Title

Parkinson's disease causality and heterogeneity: a proteogenomic view

Authors

Sergio Kaiser*[1], Luqing Zhang*[2], Brit Mollenhauer[3], Jaison Jacob[2¶], Simonne Longerich[5], Jorge Del-Aguila[5], Jacob Marcus[5], Neha Raghavan[5], David Stone[6], Olumide Fagboyegun[6], Douglas Galasko[7], Mohammed Dakna[3], Bilada Bilican[4$], Mary Dovlatyan[4], Anna Kostikova[1], Jingyao Li[4], Brant Peterson[4#], Michael Rotte[1], Vinicius Sanz[4], Tatiana Foroud[8], Karla Gonzalez[8], Samantha J. Hutten[9], Mark Frasier[9], Hirotaka Iwaki[10], Andrew Singleton[10], Ken Marek[11], Karen Crawford[12], Fiona Elwood[4§], Mirko Messa[4], Pablo Serrano-Fernandez[1]

*(contributed equally)

Affiliations

[1] Translational Medicine Department. Novartis Institutes for Biomedical Research. Basel, Switzerland.
[2] Cardiovascular and Metabolism Department. Novartis Institutes for Biomedical Research. Cambridge, USA.
[3] Department of Neurology, University Medical Center Göttingen. Göttingen, Germany.
[4] Neuroscience Department. Novartis Institutes for Biomedical Research. Cambridge USA.
[5] Genome and Biomarker Sciences. Merck Exploratory Science Center. Cambridge, USA.
[6] Department of Genetics, Cerevel Therapeutics. Cambridge, USA.
[7] Department of Neurosciences, University of Southern California, San Diego. La Jolla, USA.
[8] Department of Medical and Molecular Genetics, Indiana University School of Medicine. Indianapolis, USA.
[9] Michael J. Fox Foundation for Parkinson's Research. New York, USA.
[10] Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health. Bethesda, USA.
[11] Institute for Neurodegenerative Disorders. New Haven, USA.
[12] Laboratory of Neuroimaging, University of Southern California. Los Angeles, USA.

[¶]Current address: Moderna Genomics, Cambridge, USA
[$]Current address: Translational Genomics, Discovery Sciences BioPharmaceuticals R&D, AstraZeneca. Gothenburg, Sweden.
[#]Current address: Valo Health. Cambridge, USA.
[§]Current address: The Janssen Pharmaceutical Companies of Johnson & Johnson, Cambridge, USA

Corresponding authors

Mirko Messa, Pablo Serrano-Fernandez

Keywords

GBA, LRRK2, SomaScan, proteomics, CSF, personalized medicine, stratification, network analysis, endotype, causal analysis, mendelian randomization, pQTL, GWAS, neurodegeneration, prediction

Abstract

Pathogenesis and clinical heterogeneity in Parkinson's disease (PD) have been evaluated from genetic, pathological, and clinical perspective. Technology allowing for high-throughput proteomic analyses in cerebrospinal fluid (CSF) has opened new opportunities to scrutinize this heterogeneity. This is to date the most comprehensive proteomics profiling in CSF of PD patients and controls, both in terms of subjects (n=1103) and proteins (n=4135).
Combining CSF aptamer-based proteomics with genetics across all samples we determined the protein quantitative trait loci (pQTLs) linking genetic variants with protein abundance in CSF. Analyzing pQTLs together with summary statistics from the largest PD genome wide association study (GWAS) led us to identify 68 potential causal proteins by Mendelian randomization. The top PD causal protein, GPNMB, has colocalization support and has been reported to be upregulated in the substantia nigra of PD patients.
We also examined three subcohorts of PD patients: *LRRK2* variant carriers (*LRRK2+*), *GBA* variant carriers (*GBA+*) and idiopathic PD patients, each with their respective controls. The second goal was to identify proteomic differences between PD patients and controls within and between the three subcohorts. Statistical analyses revealed six proteins differentially expressed when comparing *GBA+* PD patients with unaffected *GBA+* controls, seven proteins when comparing *LRRK2+* PD patients with unaffected *LRRK2+* controls and 23 proteins when comparing idiopathic PD patients with healthy controls who did not carry any severe PD mutations.
Furthermore a hypothesis-free stratification of idiopathic PD patients based on their proteomic profile revealed two protein modules based on the co-expression network structure. Based on these modules, cluster analysis revealed two patient endotypes for idiopathic PD.
Differences in the CSF proteomic signature between subcohorts and between idiopathic endotypes, as well as causal targets identified using both proteomics and genetics may together influence the way we approach the identification of potential therapeutic targets in PD.

Introduction

Parkinson's disease (PD) is the second most prevalent neurodegenerative disorder with more than six million patients affected worldwide [1] (Global Burden of Disease 2015 Neurological Disorders Collaborator Group, 2017). Although symptomatic treatments are available, to date there are no disease modifying therapies, making it a major unmet medical need. PD is a clinically and pathologically heterogeneous disorder [2] (Jagadeesan et al. 2017). While PD is well known for its motor symptoms such as resting tremor, rigidity and bradykinesia, PD patients also suffer from a broad variety of non-motor and/or neurobehavioral symptoms including cognitive impairment, depression and anxiety to autonomic symptoms such as constipation, and sleep abnormalities.
PD patients experience selective degeneration and loss of dopaminergic neurons in the substantia nigra *pars compacta*. In most cases, PD is classified as idiopathic, but a growing set of genetic variants increase PD risk or accelerate its onset. Many of the identified genes are involved either

in mitochondrial or endo-lysosomal biology [3] (Blauwendraat et al. 2020a). The two most common PD risk genes are leucine rich kinase 2 (*LRRK2*) and glucosidase beta acid (*GBA*). Mutations in these genes are linked to ~10% of sporadic cases and up to 30% in specific ethnic subgroups and familial disease [4] (Bonifati 2014). Some of the genetic variants in *LRRK2* and *GBA* have been associated with specific clinical phenotypes. The dominant phenotype of PD patients carrying a *LRRK2* severe mutation is tremor, postural instability and/or gait disorder. They are also less likely to have olfactory and cognitive impairment as well as REM sleep behavior disorder [5] (Kestenbaum and Alcalay 2017). Severe *GBA* mutations, in turn, are associated with worse and faster progressing motor, cognitive, olfactory and psychiatric symptoms as compared to mild mutation carriers or idiopathic patients [6] (Thaler et al. 2018). Not to forget that, for both genes, the pathological mutations are thought to exacerbate the toxicity of alpha-synuclein, which – in an aggregated form – contributes to neuronal death and amplifies the neuroinflammatory response.

Clinical heterogeneity of PD has motivated many disease stratification efforts. Some of those have focused on clinical variables, mostly hypothesis-driven, while others have focused on molecular data, mostly hypothesis-free [7] (Qiang and Huang 2019). An association between such strata and the likelihood of carrying a mutation associated with PD risk remains elusive [8] (Ma et al. 2015).

Proteins, largely comprising the ultimate biological effectors, hold great potential as predictors, causal and/or disease-modifying targets and surrogates of disease progression and/or stratification. However, the biological and pathophysiological complexity of PD, the difficulties of collecting large amounts of standardized biological samples (especially cerebrospinal fluid; CSF) from large cohorts throughout the course of disease, and the technical limitations of high-throughput proteomic analyses hamper the identification of biomarkers at a proteomic level. The multicenter Parkinson Progression Marker Initiative (PPMI) was specifically initiated to overcome some of the limitations, particularly in terms of number of samples and clinical data, from previous single-center cohort studies [9] (Parkinson Progression Marker Initiative, 2011). In this collaboration, 1103 baseline (not longitudinal) CSF samples from PD and control participants with known status of *LRRK2* and *GBA* pathogenic variants were analyzed using the SomaScan® aptamer-based proteomics platform [10] (Gold et al. 2010). CSF is biologically closer to the brain tissue than other biological matrices and its proteomics analysis may capture more detail than the equivalent analysis in serum, plasma, or other peripheral fluids.

The main goals of this study are summarized in Figure 1: PD causal protein identification using mendelian randomization based on proteomics and genetics, identification of differences between PD patients and controls within and between subcohorts (*LRRK2*+, *GBA*+ and idiopathic), and hypothesis-free stratification of idiopathic PD patients based on their proteomic profile..

Identifying causal genes or proteins for PD by combining both the genomic and proteomic information available in the PPMI cohort builds on the substantial progress of PD genome wide association studies (GWAS) achieved during the past decades. The PPMI data base contains CSF proteomics and whole genome sequencing data from 804 patients, after quality control. To our knowledge, this is to date the largest proteomic and genetic data set for interrogating causal proteins for PD in a neurologically relevant biofluid.

Some genetic variants in *LRRK2* and *GBA* have been associated with specific clinical phenotypes. The dominant phenotype of PD patients carrying a *LRRK2* severe mutation is tremor, postural instability and/or gait disorder. They are also less likely to have olfactory and

cognitive impairment as well as REM sleep behavior disorder [5] (Kestenbaum and Alcalay 2017). In contrast, severe *GBA* mutations are associated with worse and faster progressing motor, cognitive, olfactory, and psychiatric symptoms as compared to mild mutation carriers or idiopathic patients [6] (Thaler et al. 2018). Therefore the interest in analyzing the PPMI cohort not only as a whole at proteomic level, but also separated by idiopathic and genetic subcohorts. In addition, we present a first attempt to stratify PD patients based on high-throughput CSF proteomics.

