohorts

In this study, urine samples from two independent cross- sectional cohorts were analyzed. Both studies were approved by local institutional review boards, and each participant signed an informed consent.

The first cohort was recruited at Columbia University Irving Medical Center (Columbia cohort) and its participants donated urine under a MJFF-funded *LRRK2* biomarker project from March 2016 to April 2017. This cohort contained 35 healthy individuals without pathogenic LRRK2 mutation (HC), 16 non-manifesting carriers of the LRRK2 G2019S mutation (NMC), 40 idiopathic PD patients without pathogenic LRRK2 mutation (iPD) and 28 PD patients with the pathogenic *LRRK2* G2019S mutation (LRRK2 PD) and 1 PD patient with an unknown mutation status of *LRRK2*. Motor performance was evaluated using the Unified Parkinson's Disease Rating Scale part III (UPDRS-III), and cognitive functioning was

assessed using the Montreal Cognitive Assessment (MoCA) test. Genotyping for *LRRK2* G2019S and *GBA* mutations was conducted as previously described [38].

To confirm findings from the Columbia cohort, urine from a second cohort consisting of 115 bio-banked urine samples from the Michael J. Fox Foundation for Parkinson's Research (MJFF)-funded *LRRK2* Cohort Consortium (LCC) was analyzed. The cohort used in this study was an exploratory subset of a larger cohort and contained 26 healthy individuals without pathogenic *LRRK2* mutation (HC), 37 non-manifesting carriers of the *LRRK2* G2019S mutation (NMC), 29 idiopathic PD patients without pathogenic *LRRK2* mutation (iPD) and 23 PD patients with the pathogenic *LRRK2* G2019S mutation. UPDRS-III and MoCA scores were not available for subjects from the LCC cohort.

Quality assessment

To generate the urine-specific quality marker panel, we recruited three volunteers from within the Department of Proteomics and Signal Transduction at the Max Planck Institute of Biochemistry who kindly donated 10 ml of urine at three different time points during a day and provided a written informed consent, with prior approval of the ethics committee of the Max Planck Society.

Following the collection, urinary samples were centrifuged at 2000g for 10 minutes, supernatants were harvested and pellets were resuspended in 100 µl of Urea sample solution. 100 µl of each supernatant and the entire 100 µl of the resuspended pellets were used for sample preparation as described below. A sample was flagged for potential contamination if the summed intensity of all proteins in the respective quality marker panel differed more than 2 standard deviations from the mean of all samples within the cohort.

Sample Preparation

The undiluted neat urine as well as the cleared and pelleted urine samples for the urine-specific quality marker panel were prepared using MStern Blot protocol as described previously [27]. Briefly, 100 µl of urine was first diluted in 300 µl of Urea sample solution (8 M urea in 50 mM ammonium bicarbonate (ABC)) and subsequently mixed with 30 µl of 150 mM dithiothreitol (DTT) solution (150 mM DTT, 8 M urea, 50 mM ABC) in a 96-well plate. The resulting solution was incubated for 20 min at room temperature. Reduced cysteine side chains were alkylated by adding 30 µL of iodoacetamide (IAA) solution (700 mM IAA, 8 M urea, 50 mM ABC) and incubated for 20 min in the dark. During incubation, each well of the 96-well PVDF membrane plates (MSIPS4510, Merck Millipore) was activated and equilibrated with 150 µl of 70% ethanol/water and urea sample solution, respectively. The urine samples were transferred through the PVDF membranes using a vacuum manifold (MSVMHTS00, Merck Millipore). Adsorbed proteins were washed two times with 150 µl of 50 mm ABC. Digestion was performed at 37°C for 2 hours by adding 100 µl digestion buffer (5% v/v acetonitrile (ACN)/50 mm ABC) containing 0.35 µg per well of each protease trypsin and LysC. After incubation in a

humidified incubator, the resulting peptides were collected by applying vacuum and remaining peptides were eluted twice with 75 µl of 40%/0.1%/59.9% (v/v) acetonitrile/formic acid/water. The pooled peptide solutions were dried in a vacuum centrifuge.