Methods

Study design

The clinical data and samples used in this study were obtained from the PPMI [http://www.ppmi-info.org/data] on October 1, 2020. PPMI samples were collected under a standardized protocol over 33 centers and includes clinical and imaging data as well as plasma and CSF samples. Study protocol and manuals are available online [http://www.ppmi-info.org/ study-design].
Separate subcohorts of patients with PD and their respective controls were enrolled following inclusion and exclusion criteria [11] (Marek et al. 2018). One subcohort is comprised of recently diagnosed, drug-naïve, idiopathic PD patients and healthy controls, while the second and third subcohorts are comprised of PD patients, carriers of a severe *GBA* or *LRRK2* mutation, either PD patients or unaffected controls. PD patients form these genetic subcohorts had a higher disease duration, were partially under PD medication (n=203), were over-represented for individuals of Ashkenazi Jewish descent and differed by sex distribution from the idiopathic PD patients (higher proportion of women in the genetic). The study was approved by the Institutional Review Board at each site, and participants provided written informed consent.
Genetic testing was done by the centralized PPMI genetic testing core. Non-manifesting carriers received pre-testing and post-testing genetic counselling by phone from certified genetic counsellors at the University of Indiana or site-qualified personnel. The *LRRK2* genetic testing battery includes G2019S and R1441G mutations. *GBA* genetic testing includes N370S (for all participants), and L483P, L444P, IVS2+1, and 84GG (for a subset of participants) mutations. Dual mutation carriers *(LRRK2* and *GBA)* were considered as *LRRK2* carriers for simplicity (n=1).
Six patients were diagnosed as idiopathic PD at enrolment but were re-classified during follow-up (two patients were diagnosed as multiple system atrophy and four patients did not have a final diagnose but PD had been excluded). These patients were removed from the analysis. Four patients were initially diagnosed as genetic PD, but the diagnose changed to prodromal during follow up. These patients were considered as unaffected controls in their corresponding genetic subcohort (Five *LRRK2*+ unaffected controls and one *GBA*+ unaffected control).
One subject originally classified as healthy, but later shown to have an unclear health status, was removed from the analysis. Subjects recruited into the subcohort of idiopathic PD patients and healthy controls but identified as carriers of a severe *GBA* or *LRRK2* mutation, were moved to the corresponding genetic subcohort (*GBA* n=15; *LRRK2* n=7).
The genetic screening also detected *GBA* mutations of unknown or moderate risk: A459P, E365K, T408M. Carriers of these mutations were removed from analysis (n=38).
Finally, 10 carriers of a mutation in *SCNA* (8 PD patients and 2 unaffected controls), were also removed due to lack of statistical power for analysis.
The original data set was comprised of 1190 samples out of which 32 samples were pools, which were discarded for this study, and as described above, additional six PD patients and one healthy

control were removed due to change in diagnose, 38 subjects were removed due to non-severe *GBA* mutations and 10 patients were removed for being carriers of a mutation in *SCNA*. Hence, the final data set used for analysis was comprised of 1103 proteomic samples from an equal number of subjects (no replicates) divided into three subcohorts: no mutation (idiopathic PD patients and healthy controls with no severe mutation in *GBA* or *LRRK2*), *GBA+* (PD patient carriers of severe *GBA* mutations and unaffected controls carrying the same mutations) and *LRRK2+* (PD patients carriers of severe *LRRK2* mutations and unaffected controls carrying the same mutations). The exact composition is summarized in Table 1. CSF samples from these patients were distributed from the biorepositories of the PPMI study and analyzed for proteomics.

We make an explicit distinction between "healthy" and "unaffected" controls, because there is evidence prodromal pathophysiology in *LRRK2+* and *GBA+* controls when compared to healthy controls that are non-mutation carriers [12] (Simuni et al. 2020).

Proteomics

Proteomic profiling was performed using SomaScan® in a platform version that is proprietary to Novartis and includes 4785 SOMAmers® (Slow off-rate modified aptamers) targeting 4135 human proteins. SOMAmer levels in CSF for the PPMI data set were determined at SomaLogic Inc. (Boulder, US).

SomaScan data undergo a standardization process at the vendor that includes Hybridization Normalization (controls for variability in the readout of individual microarrays), Plate Scaling (accounts for plate-by-plate variation), Median Signal Normalization (controls for total signal differences between individual samples) and Calibration (removes the variation between assay runs within and across experiments).

Relative fluorescence units are transformed to log2 scale, normalized to the median separately by dilution level across all plates. Finally, the data set is adjusted for batch effects between plates using an empirical Bayes method as implemented in the R package *sva* [13] (Leek et al. 2020).

Genetics

PPMI whole genome sequencing results were lifted over to hg19 coordinates. Biallelic SNPs on autosomes were extracted. Standard GWAS quality control (QC) was applied at both individual and SNP level. 22 patients with outlying heterozygosity and 93 patients with high identity-by-descent were excluded after QC. 306031 SNPs were removed due to missing genotype and 35907596 SNPs removed due to minor allele count less than 20. Finally, 9743041 variants and 1264 subjects passed QC.

pQTL calculation

Among the 1264 subjects who passed QC for genetics and the 1103 who passed QC for proteomics, 804 subjects overlapped. Protein expression values were ranked and inverse normal transformed. For pQTL calculation, each protein level was regressed with each independent genetic variant (SNP; MAF > 0.05), adjusted for age, sex, subcohort, protein principal components 1-4 and genetic principal components 1-10 using R package *MatrixEQTL* [14] (Shabalin and Andrey 2012). Cis-pQTLs were defined as SNPs located inside the +/- 1Mb region flanking the gene that encodes the given protein. A genome-wide threshold of $p < 5e-8$ defined a significant cis-pQTL.

Causal Analysis

The two-sample mendelian randomization method implemented in the R package *TwoSampleMR* [15,16] (Hemani et al. 2017; Hemani et al. 2018) was applied to find causal proteins for PD in CSF.

Although PPMI is a well-controlled PD study with genetic data, in order to avoid weak instrumental variable bias and take the advantage of a larger PD GWAS, we relied on the meta-analysis from Nalls et al [17]. (2019), which includes 17 datasets with PD cases ranging from 363 to 33674 individuals, and healthy controls ranging from 165 to 449056 individuals. Instrumental variables were selected for each SNP with MAF > 0.05, F-statistics larger than 10 and significant cis-pQTL $p < 5e-8$. Shared SNPs in both cis-pQTL and PD GWAS were harmonized and then clumped using an LD threshold of either $r^2<0.01$ or $r^2<0.3$. The more stringent LD threshold of $r^2<0.01$ resulted in only one instrumental variable for most proteins, therefore the results we present are from the less stringent threshold of $r^2<0.3$. The Wald ratio was used when only one instrument survived clumping, while the inverse variance weighted meta-analysis method was used when more than one instrumented SNP was available. Horizontal pleiotropy was tested using the R package *MRPRESSO* [18] (Verbanck et al. 2018).
Colocalization probability was calculated using the R package *coloc* [19] (Giambartolomei et al. 2014). Default priors of $p_1=10^{-4}$, $p_2=10^{-4}$, and $p_{12}=10^{-5}$ were used, where p_1 is the prior probability of a SNP being associated with PD, p_2 is the prior probability of a SNP being associated with CSF pQTL, and p_{12} is the prior probability of a SNP being associated with both PD and CSF pQTL. We considered $PPH_4 > 0.75$ as strong evidence for colocalization. PPH_4 is the posterior probability of one shared SNP being associated with both PD and CSF pQTL.

Clinical variables

The clinical assessment battery is described on the PPMI website (www.ppmi-info.org). In particular, PD status was assessed with the Unified Parkinson's Disease Rating Scale in the revised version published by the Movement Disorder Society (MDS-UPDRS) scores 1, 2, and 3 [20] (Goetz et al 2008). Cognitive testing comprised screening with the Montreal Cognitive Assessment (MoCA) [21] (Nasreddine et al. 2005). High resolution xy-weighted 3 tesla MRI was available for 545 PD patients and 177 controls. Caudate, putamen and striatum thicknesses were calculated as the arithmetical mean between the measure in the right and left hemisphere, respectively.
CSF was collected using standardized lumbar puncture procedures. Sample handling, shipment and storage were carried out as described in the PPMI biologics manual (http://ppmi-info.org). Besides the SomaScan analysis described earlier, data from commercially available sandwich type immunoassay kits were also used for measuring CSF total alpha-synuclein, amyloid-beta 1-42, total tau and phospho-tau (p-tau 181) protein as described previously [22,23] (Kang et al. 2013; Mollenhauer et al. 2019). Phospho-tau was measured with the Elecsys® assay run on the fully automated Roche cobas® system.
At baseline, all participants with idiopathic PD were free of PD-related medications (drug-naïve). Use of medications for PD was recorded at each visit after baseline assessment. For simplicity, we used this as a binary variable (PD medication present / absent).

Statistical analysis

We compared PD patients with controls within each subcohort (idiopathic, *GBA*+ and *LRRK2*+). A linear regression model was applied using the Bioconductor R package *limma* (Ritchie et al. 2015). The model included the following covariates: age, sex, study center and principal components 1-4. The genetic cohorts also included treatment (yes/no) as a covariate in the model to exclude treatment effects. Note that only patients in the genetic subcohorts might have been under treatment.

Additionally, a linear model with an interaction term was tested. The interaction term was between the disease status (control / PD) and the mutation status (no mutation / *GBA+* / *LRRK2+*). The covariates were the same as in the models above.
Comparisons of clinical variables between endotypes of idiopathic patients were performed using a chi2 test for categorical variables or a Mann-Whitney U-test for quantitative variables.
Predictive modeling for the idiopathic classes was performed using a partition tree with pruning as implemented in the R package *rpart* [24] (Therneau and Atkinson 2019). The model was defined on a training set (70% of the idiopathic PD patients) and tested on an independent test set (30% of the idiopathic PD patients). The pruning was based upon a 10-fold cross validation, the default for *rpart*.
All p-values were adjusted for multiple testing using false discovery rate (FDR).

Cluster analysis

Network analysis of the CSF proteome of idiopathic PD patients was carried out using the R package *WGCNA* [25] (Langfelder and Horvath, 2008). Co-expressed proteins (SOMAmers) were grouped into modules.
Consensus clustering as implemented in the R package *ConsensusClusterPlus* [26] (Wilkerson et al. 2010) used the SOMAmer modules to identify idiopathic PD patient subclasses. To avoid confounders being responsible for the differences between patient subclasses, network analysis was performed on the SOMAmer residuals of a linear regression on age, sex and study center.
Heatmaps were generated using the R package *Heatplus* [27] (Ploner 2015).
For completeness, a second – orthogonal and complementary – analytical strategy was used to identify robust idiopathic PD patient subgroups. Instead of classifying based on proteomics, we classified the patients based on clinical variables and then analyzed the overlap with the previous approach. The clinical variables used were variables relevant to PD diagnosis and progression: imaging data (caudate, putamen and striatum thickness), UPDRS scores 1, 2 and 3, MoCA and CSF levels of amyloid beta, alpha-synuclein, phospho-tau and total tau (not SomaScan data, but regular ELISA-based clinical assays).
A k-means clustering approach was applied on a uniform manifold approximation and projection (UMAP) of the variables mentioned above [28] (R package *umap*, Konopka 2020). The optimal number of clusters was determined by maximal average silhouette width as implemented in the R package *factoextra* [29] (Kassambara and Mundt 2020).

Results

Causal analysis

To establish what proteins could be causal in the different PD subcohorts we combined the PPMI whole genome sequencing and SomaScan proteomics data by performing Mendelian randomization. Our analysis reported significant cis-pQTLs for 856 SOMAmers – corresponding to 744 unique proteins (Supplementary Table 5). From these 856 significant cis-pQTLs, we identified statistically significant evidence for causation for 68 proteins in CSF (Table 2). The full table with nominal p-value smaller than 0.05 is included in Supplementary Table 1. Out of those proteins, GPNMB, FCGR2A and FCGR2B had a strong colocalization signal (see Methods), indicating the same SNP is both associated with protein level and PD risk (Figure 2 and Supplementary Table 1). When we used a more stringent clumping threshold ($r^2<0.01$) GPNMB, HLA-DAQ2 and FCGR2A passed FDR correction for significance. The full table with nominal p-value less than 0.05 are listed in Supplementary Table 2.

Differential protein expression in subcohorts of PD patients

To identify proteins differentially expressed in PD, we compared SOMAmers in each of the subcohorts (*GBA+*, *LRRK2+* and idiopathic) to their corresponding controls. Our statistical analyses revealed six differentially expressed SOMAmers for *GBA+* PD, seven SOMAmers for *LRRK2+* PD and 23 SOMAmers for idiopathic PD. Directionality of the change and adjusted p-values for each of these markers are reported in Table 3.

For each subcohort, several identified markers confirmed previously reported proteins dysregulated in PD. Interestingly, there was little overlap between proteins dysregulated in *GBA+*, *LRRK2+* and idiopathic PD subcohorts (SEMG2 and DLK1 were shared by *GBA+* and the idiopathic subcohort) though in each list there is a high percentage (4/6 in *GBA+*, 4/7 in *LRRK2+* and 10/23 in the idiopathic subcohort) of markers previously reported in relation to PD (Table 3).

Only one protein, CTSB, passed the FDR significance threshold for the interaction between disease status and mutation status. As shown above, CTSB was also differentially expressed in the *LRRK2+* PD subcohort and in the idiopathic PD subcohort.

Additional analysis comparing all PD subcohorts with all controls, using subcohort membership (no mutation/*GBA+*/*LRRK2+*) and treatment status (yes/no) as additional covariates resulted in 129 SOMAmers, tagging 122 distinct proteins, passing FDR correction (Supplementary Table 3).

Identification of subtypes of idiopathic PD patients

Identification of endotypes. To determine if any distinct endotypes were present in the idiopathic subcohort we performed a network analysis on the CSF proteome of idiopathic PD patients. Two modules of co-expressed proteins were identified. They comprised 889 and 600 SOMAmers, respectively. Applying consensus clustering on these two protein modules split the idiopathic PD subcohort (n=350 patients) into two proteome-based patient subclasses: endotype 1 (n=185 patients) and 2 (n=165 patients) (Figure 3A). As seen in the tracking plot (Figure 3B) these endotypes suffer only negligible changes as the number of modeled subclasses increases.

Identification of phenotypes. The stratification of idiopathic PD patients based on clinical variables showed an almost orthogonal alignment between imaging data and UPDRS scores on one side, and CSF levels of amyloid beta, alpha-synuclein, phospho-tau and total tau on the other side (Figure 4A).

The optimal number of clusters on the UMAP projection was calculated to be two (Figure 4B). A k-means approach for two clusters – here referred as phenotypes 1 and 2 for convenience – created a split as seen in Figure 4C. In Figure 4D we can see the endotypes as defined by the previous approach based on proteomics as projected on the same UMAP layout, for comparison.

Relationship between endotypes and phenotypes. With an accuracy of 0.761, the overlap between phenotypes and endotypes is a remarkably good one. Moreover, a predictive model for the endotypes was built based on clinical parameters, avoiding the re-use of the same proteomic data involved in the definition of the endotypes. Patients with CSF phospho-tau ≥ 11pg/mL (as measured with the Elecsys® assay) were enriched for endotype 2, and patients with CSF phospho-tau < 11pg/mL were enriched for endotype 1 (Figure 3C). The model accuracy in the training (n=244 patients) and in the independent test (n=106 patients) sets, was 0.82 and 0.73, respectively. The estimated area under the curve (AUC) for the test set was 0.77.

There were no significant differences between patient endotypes in MRI-defined caudate, striatum or putamen thickness, UPDRS scores or MoCA scores. Significant differences were

limited to CSF levels of amyloid beta, phospho-tau and alpha-synuclein, all of them lower for endotype 2 (Figure 5). Endotypes did not significantly differ in age or sex.

Proteins differentially expressed in endotypes (CSF SomaScan). To identify the unique proteins significantly dysregulated in each endotype, a linear model was used to identify the differences between each of these two endotypes and the healthy controls to the idiopathic PD subcohort. For endotype 1, five markers were significantly different compared to the control group (CNTFR, LPO, MMP10, RIPK2 and VEGFA). LPO, RIPK2 and VEGFA were also part of the differences between healthy controls and the whole idiopathic group (see above). Endotype 2, however, showed 200 differentially expressed SOMAmers, 197 unique proteins (see Supplementary Table 4). Among those proteins, AK1, CCL14, FRZB, GPI, HAMP, LPO, NETO1, PTPRR, RAB31, RELT, RIPK2, ROBO3, RSPO4, SHANK1, SPINK9, VEGFA and VIP were dysregulated for the whole idiopathic group as compared to healthy controls. Differentially expressed SOMAmers between endotypes added up to 155, 153 unique proteins (see Supplementary Table 4).

Discussion

The clinical, pathological and genetic variability observed in PD poses a severe challenge in the development of disease modifying therapies to improve the life of patients worldwide. In this study, by using the PPMI patient cohort, we combined genetic and proteomic approaches to build more foundations to help classifying PD patients by their molecular signatures.
Several proteins could be identified pointing towards causal relationship with PD, when combining all subcohorts. At the same time, results indicate that the pathophysiology of PD might differ on a molecular level between patient subgroups defined by *GBA+* and *LRRK2+* PD as well as two separate endotypes of idiopathic PD as derived from their proteomic profiles. While there are neurodegeneration markers in all subgroups, these markers differ across comparison groups.

PD causal proteins

Identification of causal proteins for a specific indication is extremely important when looking at new targets. 68 causal proteins were identified as significant in our PPMI CSF sample population with the top hit being GPNMB (aka HGFIN, Osteoactivin).
This gene encodes for a transmembrane glycoprotein broadly expressed through most cell types with high expression in macrophages and microglia [30] (Saade et al. 2021). It can be found in the plasma membrane, in intracellular compartments such as endosomes and lysosomes [31] (Tomihari et al. 2009) or secreted after cleavage [32] (Rose et al. 2010). Functionally, GPNMB has been reported to be involved in differentiation and maturation of different cell types [33-35](Abdelmagid et al. 2008, Furochi et al. 2007, Bandari et al. 2003) and most recently it has been associated with inflammation both in the brain and in the periphery [30] (Saade et al. 2021), although its specific role is not yet fully understood. Several studies reported GPNMB being impacted during neurodegeneration. While protein levels have been found to be elevated specifically in the substantia nigra of PD patients [36] (Moloney et al. 2018), the dysregulation is not restricted to PD. An increase in both mRNA and protein levels have been also detected in preclinical models and clinical samples of Alzheimer's disease, amyotrophic lateral sclerosis and Gaucher's disease [37-39] (Hüttenrauch et al. 2018, Tanaka et al. 2012, Kramer et al. 2016).
In PD, increase in GPNMB levels has been associated with deficiency in *GBA* activity. The elevated levels of GPNMB in the substantia nigra of PD patients correlates with an increase of

glycosphingolipids in the same area, a phenomenon due to decreased GCase activity [36,40] (Moloney et al. 2018, Rocha 2015a). Interestingly, in rodent models mimicking synucleopathy the levels of GPNMB were not affected unless GCase activity was impaired, supporting the conclusion that GPNMB levels are impacted by the lipid load at lysosomal level and not by alpha-synuclein toxicity alone. Hence, increased GPNMB levels may reflect the inflammatory status in the brain of PD patients and at the same time could be relevant in the context of PD lipidopathy.

The increased protein levels, together with the high expression of GPNMB in a variety of immune cells [41] (Tanaka et al. 2012), point to a role for this protein as a regulator of inflammation in PD and other neurodegenerations [38,42] (Ahn et al. 2002; Ono et al. 2016). GPNMB could halt inflammation probably via its ability of binding NKA, a known regulator of neuroinflammation [43] (Kinoshita et al. 2017). Nevertheless, the mechanism of action has not been fully characterized yet and there are evidence suggesting GPNMB may have a pro-inflammatory role thus an increase in its levels could in the end exacerbate the neuronal damage and neurodegeneration [44] (Shi et al. 2014). Based on this hypothesis Brendza et al. recently tested if genetic ablation of GPNMB has any protective effects in PD mouse models [45] (Brendza et al. 2021). As in our results, they also observed increased GPNMB protein levels in PD patients; specifically, in the substantia nigra (both in microglia and neurons). Nevertheless, knocking out *GPNMB* didn't show neuroprotection in the rodent models used by the authors (synucleinopathy and demyelination mouse models).

Also, in our proteomics results, GPNMB is one of the 122 proteins differentially expressed between PD and controls, for all three subcohorts combined (see Supplementary Table 3).

A second lysosomal protein, Cathepsin B (CTSB), was also identified as a causal protein in our analysis. CTSB can cleave alpha-synuclein fibrils helping to decrease the probability of aggregate formation [46] (McGlinchey and Lee 2015). It is also reported as a potential contributor to PD risk in presence of *GBA* mutations. Neurons derived from *GBA*-mutant induced pluripotent stem cells show a lower expression of this gene [47] (Blauwendraat et al. 2020b). However, in our analysis it was not specifically related to *GBA*+ PD patients. Interestingly, CTSB was overabundant in *LRRK2*+ PD patients and it was the only significant marker for the interaction term between disease status and mutation status. A link between LRRK2 and CTSB (and other lysosomal proteins) was previously identified via transcriptomics analysis of a G2019-*LRRK2* recombinant cell line where CTSB shown to be dysregulated [48] (Obergasteiger et al. 2020).