Peptides resuspended in 0.1% trifluoroacetic acid (TFA) were desalted on C18 StageTips as described in [29]. The StageTips were centrifuged at 1,000g for washing with 0.1% TFA and elution with 80% ACN/0.1% TFA. The eluate was evaporated to dryness using a vacuum centrifuge and peptides were resuspended in 10µl buffer A* (2% ACN/0.1% TFA and stored at -20°C. Samples were thawed shortly before mass spectrometric analysis and shaken for 2 minutes at 2000rpm (thermomixer C, Eppendorf). Peptide concentrations were measured optically at 280nm (Nanodrop 2000, Thermo Scientific) and subsequently equalized using buffer A*. 500ng peptide was subjected to LC-MS/MS analysis.

Cohort-specific libraries for data-independent analyses were generated by pooling of 25 randomly selected samples of each cohort. Sample pools were fractionated into 24 fractions each by high pH (pH 10) reversed-phase chromatography as described earlier [84]. Fractions were concatenated automatically by shifting the collection tube every 120 seconds and subsequently dried in a vacuum centrifuge and resuspended in buffer A*.

To increase the depth of each library, extracellular vesicles (EV) were isolated from pooled urine samples of each cohort by ultra-centrifugation as described earlier (Add Andy West paper here). Briefly, 8.5 ml of 6 urine samples per group (*LRRK2*-/PD-, *LRRK2*+/PD-, *LRRK2*-/PD+ and *LRRK2*+/PD+) were pooled were centrifuged at 10,000g for 30 min at 4 °C and supernatant was transferred and then centrifuged again at 100,000g for 1h at 4 °C. Supernatants were discarded and pellets were washed by adding 30 mL PBS and centrifugation at 100,000g for 1 h at 4 °C. Supernatant was discarded and pellets were resuspended in100 µl of a sodium deoxycholate-based lysis buffer containing chloroacetamide (PreOmics GmbH) and heated to 95°C for 10 min for reduction and alkylation. After cooling to room temperature, 0.75 µg of each protease trypsin and 0 LysC were added to each sample and digestion was performed at 37°C overnight. Peptides were desalted with SDB-RPS (styrenedivinylbenzene- reverse phase sulfonate) StageTips. Samples were mixed with 5 volumes of 1% TFA/isopropanol for loading on StageTips and subsequently washed once with 1%TFA/isopropanol and once with 0.2% TFA as described earlier [29]. Peptides were eluted 80%/5% ACN/ammonium hydroxide. The eluate was completely dried using a vacuum centrifuge and resuspended in 0.1% formic acid. Peptides were then separated into 8 fractions by high pH reversed-phase chromatography as described above for the libraries.

To determine coefficients of variation for the Columbia cohort, urine from five donors in triplicates on one plate were subjected to sample preparation (intra-plate) and this was repeated on three different plates (inter-

plate). For the LCC cohort, urine from three donors in duplicates on one plate were subjected to sample preparation (intra-plate). Urine form nine other subjects were prepared on two different plates (inter-plate).

LC-MS/MS analysis

LC-MS/MS analysis was performed on an EASY-nLC 1200 coupled to a Q Exactive HF-X Orbitrap mass spectrometer via a nano-electrospray ion source (all Thermo Fisher Scientific). Purified peptides were separated at 60 °C on 50cm columns with an inner diameter of 75µm packed in-house with ReproSil-Pur C18-AQ 1.9µm resin (Dr.Maisch GmbH). Mobile phases A and B were 99.9/0.1% water/formic acid (v/v) and 80/20/0.1% acetonitrile/water/formic acid (v/v/v). For the LCC cohort, the flow rate was constant at 300 nl/min and the initial concentration of 5% B was linearly increased to 30% B within 36 minutes, and then increased further to 95% within 6 min with a 3 min plateau at the end. For the Columbia cohort, the flow rate was constant at 350 nl/min and the initial concentration of 5% B was linearly increased to 30% B within 35 minutes, and then increased further to 95% within 5 min with a 5 min plateau at the end.