Our results on GPNMB and CTSB are in line with two independent recent reports [49,50] (Kia et al. 2021, Storm et al. 2021) proposing these two genes as causally related to PD based on mendelian randomization evidence from eQTL data in blood and brain tissue, both supported by colocalization and metagenome wide association.

Among the list of positive hits, we also identify *FCGR2A* and *FCGR2B*. These two Fc gamma receptor genes are part of a cluster of four (together with *FCGR3A* and *FCGR3B*) that map in direct neighborhood on the human genome, are expressed in a variety of immune response cells and are involved in several processes spanning from phagocytosis to inflammation [51] (Mellor et al. 2013). Both showed strong colocalization and mendelian randomization signals and they could be independently causal for PD.

FCGR2A variants have been recently proposed as putatively causal of PD [52] (Schilder and Raj 2021). In vitro experiments showed that alpha-synuclein aggregates specifically interact with FCGR2B at the surface of microglia cells [53] (Choi et al. 2015). Such interaction inhibits phagocytosis in microglia by activation of the phosphatase SHP-1 and may trigger

neurodegeneration [53] (Choi et al. 2015). The inhibition is reversed by knocking down FCGR2B in microglia.
Even though our results for these two genes are supported by previously published evidence, it is important to consider two potential limitations. First, these two genes are homologs and therefore some cross-reactivity between the individual SOMAmers cannot be completely ruled out. Second, being close neighbors in the human genome, linkage disequilibrium might explain why both genes are hits in a mendelian randomization.
Three out of the top four causal proteins; GPNMB, FCGR2B and CTSB are expressed in the microglia of the substantia nigra [49,53] (Choi et al. 2015, Kia et al. 2021) and the fourth protein, FCGR2A, probably too, since it contains consensus SNPs for PD fine-mapping, all of which exclusively overlap with microglia-specific epigenomic peaks [52] (Schilder and Raj 2021).

GBA mutation carriers

The lysosomal glucocerebrosidase *GBA* is one of the genes whose mutations are among the most known risk factors for development of PD and dementia with Lewy Body [54] (Bastien et al. 2021). In the subcohort of *GBA*+ PD, CALCA, DLK1, GCH1 and IL17A were among the proteins that show significantly different abundance as compared with unaffected *GBA*+ controls. While none of them have been previously linked specifically with *GBA*+ PD patients, all of them have been reported in association with PD.
CALCA (aka *CGRP*) encodes the hormone calcitonin and, via alternative splicing, for the calcitonin-related peptide alpha. CALCA is known to affect dopamine release and metabolism in the brain [55,56](Deutch & Roth 1987; Drumheller et al. 1992) and is significantly elevated in CSF of depressed PD patients as compared to healthy controls with major depressive disorder [57] (Svenningsson et al. 2017). The directionality was confirmed by our study, with increased levels of CALCA in *GBA*+ PD patients (p=0.022). DLK1 is a transmembrane protein initially described to be involved in the differentiation of peripheral cell types [58] (12101250). It has been shown to modulate meso-diencephalic dopaminergic neuronal differentiation and patterning as well as axon growth in animal models together with the well-known PD target *Nurr1* [59] (Jacobs et al. 2009). Significantly lower levels of DLK1 in *GBA*+ PD patients, as seen here (p=0.025), could be considered indicative of dopaminergic neurodegeneration. GCH1 is an enzyme involved in the synthesis of dopamine [60] (Kurian et al. 2011). Common and rare *GCH1* variants are associated with PD [61] (Xu et al. 2017) and with increased GCH1 expression in brain regions relevant for PD [62] (Rudakou et al. 2019); here we also found GCH1 elevated in *GBA*+ PD patients (p=0.025).
Bolte and Lukens [63] (2018) have recently proposed two separate mechanisms for T cell-mediated neurodegeneration in PD. One mechanism would involve IL17-producing CD4+ T cells (Th17 cells). Engagement of the IL17 receptor (IL17R) on neurons would cause altered NF-kB activation and subsequent neurodegeneration. The overabundance of IL17A in *GBA*+ PD patients (p=0.025) could suggest that neurodegeneration in these mutation carriers is mainly driven by this mechanism.

LRRK2 mutation carriers

Among the markers differentially expressed for *LRRK2*+ PD as compared to unaffected *LRRK2*+ controls; ARSA, CTSB, SMPD1 and TENM4 stand out for their association with PD. The role of ARSA in the context of PD has been recently reviewed by Paciotti et al [64]. (2020). In vitro and in vivo studies indicate that ARSA is a lysosomal chaperone interacting with alpha-synuclein in the cytoplasm via a non-lysosomal mediated function, preventing its aggregation, secretion and cell-to-cell propagation. The binding affinity of ARSA to alpha-synuclein is high

for an *ARSA* variant, which has been suggested to have a protective role in PD. Conversely, the affinity of a known pathogenic variant is low (Paciotti et al. 2020). There is an ongoing clinical trial to test the efficacy of ARSA as a therapeutic agent for PD (NCT01801709). We found ARSA elevated (p=0.023) among *LRRK2*+ PD patients. This result could be interpreted as a compensatory effect to prevent the formation aggregates.

CTSB and SMPD1 play important roles in the autophagy / lysosomal pathway in PD [65] (Senkevich and Gan-Or, 2020). *SMPD1* genetic variants are known to be associated with PD risk [66] (Mao et al. 2017) and the SMPD1 protein was overabundant in *LRRK2*+ PD patients here (p=0.037). Variants in the *CTSB* locus are known to modify PD risk in *GBA*+. These variants have been shown to decrease expression levels of *CTSB*, which may lead to lower lysosomal protease activity and increased accumulation of protein aggregates in neurons [63] (Blauwendraat et al. 2020b). We found CTSB significantly elevated in *LRRK2*+ PD patients (p=0.037), and no significant differences for the other groups. CTSB was the only marker that passed the significance threshold for an interaction term against non-mutation carriers (p=0.027). How *CTSB* might play a role in PD in association with *LRRK2* is not well understood, but in animal models it has been shown that on astrocyte lysosomes, presence of LRRK2 was correlated with absence of CTSB [67] (Bonet-Ponce et al. 2020). Also, considering that our causal analysis also pointed to CTSB, this is a protein that should be called attention to.

TENM4 encodes a transmembrane protein primarily expressed in the brain and is involved in axon guidance and central myelination [68] (Hor et al. 2015). Loss-of-function and damaging missense variants in *TENM4* are associated with early onset PD [69,70] (Pu et al. 2020; Liang et al. 2021). Accordingly, we saw significantly reduced levels of TENM4 in *LRRK2*+ PD patients (p=0.037). Several studies have linked missense mutations of TENM4 with essential tremor, the most common form of motor disorder worldwide. Although, essential tremor is genetically different from PD, the two indications share clinical manifestations and there are data reporting a greater risk of developing PD in essential tremor patients [71] (Algarni and Fasano 2018).

Idiopathic PD patients

Most of the PD cases worldwide are idiopathic PD with only ~10% being associated with a genetic background. Understanding the differential protein expression in idiopathic PD patients has an obvious potential for patient stratification and future development of more individually targeted therapies. Our findings for the subgroup of idiopathic PD patients, revealed also several markers associated with PD, when compared to healthy controls: AK1, DLK1, GPI, LPO, RAB31, RIP2K, SHANK1, TXN, VEGFA and VIP.

The AK1 kinase is a small enzyme involved in nucleotide metabolism [72] (Lanning et al. 2014) and it is mainly expressed in the brain, both in neurons and astrocytes [73] (Garcia-Esparcia et al. 2015). Transcriptomics profiling showed how *AK1* mRNA is down-regulated in the substantia nigra of PD patients probably as a consequence of dopaminergic neuronal death [73] (Garcia-Esparcia et al. 2015). The downregulation could also participate in the neurodegeneration by decreasing the AMPK function and thus impacting mitochondrial quality control [74] (Ionescu 2019). Interestingly, the same study showed *AK1* is up-regulated in the frontal cortex of end-stage PD patients probably as a mechanism to compensate for its reduced expression in other brain regions [73] (Garcia-Esparcia et al. 2015). We observed elevated AK1 levels in idiopathic PD patients (p=0.049), but since CSF reflects abundance across the central nervous system, the difference is difficult to interpret in this case.

The glucose metabolism enzyme GPI was also elevated in idiopathic PD patients (p=0.006). GPI is a modifier of neurodegeneration in animal models of PD, its overexpression in dopaminergic

neurons results in significant protection from alpha-synuclein-induced neurotoxicity. Also, in mouse neurons, knocking down GPI reverses said protection [75] (Knight et al. 2014). The increased GPI levels we saw here may be reflecting a compensatory mechanism. The increased glucose metabolism in worm and flies overexpressing GPI [75] (Knight et al. 2014) could be involved in the reduction of overall oxidative stress and impacting the mitochondrial quality control; both factors necessary for cellular homeostasis and aggregate removal via the endo-lysosomal pathway.

The metabolic enzyme TXN promotes cell proliferation and has anti-apoptotic functions, which makes it a good candidate for a neurodegeneration marker. Also, single cell-transcriptomics of human dopaminergic neurons carrying the A53T alpha-synuclein mutation found *TXN*, together with other glycolysis genes, upregulated upon mitochondrial damage inducing oxidative stress [76] (Fernandes et al. 2020). In a different PD in vivo model, downregulation of *Txn-1* amplified the MPTP-induced neurodegeneration [77] (Zeng et al. 2014) and its levels were found decreased in amnesic mild cognitive impairment [78] (Domenico et al. 2010), same as we found here for idiopathic PD (p=0.023).

We have already discussed how oxidative stress is considered as a contributing factor to the development of PD. Another protein involved in oxidative stress, LPO, is found in the substantia nigra and it has been proposed to be involved in neurodegeneration [79] (Fernández-Espejo et al. 2021). We found decreased levels of LPO in idiopathic PD patients (p=1.5e-6), opposite to a recent report where LPO levels in the CSF of idiopathic PD patients were increased as compared to controls [79](Fernández-Espejo et al. 2021).

Although there is no direct link known between HAMP and PD risk, it is interesting to see how both LPO and HAMP are tied to iron metabolism (HAMP being an iron regulatory hormone, and LPO being a heme peroxidase). Deposits of alpha-synuclein aggregates colocalize with iron distribution in the brain and specifically in the substantia nigra of PD patients [80] (Rizzollo et al. 2021). Iron dysregulation is associated with iron-induced oxidative stress and lipid peroxidation and is also implicated in neuroinflammation by mediating proinflammatory cytokine release in glial cells [82] (Ndayisaba et al. 2019), is correlated with motor disability in PD patients [83] (Mochizuki et al. 2020) most likely through neurodegeneration [83] (Viktorinova et al. 2021).

DLK1 has already been discussed for *GBA* mutation carriers; the same reduced levels in PD patients are also seen in the idiopathic subcohort (p=0.032).

Downregulation in idiopathic PD patients was also observed for the small GTPase RAB31 (p=0.044). This RAB protein participates in exosome biogenesis [84] (Wei et al. 2021) and it could be involved, together with the LRRK2 substrate RAB8, in alpha-synuclein spreading in PD patients [86] (Kumar et al. 2020).

Another LRRK2 substrate differentially expressed in idiopathic PD subcohort was RIPK2. Its phosphorylation by LRRK2 promotes inflammatory cytokine induction through the Nod1/2-Rip2 pathway [86] (Yan and Liu 2017). LRRK2 deficiency leads to less activation of RIPK2. Although we saw reduced levels of RIPK2 in PD patients for all the subcohorts including *LRRK2*+ PD, this difference was only significant for the idiopathic PD subcohort (p=4.6e-5).

The postsynaptic density protein SHANK1 was also less abundant in idiopathic in PD (p=0.023). SHANK1 is a well-known marker in psychiatric indications and more recently in Alzheimer's disease [87] (Guilmatre et al. 2014). It has been suggested to be regulated by PINK1, for which rare variants are known to be causal for PD [88] (Hernández et al. 2019). Even if the cytoplasmic role of PINK1 is still controversial, the authors suggested that the mitochondrial kinase may modulate dendritic spine morphology in hippocampal neurons. In *PINK* knockdown neurons the

expression of *PSD95* and *SHANK1* is decreased compared to controls [88] (Hernández et al. 2019). Because synaptic strength is dependent on the spine architecture [90] (Murthy 1998) a decrease in the levels of these two proteins, enriched at the postsynapse [90,91] (Jiang et al. 2013, Yoo et al. 2019), may have a negative consequence in synaptic plasticity as also previously reported in in vitro models [92,93] (Mao et al. 2015; Coley et al. 2019). Conversely, reduced *PINK1* expression leads to an increased level of actin binding protein found in dendritic spines with a consequent variation in spine morphology and number [88] (Hernández et al. (2019). The mechanism of action for this pathway is not fully understood yet, thus it is not clear if the postsynaptic density modulation is a direct result of SHANK1 and PSD95 phosphorylation or an indirect process. A possible relation between SHANK1 levels and neurodegeneration was previously reported for Alzheimer's disease where protein levels were decreased in both mouse model and human brain samples [87,94] (Guilmatre et al. 2014, Pham et al. 2010). It remains to be elucidated if the synaptic plasticity modulation and consequent neuronal degeneration are due to a direct role of PINK on the postsynaptic structure or the result of a secondary mitochondrial damage that impacts the overall neuronal architecture.

VEGFA is neuroprotective / neurotrophic factor in the brain, and it has been proposed as a disease modifying protein for gene therapy in animal models of PD [96] (Axelsen and Woldbye, 2018). Also, an association has been established between a *VEGFA* variant and PD risk [96] (Wu et al. 2016). The decreased levels of VEGFA we observe in idiopathic PD patients (p=0.006) may reflect ongoing neurodegeneration. VIP enhances striatal plasticity and prevents dopaminergic cell loss in Parkinsonian rats [97] (Korkmaz et al. 2012) and has been directly proposed as a therapeutic target against inflammation-induced neurodegeneration in PD [99] (Gonzalez-Rey et al. 2005). As for VEGFA, the decreased levels of VIP in idiopathic PD patients we observe (p=0.023) may reflect the neurodegeneration process.

Heterogeneity of idiopathic PD

The two protein modules identified by network analysis split the idiopathic subcohort in two robust classes or endotypes of idiopathic PD patients (Figure 3A). The robustness of these endotypes was reflected by their stability as assessed using consensus clustering (Figure 3B), and by the high performance (0.73 accuracy and 0.77 AUC in the test set) of the endotype predictive model based on a partition tree (Figure 3C). This partition tree approach suggest that clinical variables alone may provide a robust method for stratification, i.e., without SomaScan analysis of CSF. Importantly, CSF phospho-tau levels sufficed to discriminate endotype 1 from 2. Differences in clinical variables between the two endotypes were limited to CSF levels of amyloid beta, phospho-tau and alpha-synuclein, all of them lower for endotype 2 (Figure 4). Despite the two endotypes being similar in relative size, endotype 1 differs from healthy controls by merely four proteins in the SomaScan analysis, while endotype 2 differs by 276 proteins, independently of age, sex or study center.

In contrast to these endotypes, we also identified what we called "phenotypes" based on clinical variables. Endotypes and phenotypes were identified from different input variables (SomaScan proteomics in the first case, clinical variables relevant to PD diagnosis and progression in the second case) in an unsupervised manner and yet the overlap between endotypes and phenotypes was remarkable (accuracy 0.761). To our understanding this is indicative of the existence of true underlying subtypes of idiopathic PD patients, which may not be perfectly dissected by neither the endotype nor the phenotype approach but are robustly reflected by similar results in both cases.

Summary

Molecular characterization of human patients is a challenging goal when dealing with a heterogeneous indication like PD. A reasonably powered analysis of high-throughput proteomics of the CSF required a sufficiently large number of samples, particularly if we take into consideration the presence of PD subcohorts (two genetic, one idiopathic) and the set of biological and technical variables that could affect the results. Here, we performed the first proteogenomic characterization of CSF of PD patients and controls from the PPMI cohort revealing a combination of known risk or progression markers of PD and new candidates that were not previously cited in the context of PD. Differences between the genetic (*GBA*+ and *LRRK2*+) and the idiopathic PD subcohorts when compared to controls, as well as subclass differences within the idiopathic PD subcohort, suggest that neurodegeneration and neuroinflammation patterns might be different in each of them. Furthermore, an integrated analysis of the genomics and CSF proteomics of the combined cohorts reveals known and new causal proteins. Although future studies will be needed to address biological questions and bring additional validation, our findings could be pivotal to identify new therapeutic targets and shape personalized medicine in the neurodegenerative field.

Declarations of interest

The study was supported by the Novartis Institutes for Biomedical Research and Merck. Protein measurements were performed at SomaLogic. A.K., B.B., B.P. F.E., J.J., J.L., L.Z., M. Dovlatyan, M.M., M.R., P.S-F., S.K. and V.S. are or have been employees of Novartis; J.J., S.K., M.M. are also stockholders of Novartis. D.S., J.D-A., J.M., N.R. and S.L. are or have been employees of Merck. All other authors have no conflict of interests to declare.

Acknowledgements

The authors would like to thank Rose Case (Indiana University Genetics Biobank, Indiana University School of Medicine, Indianapolis, USA) for her key role in sample management and Myung Shin (Genome and Biomarker Sciences. Merck Exploratory Science Center. Cambridge, USA) and Faraz Faghri (Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health. Bethesda, USA) for constructive feedback.

References

1. Group GBDNDC. Global, regional, and national burden of neurological disorders during 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet Neurol* 2017; **16**(11): 877-97.
2. Jagadeesan AJ, Murugesan R, Vimala Devi S, et al. Current trends in etiology, prognosis and therapeutic aspects of Parkinson's disease: a review. *Acta Biomed* 2017; **88**(3): 249-62.
3. Blauwendraat C, Nalls MA, Singleton AB. The genetic architecture of Parkinson's disease. *Lancet Neurol* 2020; **19**(2): 170-8.
4. Bonifati V. Genetics of Parkinson's disease--state of the art, 2013. *Parkinsonism Relat Disord* 2014; **20 Suppl 1**: S23-8.
5. Kestenbaum M, Alcalay RN. Clinical Features of LRRK2 Carriers with Parkinson's Disease. *Adv Neurobiol* 2017; **14**: 31-48.
6. Thaler A, Bregman N, Gurevich T, et al. Parkinson's disease phenotype is influenced by the severity of the mutations in the GBA gene. *Parkinsonism Relat Disord* 2018; **55**: 45-9.
7. Qian E, Huang Y. Subtyping of Parkinson's Disease - Where Are We Up To? *Aging Dis* 2019; **10**(5): 1130-9.
8. Ma LY, Chan P, Gu ZQ, Li FF, Feng T. Heterogeneity among patients with Parkinson's disease: cluster analysis and genetic association. *J Neurol Sci* 2015; **351**(1-2): 41-5.
9. Parkinson Progression Marker I. The Parkinson Progression Marker Initiative (PPMI). *Prog Neurobiol* 2011; **95**(4): 629-35.
10. Gold L, Ayers D, Bertino J, et al. Aptamer-based multiplexed proteomic technology for biomarker discovery. *PLoS One* 2010; **5**(12): e15004.
11. Marek K, Chowdhury S, Siderowf A, et al. The Parkinson's progression markers initiative (PPMI) - establishing a PD biomarker cohort. *Ann Clin Transl Neurol* 2018; **5**(12): 1460-77.
12. Simuni T, Uribe L, Cho HR, et al. Clinical and dopamine transporter imaging characteristics of non-manifest LRRK2 and GBA mutation carriers in the Parkinson's Progression Markers Initiative (PPMI): a cross-sectional study. *Lancet Neurol* 2020; **19**(1): 71-80.
13. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* 2012; **28**(6): 882-3.
14. Shabalin AA. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics* 2012; **28**(10): 1353-8.
15. Hemani G, Zheng J, Elsworth B, et al. The MR-Base platform supports systematic causal inference across the human phenome. *Elife* 2018; **7**.
16. Hemani G, Tilling K, Davey Smith G. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. *PLoS Genet* 2017; **13**(11): e1007081.
17. Nalls MA, Blauwendraat C, Vallerga CL, et al. Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet Neurol* 2019; **18**(12): 1091-102.
18. Verbanck M, Chen CY, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat Genet* 2018; **50**(5): 693-8.
19. Giambartolomei C, Vukcevic D, Schadt EE, et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet* 2014; **10**(5): e1004383.
20. Goetz CG, Tilley BC, Shaftman SR, et al. Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): scale presentation and clinimetric testing results. *Mov Disord* 2008; **23**(15): 2129-70.
21. Nasreddine ZS, Phillips NA, Bedirian V, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *J Am Geriatr Soc* 2005; **53**(4): 695-9.
22. Kang JH, Irwin DJ, Chen-Plotkin AS, et al. Association of cerebrospinal fluid beta-amyloid 1-42, T-tau, P-tau181, and alpha-synuclein levels with clinical features of drug-naive patients with early Parkinson disease. *JAMA Neurol* 2013; **70**(10): 1277-87.
23. Mollenhauer B, Caspell-Garcia CJ, Coffey CS, et al. Longitudinal analyses of cerebrospinal fluid alpha-Synuclein in prodromal and early Parkinson's disease. *Mov Disord* 2019; **34**(9): 1354-64.