MS data was acquired in the data-independent acquisition (DIA) scan mode for single-shot patient samples, using the MaxQuant Live software and spectral processing with phase-constrained spectrum deconvolution (phi-SDM) [85, 86]. 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. For the Columbia cohort, additionally two BoxCar scans with 12 isolation windows each and a resolution of 60,000 at m/z 200 were acquired [87] . Full MS events were followed by 33 MS/MS windows (LCC cohort) or 58 MS/MS windows (Columbia cohort) per cycle in the range of m/z 300–1,650 at a resolution of 15,000 at m/z 200. For the LCC cohort, higher-energy collisional dissociation MS/MS scans were acquired with a stepped normalized collision energy of 25/27.5/30 and ions were accumulated to reach an AGC target value of 3e6 or for a maximum of 30 ms. For the Columbia cohort, higher-energy collisional dissociation MS/MS scans were acquired with a normalized collision energy of 27 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 with a top12 data-dependent acquisition (DDA) scan mode. Full MS scans were acquired in the range of m/z 300–1,650 at a resolution of 60,000 (Columbia cohort) or 120,000 (LCC cohort) at m/z 200. The automatic gain control (AGC) target was set to 3e6. Higher-energy collisional dissociation MS/MS scans were acquired with a normalized collision energy of 27 at a resolution of 15,000 at m/z 200. Precursor ions with charge states of 2-7 were isolated in a 1.4 Th window and accumulated to reach an AGC target value of 1e5 or for a maximum of 60 ms. Precursors were dynamically excluded for 20 s after the first fragmentation event.

Mass spectrometry data processing

The MS data of the fractionated pools (DDA MS data, 24 neat pool urine and 8 EV fractions) and the single shot subject samples (DIA MS data, 165 and 132 samples in Columbia and LCC, respectively) were used to generate a DDA-library and direct-DIA-library, respectively, which were computationally merged into two cohort-specific hybrid libraries using Spectronaut version 13.9.191106.43655 (Biognosys AG). For all experiments except the machine learning, the two cohorts were quantified separately in Spectronaut. A minimum of 3 and a maximum of 10 fragments was required for each peptide in the library. The hybrid spectral libraries were subsequently used to search the MS data of the single shot patient samples in the Spectronaut software. 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. 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 cutoffs for both library generation and DIA analyses were set to 0.01. For generation of the global correlation map, the individual protein correlations with clinical parameters, and the machine learning, the Q-value data filtering setting in Spectronaut was set to 'Qvalue' to use every peptide passing the Q-value threshold for the protein group quantification. 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 'Qvalue percentile' setting results in a complete data matrix with no missing values, as the noise is quantified and reported if the peptide did not pass the Qvalue threshold.

Bioinformatics data analysis

The Perseus software package versions 1.6.0.7 and 1.6.1.3 and GraphPad Prism version 7.03 were used for the data analysis [88]. 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 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. The protein abundance levels were cross-correlated to generate a matrix of correlation coefficients. Unsupervised hierarchical clustering was performed using Perseus and proteins were clustered based on Pearson correlation scores. 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-), *GBA* status (only Columbia cohort LRRK2+ vs. LRRK2-), and PD status (only LRRK2+ vs. LRRK2-) were set as confounding factors. The FDR was set to 5% after Benjamini-Hochberg

correction. GO annotations were matched to the proteome data based on Uniprot and Ensemble 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-III and MoCA scores was performed in Perseus.

Machine learning

Data processing and machine learning was performed in Python (Version 3.7.3). Missing values were not imputed and protein intensities were normalized using the ScandardScaler method from the scikit-learn package (0.21.3). The XGBoost package (Version 0.90) was used to classify the samples and results were plotted using the bokeh library (2.1.1). Features were selected using a decision tree. Samples from both Columbia and LCC cohorts were used for the model and cross-validated using four stratified training/test splits and 15 repeats were applied. To assess sensitivity and specificity of the model, the results of the test sets were summed and averaged from 15 repeats.

ACKNOWLEDGEMENTS

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, visit www.michaeljfox.org/lcc .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, Christian Deiml and Alexander Strasser for helpful discussions and technical assistance. We further thank Hanno Steen, Dario Alessi and Suzanne Pfeffer and their group members for helpful discussion. We thank employees of the Michael J. Fox Foundation for Parkinson's research for helpful discussions.

FUNDING

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.

AUTHOR CONTRIBUTIONS

SVW and OK designed the experiments, performed, analyzed and interpreted all data. MTS helped with the machine learning. SP, MS, KM and RNA were responsible for sample collection and their distribution. SP, KM and RNA helped with interpretation of the results. MM supervised and guided the project, interpreted results and wrote the manuscript with SVW and OK.

COMPETING INTERESTS

The authors declare no competing interests.

REFERENCES

1. De Lau, L.M. and M.M.J.T.L.N. Breteler, *Epidemiology of Parkinson's disease.* 2006. **5**(6): p. 525-535.
2. Reeve, A., E. Simcox, and D.J.A.r.r. Turnbull, *Ageing and Parkinson's disease: why is advancing age the biggest risk factor?* 2014. **14**: p. 19-30.
3. Tysnes, O.-B. and A.J.J.o.N.T. Storstein, *Epidemiology of Parkinson's disease.* 2017. **124**(8): p. 901-905.
4. West, A.B.J.M.d., *Ten Y ears and C ounting: M oving L eucine-R ich Repeat K inase 2 I nhibitors to the C linic.* 2015. **30**(2): p. 180-189.
5. Steger, M., et al., *Systematic proteomic analysis of LRRK2-mediated Rab GTPase phosphorylation establishes a connection to ciliogenesis.* 2017. **6**: p. e31012.
6. Steger, M., et al., *Phosphoproteomics reveals that Parkinson's disease kinase LRRK2 regulates a subset of Rab GTPases.* 2016. **5**: p. e12813.
7. Karayel, O., et al., *Accurate MS-based Rab10 phosphorylation stoichiometry determination for LRRK2 activity in Parkinson's disease.* 2019: p. 819607.
8. Tolosa, E., et al., *LRRK2 in Parkinson disease: challenges of clinical trials.* 2020: p. 1-11.
9. Nalls, M.A., et al., *Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies.* 2019. **18**(12): p. 1091-1102.
10. Di Maio, R., et al., *LRRK2 activation in idiopathic Parkinson's disease.* 2018. **10**(451): p. eaar5429.
11. Alcalay, R.N., et al., *Higher urine bis (Monoacylglycerol) phosphate levels in LRRK2 G2019S mutation carriers: implications for therapeutic development.* 2020. **35**(1): p. 134-141.
12. Decramer, S., et al., *Urine in clinical proteomics.* 2008. **7**(10): p. 1850-1862.
13. An, M., Y.J.G. Gao, proteomics, and bioinformatics, *Urinary biomarkers of brain diseases.* 2015. **13**(6): p. 345-354.
14. Kentsis, A., et al., *Urine proteomics for profiling of human disease using high accuracy mass spectrometry.* 2009. **3**(9): p. 1052-1061.
15. Kentsis, A., et al., *Urine proteomics for discovery of improved diagnostic markers of Kawasaki disease.* 2013. **5**(2): p. 210-220.
16. Ferrari, E., et al., *Urinary proteomics profiles are useful for detection of cancer biomarkers and changes induced by therapeutic procedures.* 2019. **24**(4): p. 794.
17. Adachi, J., et al., *The human urinary proteome contains more than 1500 proteins, including a large proportion of membrane proteins.* 2006. **7**(9): p. R80.
18. Nguyen, M.T., et al., *Early prediction of acute renal injury using urinary proteomics.* 2005. **25**(4): p. 318-326.
19. Zimmerli, L.U., et al., *Urinary proteomic biomarkers in coronary artery disease.* 2008. **7**(2): p. 290-298.
20. Duangkumpha, K., et al., *Urine proteomics study reveals potential biomarkers for the differential diagnosis of cholangiocarcinoma and periductal fibrosis.* 2019. **14**(8).
21. Kaiser, T., et al., *Proteomics applied to the clinical follow-up of patients after allogeneic hematopoietic stem cell transplantation.* 2004. **104**(2): p. 340-349.
22. Metzger, J., et al., *Urine proteomic analysis differentiates cholangiocarcinoma from primary sclerosing cholangitis and other benign biliary disorders.* 2013. **62**(1): p. 122-130.