24. Therneau T, Atkinson B, Ripley B. rpart: Recursive Partitioning and Regression Trees. R package version 4.1-15. 2019.
25. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 2008; **9**: 559.
26. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics* 2010; **26**(12): 1572-3.
27. Ploner A. Heatplus: Heatmaps with row and/or column covariates and colored clusters. R package. https://github.com/alexploner/Heatplus: Bioconductor; 2015.
28. Konopka T. umap: Uniform Manifold Approximation and Projection. https://cran.r-project.org/web/packages/umap/index.html; 2020.
29. Kassambara A, Mundt F. factoextra: Extract and Visualize the Results of Multivariate Data Analyses. https://cran.r-project.org/web/packages/factoextra/index.html; 2020.
30. Saade M, Araujo de Souza G, Scavone C, Kinoshita PF. The Role of GPNMB in Inflammation. *Front Immunol* 2021; **12**: 674739.
31. Tomihari M, Hwang SH, Chung JS, Cruz PD, Jr., Ariizumi K. Gpnmb is a melanosome-associated glycoprotein that contributes to melanocyte/keratinocyte adhesion in a RGD-dependent fashion. *Exp Dermatol* 2009; **18**(7): 586-95.
32. Rose AA, Annis MG, Dong Z, et al. ADAM10 releases a soluble form of the GPNMB/Osteoactivin extracellular domain with angiogenic properties. *PLoS One* 2010; **5**(8): e12093.
33. Abdelmagid SM, Barbe MF, Rico MC, et al. Osteoactivin, an anabolic factor that regulates osteoblast differentiation and function. *Exp Cell Res* 2008; **314**(13): 2334-51.
34. Furochi H, Tamura S, Mameoka M, et al. Osteoactivin fragments produced by ectodomain shedding induce MMP-3 expression via ERK pathway in mouse NIH-3T3 fibroblasts. *FEBS Lett* 2007; **581**(30): 5743-50.
35. Bandari PS, Qian J, Yehia G, et al. Hematopoietic growth factor inducible neurokinin-1 type: a transmembrane protein that is similar to neurokinin 1 interacts with substance P. *Regul Pept* 2003; **111**(1-3): 169-78.
36. Moloney EB, Moskites A, Ferrari EJ, Isacson O, Hallett PJ. The glycoprotein GPNMB is selectively elevated in the substantia nigra of Parkinson's disease patients and increases after lysosomal stress. *Neurobiol Dis* 2018; **120**: 1-11.
37. Huttenrauch M, Ogorek I, Klafki H, et al. Glycoprotein NMB: a novel Alzheimer's disease associated marker expressed in a subset of activated microglia. *Acta Neuropathol Commun* 2018; **6**(1): 108.
38. Tanaka H, Shimazawa M, Kimura M, et al. The potential of GPNMB as novel neuroprotective factor in amyotrophic lateral sclerosis. *Sci Rep* 2012; **2**: 573.
39. Kramer G, Wegdam W, Donker-Koopman W, et al. Elevation of glycoprotein nonmetastatic melanoma protein B in type 1 Gaucher disease patients and mouse models. *FEBS Open Bio* 2016; **6**(9): 902-13.
40. Rocha EM, Smith GA, Park E, et al. Progressive decline of glucocerebrosidase in aging and Parkinson's disease. *Ann Clin Transl Neurol* 2015; **2**(4): 433-8.
41. Ahn JH, Lee Y, Jeon C, et al. Identification of the genes differentially expressed in human dendritic cell subsets by cDNA subtraction and microarray analysis. *Blood* 2002; **100**(5): 1742-54.
42. Ono Y, Tsuruma K, Takata M, Shimazawa M, Hara H. Glycoprotein nonmetastatic melanoma protein B extracellular fragment shows neuroprotective effects and activates the PI3K/Akt and MEK/ERK pathways via the Na+/K+-ATPase. *Sci Rep* 2016; **6**: 23241.
43. Kinoshita PF, Yshii LM, Orellana AMM, et al. Alpha 2 Na(+),K(+)-ATPase silencing induces loss of inflammatory response and ouabain protection in glial cells. *Sci Rep* 2017; **7**(1): 4894.
44. Shi F, Duan S, Cui J, et al. Induction of matrix metalloproteinase-3 (MMP-3) expression in the microglia by lipopolysaccharide (LPS) via upregulation of glycoprotein nonmetastatic melanoma B (GPNMB) expression. *J Mol Neurosci* 2014; **54**(2): 234-42.
45. Brendza R, Lin H, Stark K, et al. Genetic ablation of Gpnmb does not alter synuclein-related pathology. *Neurobiol Dis* 2021; **159**: 105494.
46. McGlinchey RP, Lee JC. Cysteine cathepsins are essential in lysosomal degradation of alpha-synuclein. *Proc Natl Acad Sci U S A* 2015; **112**(30): 9322-7.
47. Blauwendraat C, Reed X, Krohn L, et al. Genetic modifiers of risk and age at onset in GBA associated Parkinson's disease and Lewy body dementia. *Brain* 2020; **143**(1): 234-48.

48. Obergasteiger J, Frapporti G, Lamonaca G, et al. Kinase inhibition of G2019S-LRRK2 enhances autolysosome formation and function to reduce endogenous alpha-synuclein intracellular inclusions. *Cell Death Discov* 2020; **6**: 45.
49. Kia DA, Zhang D, Guelfi S, et al. Identification of Candidate Parkinson Disease Genes by Integrating Genome-Wide Association Study, Expression, and Epigenetic Data Sets. *JAMA Neurol* 2021; **78**(4): 464-72.
50. Storm CS, Kia DA, Almramhi MM, et al. Finding genetically-supported drug targets for Parkinson's disease using Mendelian randomization of the druggable genome. *Nat Commun* 2021; **12**(1): 7342.
51. Mellor JD, Brown MP, Irving HR, Zalcberg JR, Dobrovic A. A critical review of the role of Fc gamma receptor polymorphisms in the response to monoclonal antibodies in cancer. *J Hematol Oncol* 2013; **6**: 1.
52. Schilder BM, Raj T. Fine-mapping of Parkinson's disease susceptibility loci identifies putative causal variants. *Hum Mol Genet* 2021.
53. Choi YR, Kang SJ, Kim JM, et al. FcgammaRIIB mediates the inhibitory effect of aggregated alpha-synuclein on microglial phagocytosis. *Neurobiol Dis* 2015; **83**: 90-9.
54. Bastien J, Menon S, Messa M, Nyfeler B. Molecular targets and approaches to restore autophagy and lysosomal capacity in neurodegenerative disorders. *Mol Aspects Med* 2021; **82**: 101018.
55. Deutch AY, Roth RH. Calcitonin gene-related peptide in the ventral tegmental area: selective modulation of prefrontal cortical dopamine metabolism. *Neurosci Lett* 1987; **74**(2): 169-74.
56. Drumheller A, Menard D, Fournier A, Jolicoeur FB. Neurochemical effects of CGRP. *Ann N Y Acad Sci* 1992; **657**: 546-8.
57. Svenningsson P, Palhagen S, Mathe AA. Neuropeptide Y and Calcitonin Gene-Related Peptide in Cerebrospinal Fluid in Parkinson's Disease with Comorbid Depression versus Patients with Major Depressive Disorder. *Front Psychiatry* 2017; **8**: 102.
58. Moon YS, Smas CM, Lee K, et al. Mice lacking paternally expressed Pref-1/Dlk1 display growth retardation and accelerated adiposity. *Mol Cell Biol* 2002; **22**(15): 5585-92.
59. Jacobs FM, van der Linden AJ, Wang Y, et al. Identification of Dlk1, Ptpru and Klhl1 as novel Nurr1 target genes in meso-diencephalic dopamine neurons. *Development* 2009; **136**(14): 2363-73.
60. Kurian MA, Gissen P, Smith M, Heales S, Jr., Clayton PT. The monoamine neurotransmitter disorders: an expanding range of neurological syndromes. *Lancet Neurol* 2011; **10**(8): 721-33.
61. Xu Q, Li K, Sun Q, et al. Rare GCH1 heterozygous variants contributing to Parkinson's disease. *Brain* 2017; **140**(7): e41.
62. Rudakou U, Ouled Amar Bencheikh B, Ruskey JA, et al. Common and rare GCH1 variants are associated with Parkinson's disease. *Neurobiol Aging* 2019; **73**: 231 e1- e6.
63. Bolte AC, Lukens JR. Th17 Cells in Parkinson's Disease: The Bane of the Midbrain. *Cell Stem Cell* 2018; **23**(1): 5-6.
64. Paciotti S, Albi E, Parnetti L, Beccari T. Lysosomal Ceramide Metabolism Disorders: Implications in Parkinson's Disease. *J Clin Med* 2020; **9**(2).
65. Senkevich K, Gan-Or Z. Autophagy lysosomal pathway dysfunction in Parkinson's disease; evidence from human genetics. *Parkinsonism Relat Disord* 2020; **73**: 60-71.
66. Mao CY, Yang J, Wang H, et al. SMPD1 variants in Chinese Han patients with sporadic Parkinson's disease. *Parkinsonism Relat Disord* 2017; **34**: 59-61.
67. Bonet-Ponce L, Beilina A, Williamson CD, et al. LRRK2 mediates tubulation and vesicle sorting from lysosomes. *Sci Adv* 2020; **6**(46).
68. Hor H, Francescatto L, Bartesaghi L, et al. Missense mutations in TENM4, a regulator of axon guidance and central myelination, cause essential tremor. *Hum Mol Genet* 2015; **24**(20): 5677-86.
69. Pu JL, Gao T, Si XL, et al. Parkinson's Disease in Teneurin Transmembrane Protein 4 (TENM4) Mutation Carriers. *Front Genet* 2020; **11**: 598064.
70. Liang D, Zhao Y, Pan H, et al. Rare variant analysis of essential tremor-associated genes in early-onset Parkinson's disease. *Ann Clin Transl Neurol* 2021; **8**(1): 119-25.
71. Algarni M, Fasano A. The overlap between Essential tremor and Parkinson disease. *Parkinsonism Relat Disord* 2018; **46 Suppl 1**: S101-S4.