23. Tantipaiboonwong, P., et al., *Different techniques for urinary protein analysis of normal and lung cancer patients.* 2005. **5**(4): p. 1140-1149.
24. Ward, D.G., et al., *Proteomic profiling of urine for the detection of colon cancer.* 2008. **6**(1): p. 19.
25. Mischak, H., et al., *Proteomic analysis for the assessment of diabetic renal damage in humans.* 2004. **107**(5): p. 485-495.
26. Nagaraj, N. and M.J.J.o.p.r. Mann, *Quantitative analysis of the intra-and inter-individual variability of the normal urinary proteome.* 2011. **10**(2): p. 637-645.
27. Berger, S.T., et al., *MStern blotting–high throughput polyvinylidene fluoride (PVDF) membrane-based proteomic sample preparation for 96-well plates.* 2015. **14**(10): p. 2814-2823.
28. Batth, T.S., et al., *Protein aggregation capture on microparticles enables multipurpose proteomics sample preparation.* 2019. **18**(5): p. 1027-1035.
29. Kulak, N.A., et al., *Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells.* 2014. **11**(3): p. 319.
30. Ding, H., et al., *Urine Proteomics: Evaluation of Different Sample Preparation Workflows for Quantitative, Reproducible, and Improved Depth of Analysis.* 2020. **19**(4): p. 1857-1862.
31. Albrechtsen, N.J.W., et al., *Plasma proteome profiling reveals dynamics of inflammatory and lipid homeostasis markers after Roux-en-Y gastric bypass surgery.* 2018. **7**(6): p. 601-612. e3.
32. Geyer, P.E., et al., *Proteomics reveals the effects of sustained weight loss on the human plasma proteome.* 2016. **12**(12).
33. Geyer, P.E., et al., *Plasma proteome profiling to assess human health and disease.* 2016. **2**(3): p. 185-195.
34. Geyer, P.E., et al., *Plasma Proteome Profiling to detect and avoid sample-related biases in biomarker studies.* 2019. **11**(11).
35. Niu, L., et al., *Plasma proteome profiling discovers novel proteins associated with non-alcoholic fatty liver disease.* 2019. **15**(3).
36. Geyer, P.E., et al., *Revisiting biomarker discovery by plasma proteomics.* 2017. **13**(9).
37. Melachroinou, K., et al., *Elevated In Vitro Kinase Activity in Peripheral Blood Mononuclear Cells of Leucine-Rich Repeat Kinase 2 G2019S Carriers: A Novel Enzyme-Linked Immunosorbent Assay–Based Method.*
38. Alcalay, R.N., et al., *Glucocerebrosidase activity in Parkinson's disease with and without GBA mutations.* 2015. **138**(9): p. 2648-2658.
39. Ludwig, C., et al., *Data-independent acquisition-based SWATH-MS for quantitative proteomics: a tutorial.* 2018. **14**(8): p. e8126.
40. Gillet, L.C., et al., *Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis.* 2012. **11**(6).
41. Guo, Z., et al., *A proteomic analysis of individual and gender variations in normal human urine and cerebrospinal fluid using iTRAQ quantification.* 2015. **10**(7).
42. Uhlén, M., et al., *Tissue-based map of the human proteome.* 2015. **347**(6220): p. 1260419.
43. Jain, S.J.P. and r. disorders, *Multi-organ autonomic dysfunction in Parkinson disease.* 2011. **17**(2): p. 77-83.
44. Poewe, W.J.E.j.o.n., *Non-motor symptoms in Parkinson's disease.* 2008. **15**: p. 14-20.
45. Cook, C., C. Stetler, and L.J.C.S.H.p.i.m. Petrucelli, *Disruption of protein quality control in Parkinson's disease.* 2012. **2**(5): p. a009423.

46. Torsney, K.M., et al., *Bone health in Parkinson's disease: a systematic review and meta-analysis.* 2014. **85**(10): p. 1159-1166.
47. Ysselstein, D., et al., *LRRK2 kinase activity regulates lysosomal glucocerebrosidase in neurons derived from Parkinson's disease patients.* 2019. **10**(1): p. 1-9.
48. Ferrazza, R., et al., *LRRK2 deficiency impacts ceramide metabolism in brain.* 2016. **478**(3): p. 1141-1146.
49. Spitz, M., et al., *Association of LRRK2 and GBA mutations in a Brazilian family with Parkinson's disease*. 2015. **21**(7): p. 825-826.
50. Duran, R., et al., *The glucocerobrosidase E326K variant predisposes to Parkinson's disease, but does not cause Gaucher's disease*. 2013. **28**(2): p. 232-236.
51. Yahalom, G., et al., *Carriers of both GBA and LRRK2 mutations, compared to carriers of either, in Parkinson's disease: Risk estimates and genotype-phenotype correlations*. 2019. **62**: p. 179-184.
52. Chen-Plotkin, A.S., et al., *Finding useful biomarkers for Parkinson's disease.* 2018. **10**(454).
53. Fields, C.R., N. Bengoa-Vergniory, and R.J.F.i.M.N. Wade-Martins, *Targeting alpha-synuclein as a therapy for Parkinson's Disease.* 2019. **12**.
54. Bader, J.M., et al., *Proteome profiling in cerebrospinal fluid reveals novel biomarkers of Alzheimer's disease.* 2020. **16**(6): p. e9356.
55. Shao, C., et al., *Comprehensive analysis of individual variation in the urinary proteome revealed significant gender differences*. 2019. **18**(6): p. 1110-1122.
56. Oh, Y., et al., *Cyclophilin B protects SH-SY5Y human neuroblastoma cells against MPP+-induced neurotoxicity via JNK pathway*. 2016. **478**(3): p. 1396-1402.
57. M Fayaz, S., et al., *CypD: the key to the death door*. 2015. **14**(5): p. 654-663.
58. Handa, K., et al., *Bone loss caused by dopaminergic degeneration and levodopa treatment in parkinson's disease model mice.* 2019. **9**(1): p. 1-16.
59. Paul, G. and A.M.J.E.J.o.N. Sullivan, *Trophic factors for Parkinson's disease: Where are we and where do we go from here?* 2019. **49**(4): p. 440-452.
60. Henderson-Smith, A., et al., *Next-generation profiling to identify the molecular etiology of Parkinson dementia.* 2016. **2**(3): p. e75.
61. Cocco, C., et al., *VGF peptides as novel biomarkers in Parkinson's disease.* 2020. **379**(1): p. 93-107.
62. Cocco, C., et al., *Distribution of VGF peptides in the human cortex and their selective changes in Parkinson's and Alzheimer's diseases.* 2010. **217**(6): p. 683-693.
63. Pasinetti, G., et al., *Identification of potential CSF biomarkers in ALS.* 2006. **66**(8): p. 1218-1222.
64. Carrette, O., et al., *A panel of cerebrospinal fluid potential biomarkers for the diagnosis of Alzheimer's disease.* 2003. **3**(8): p. 1486-1494.
65. Emamzadeh, F.N.J.J.o.M.N., *Role of Apolipoproteins and α-Synuclein in Parkinson's Disease.* 2017. **62**(3-4): p. 344-355.
66. Wang, E.S., et al., *Tetranectin and apolipoprotein A-I in cerebrospinal fluid as potential biomarkers for Parkinson's disease.* 2010. **122**(5): p. 350-359.
67. Swanson, C.R., et al., *Plasma apolipoprotein A1 associates with age at onset and motor severity in early Parkinson's disease patients.* 2015. **30**(12): p. 1648-1656.
68. Qiang, J.K., et al., *Plasma apolipoprotein A1 as a biomarker for Parkinson disease*. 2013. **74**(1): p. 119-127.