72. Lanning NJ, Looyenga BD, Kauffman AL, et al. A mitochondrial RNAi screen defines cellular bioenergetic determinants and identifies an adenylate kinase as a key regulator of ATP levels. *Cell Rep* 2014; **7**(3): 907-17.
73. Garcia-Esparcia P, Hernandez-Ortega K, Ansoleaga B, Carmona M, Ferrer I. Purine metabolism gene deregulation in Parkinson's disease. *Neuropathol Appl Neurobiol* 2015; **41**(7): 926-40.
74. Ionescu MI. Adenylate Kinase: A Ubiquitous Enzyme Correlated with Medical Conditions. *Protein J* 2019; **38**(2): 120-33.
75. Knight AL, Yan X, Hamamichi S, et al. The glycolytic enzyme, GPI, is a functionally conserved modifier of dopaminergic neurodegeneration in Parkinson's models. *Cell Metab* 2014; **20**(1): 145-57.
76. Fernandes HJR, Patikas N, Foskolou S, et al. Single-Cell Transcriptomics of Parkinson's Disease Human In Vitro Models Reveals Dopamine Neuron-Specific Stress Responses. *Cell Rep* 2020; **33**(2): 108263.
77. Zeng XS, Jia JJ, Kwon Y, Wang SD, Bai J. The role of thioredoxin-1 in suppression of endoplasmic reticulum stress in Parkinson disease. *Free Radic Biol Med* 2014; **67**: 10-8.
78. Di Domenico F, Sultana R, Tiu GF, et al. Protein levels of heat shock proteins 27, 32, 60, 70, 90 and thioredoxin-1 in amnestic mild cognitive impairment: an investigation on the role of cellular stress response in the progression of Alzheimer disease. *Brain Res* 2010; **1333**: 72-81.
79. Fernandez-Espejo E, Rodriguez de Fonseca F, Suarez J, Martin de Pablos A. Cerebrospinal fluid lactoperoxidase level is enhanced in idiopathic Parkinson's disease, and correlates with levodopa equivalent daily dose. *Brain Res* 2021; **1761**: 147411.
80. Rizzollo F, More S, Vangheluwe P, Agostinis P. The lysosome as a master regulator of iron metabolism. *Trends Biochem Sci* 2021; **46**(12): 960-75.
81. Ndayisaba A, Kaindlstorfer C, Wenning GK. Iron in Neurodegeneration - Cause or Consequence? *Front Neurosci* 2019; **13**: 180.
82. Mochizuki H, Choong CJ, Baba K. Parkinson's disease and iron. *J Neural Transm (Vienna)* 2020; **127**(2): 181-7.
83. Viktorinova A, Durfinova M. Mini-Review: Is iron-mediated cell death (ferroptosis) an identical factor contributing to the pathogenesis of some neurodegenerative diseases? *Neurosci Lett* 2021; **745**: 135627.
84. Wei D, Zhan W, Gao Y, et al. RAB31 marks and controls an ESCRT-independent exosome pathway. *Cell Res* 2021; **31**(2): 157-77.
85. Kumar R, Donakonda S, Muller SA, Botzel K, Hoglinger GU, Koeglsperger T. FGF2 Affects Parkinson's Disease-Associated Molecular Networks Through Exosomal Rab8b/Rab31. *Front Genet* 2020; **11**: 572058.
86. Yan R, Liu Z. LRRK2 enhances Nod1/2-mediated inflammatory cytokine production by promoting Rip2 phosphorylation. *Protein Cell* 2017; **8**(1): 55-66.
87. Guilmatre A, Huguet G, Delorme R, Bourgeron T. The emerging role of SHANK genes in neuropsychiatric disorders. *Dev Neurobiol* 2014; **74**(2): 113-22.
88. Hernandez CJ, Baez-Becerra C, Contreras-Zarate MJ, Arboleda H, Arboleda G. PINK1 Silencing Modifies Dendritic Spine Dynamics of Mouse Hippocampal Neurons. *J Mol Neurosci* 2019; **69**(4): 570-9.
89. Murthy VN. Synaptic plasticity: step-wise strengthening. *Curr Biol* 1998; **8**(18): R650-3.
90. Jiang YH, Ehlers MD. Modeling autism by SHANK gene mutations in mice. *Neuron* 2013; **78**(1): 8-27.
91. Yoo KS, Lee K, Oh JY, et al. Postsynaptic density protein 95 (PSD-95) is transported by KIF5 to dendritic regions. *Mol Brain* 2019; **12**(1): 97.
92. Mao W, Watanabe T, Cho S, et al. Shank1 regulates excitatory synaptic transmission in mouse hippocampal parvalbumin-expressing inhibitory interneurons. *Eur J Neurosci* 2015; **41**(8): 1025-35.
93. Coley AA, Gao WJ. PSD-95 deficiency disrupts PFC-associated function and behavior during neurodevelopment. *Sci Rep* 2019; **9**(1): 9486.
94. Pham E, Crews L, Ubhi K, et al. Progressive accumulation of amyloid-beta oligomers in Alzheimer's disease and in amyloid precursor protein transgenic mice is accompanied by selective alterations in synaptic scaffold proteins. *FEBS J* 2010; **277**(14): 3051-67.

95. Axelsen TM, Woldbye DPD. Gene Therapy for Parkinson's Disease, An Update. *J Parkinsons Dis* 2018; **8**(2): 195-215.
96. Wu Y, Zhang Y, Han X, Li X, Xue L, Xie A. Association of VEGF gene polymorphisms with sporadic Parkinson's disease in Chinese Han population. *Neurol Sci* 2016; **37**(12): 1923-9.
97. Korkmaz O, Ay H, Ulupinar E, Tuncel N. Vasoactive intestinal peptide enhances striatal plasticity and prevents dopaminergic cell loss in Parkinsonian rats. *J Mol Neurosci* 2012; **48**(3): 565-73.
98. Gonzalez-Rey E, Chorny A, Fernandez-Martin A, Varela N, Delgado M. Vasoactive intestinal peptide family as a therapeutic target for Parkinson's disease. *Expert Opin Ther Targets* 2005; **9**(5): 923-9.

Figures

Figure 1

Analytical Design

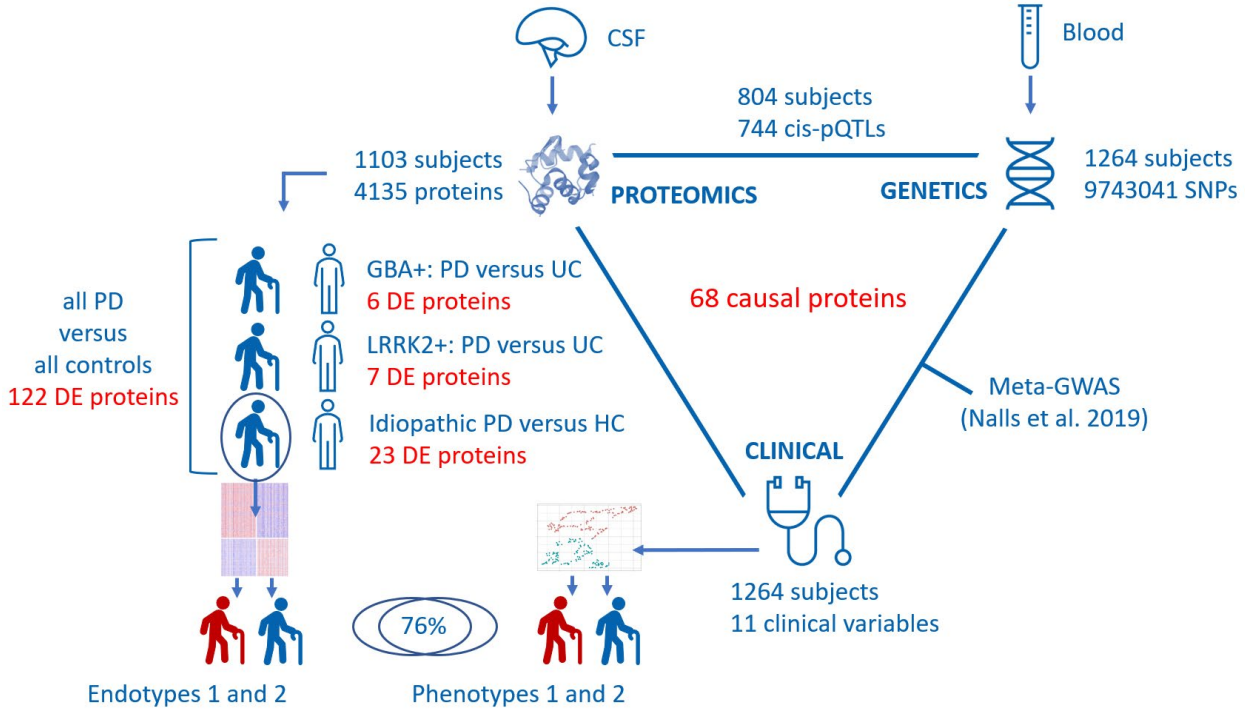

Figure 1. Full analytical design: 1103 subjects were analyzed with SomaScan for 4135 unique proteins in CSF. The comparison between *GBA*+ PD patients and *GBA*+ unaffected controls (UC) retrieved six differentially expressed (DE) proteins. The comparison between *LRRK2*+ PD patients and *LRRK2*+ UC retrieved seven DE proteins. The comparison between idiopathic PD patients and HC non-mutation carriers retrieved 23 DE proteins. Patients and controls were also combined and compared, which retrieved 122 DE proteins. Idiopathic PD patients were further analyzed, and two endotypes were identified based on CSF proteomics. 1264 subjects were sequenced genome wide to detect a total of 9743041 SNPs. For the 804 patients that had both genomic and proteomic data, a pQTL analysis was performed that identified 744 unique proteins with a significant cis-pQTL. The pQTLs combined with a meta GWAS for PD performed by Nalls et al. (2019), led to the proposal of 68 unique CSF proteins presumed to be causal for PD. Phenotyping for idiopathic PD patients took place based on eleven clinical variables that lead to two distinct phenotypes. Endotypes and phenotypes overlapped with an accuracy of 76%.