69. Borghini, I., et al., *Characterization of subpopulations of lipoprotein particles isolated from human cerebrospinal fluid.* 1995. **1255**(2): p. 192-200.
70. Koch, S., et al., *Characterization of four lipoprotein classes in human cerebrospinal fluid.* 2001. **42**(7): p. 1143-1151.
71. Petit-Turcotte, C., et al., *Apolipoprotein CI expression in the brain in Alzheimer's disease.* 2001. **8**(6): p. 953-963.
72. Alessi, D.R. and E.J.S. Sammler, *LRRK2 kinase in Parkinson's disease.* 2018. **360**(6384): p. 36-37.
73. !!! INVALID CITATION !!! [60-62].
74. Henry, A.G., et al., *Pathogenic LRRK2 mutations, through increased kinase activity, produce enlarged lysosomes with reduced degradative capacity and increase ATP13A2 expression.* 2015. **24**(21): p. 6013-6028.
75. Biskup, S., et al., *Dynamic and redundant regulation of LRRK2 and LRRK1 expression.* 2007. **8**(1): p. 102.
76. Berndsen, K., et al., *PPM1H phosphatase counteracts LRRK2 signaling by selectively dephosphorylating Rab proteins.* 2019. **8**.
77. Plotegher, N., et al., *Ceramides in Parkinson's disease: from recent evidence to new hypotheses.* 2019. **13**: p. 330.
78. Lin, G., et al., *Sphingolipids in the Pathogenesis of Parkinson's Disease and Parkinsonism.* 2019. **30**(2): p. 106-117.
79. Yamada, M., et al., *Furin inhibitor protects against neuronal cell death induced by activated NMDA receptors.* 2018. **8**(1): p. 1-9.
80. Maksoud, E., E.H. Liao, and A.P.J.C.r. Haghighi, *A neuron-glial trans-signaling cascade mediates LRRK2-induced neurodegeneration.* 2019. **26**(7): p. 1774-1786. e4.
81. Xu, J.-C., et al., *The extracellular matrix glycoprotein tenascin-R regulates neurogenesis during development and in the adult dentate gyrus of mice.* 2014. **127**(3): p. 641-652.
82. Morawski, M., et al., *ECM in brain aging and dementia*, in *Progress in brain research*. 2014, Elsevier. p. 207-227.
83. Farlow, J.L., et al., *Whole-exome sequencing in familial Parkinson disease.* 2016. **73**(1): p. 68-75.
84. Kulak, N.A., et al., *Loss-less nano-fractionator for high sensitivity, high coverage proteomics.* 2017. **16**(4): p. 694-705.
85. Wichmann, C., et al., *MaxQuant. Live enables global targeting of more than 25,000 peptides.* 2019. **18**(5): p. 982-994.
86. Grinfeld, D., et al., *Phase-constrained spectrum deconvolution for Fourier transform mass spectrometry.* 2017. **89**(2): p. 1202-1211.
87. Meier, F., et al., *BoxCar acquisition method enables single-shot proteomics at a depth of 10,000 proteins in 100 minutes.* 2018. **15**: p. 440-448.
88. Tyanova, S., et al., *The Perseus computational platform for comprehensive analysis of (prote) omics data.* 2016. **13**(9): p. 731.