Figure 2

Causal Analysis (best protein candidate)

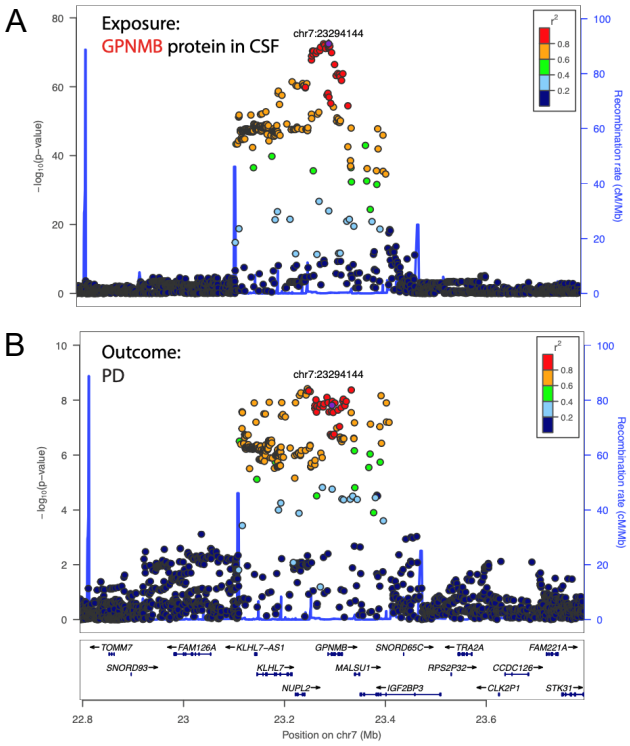

Figure 2. A: Locus visualization of GPNMB pQTL hits suggest a strong association between GPNMB SOMAmer levels and its cis-SNPs. Colors indicate the linkage disequilibrium (LD) correlation of other SNPs with chr7:23294144 (rs858275).**B**: Locus visualization of GWAS hits in the GPNMB locus for the risk of developing Parkinson's disease. The y axis is the -log10 nominal p-value of the GWAS results.

Figure 3

Idiopathic PD endotypes

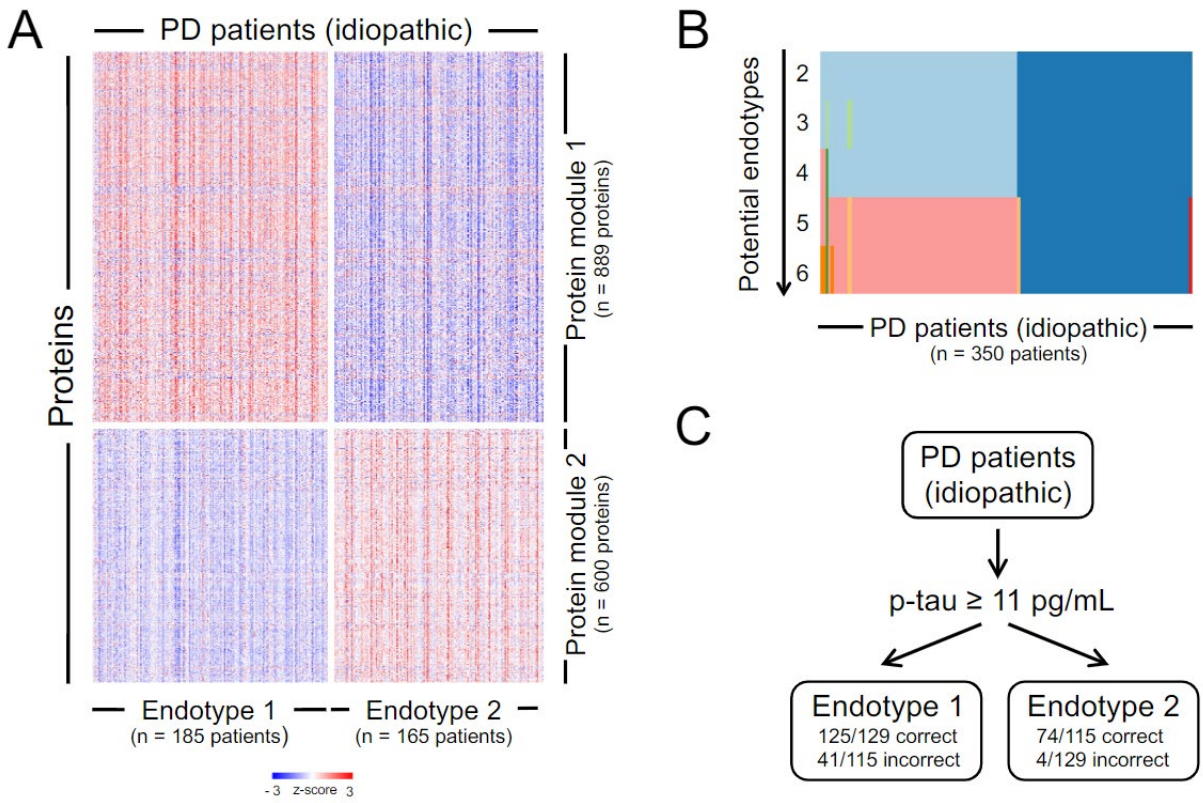

Figure 3. A: Heatmap of z-scores of the protein values as measured with SomaScan, corresponding to the two modules identified using Weighted Gene Co-expression Network Analysis (WGCNA). The proteins in these modules are used for cluster analysis using Consensus Clustering, which retrieves two clusters (endotypes) of idiopathic PD patients. Patients are shown in the x-axis, separated by endotype, while proteins are shown in the y-axis, separated by module. **B**: Tracking plot depicting how the idiopathic PD patients are assigned to specific endotypes by Consensus Clustering as the number of potential endotypes increases. **C**: Partition Tree predicting endotype membership of the idiopathic PD patients based on clinical variables only. One node suffices to separate patients into endotypes based on phospho-tau levels (p-tau) in CSF as measured with a clinical assay, the cutoff being 11 pg/mL.

Figure 4

Idiopathic PD endotypes and phenotypes.

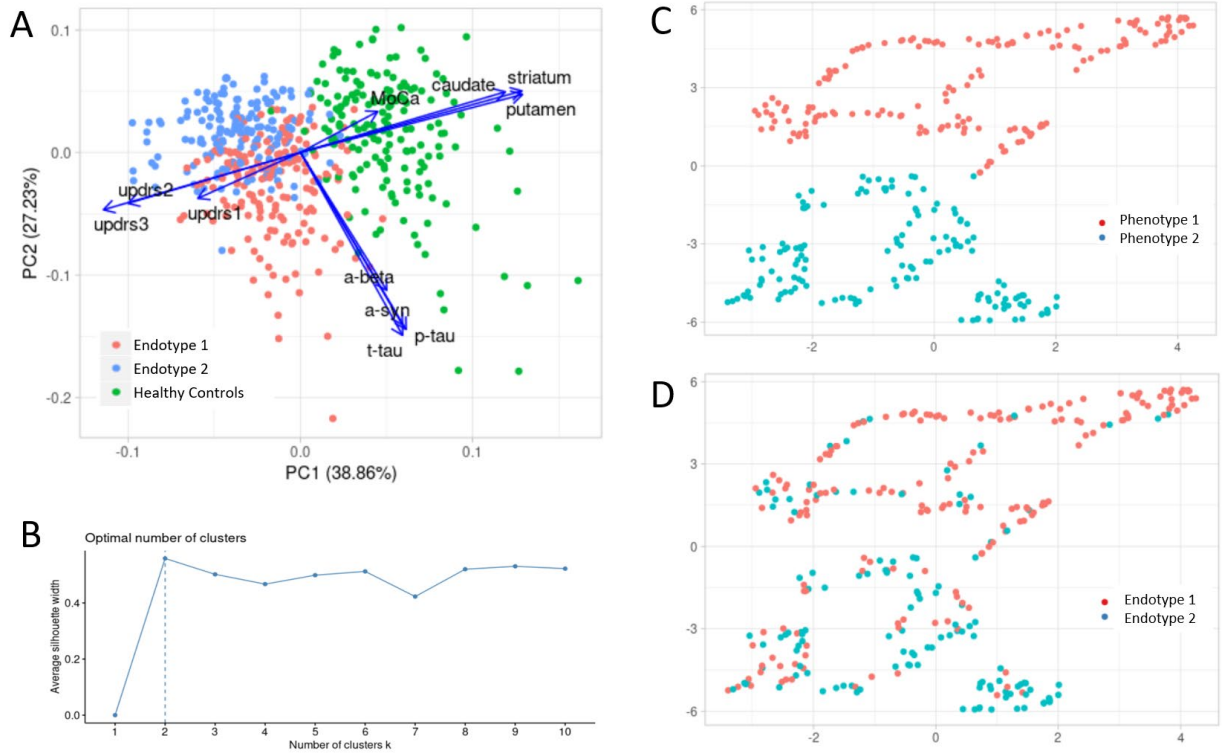

Figure 4. A: Principal Component Analysis of the clinical variables corresponding to mean caudate, striatum and putamen thickness, UPDRS scores 1, 2 and 3, MoCA and CSF protein levels of alpha-synuclein (a-syn), amyloid beta (a-beta), phospho-tau (p-tau) and total tau (t-tau), as measured with clinical assays. The axes show principal components 1 and 2 with the percentage of variance explained in parentheses. The arrows represent the loadings of the clinical variables on the principal component projection. Colors distinguish between controls and endotypes 1 and 2. **B**: Average silhouette width for different potential numbers of clusters as retrieved by k-means over the Uniform Manifold Approximation and Projection (UMAP) shown in 4C. The optimal number of clusters (two) is highlighted with a dotted line. **C**: UMAP of the idiopathic PD patients using the same clinical variables described in A, colored by cluster (phenotype) membership. **D**: Same UMAP as in C, colored by SomaScan endotypes for comparison (76% accuracy in the overlap).

Figure 5

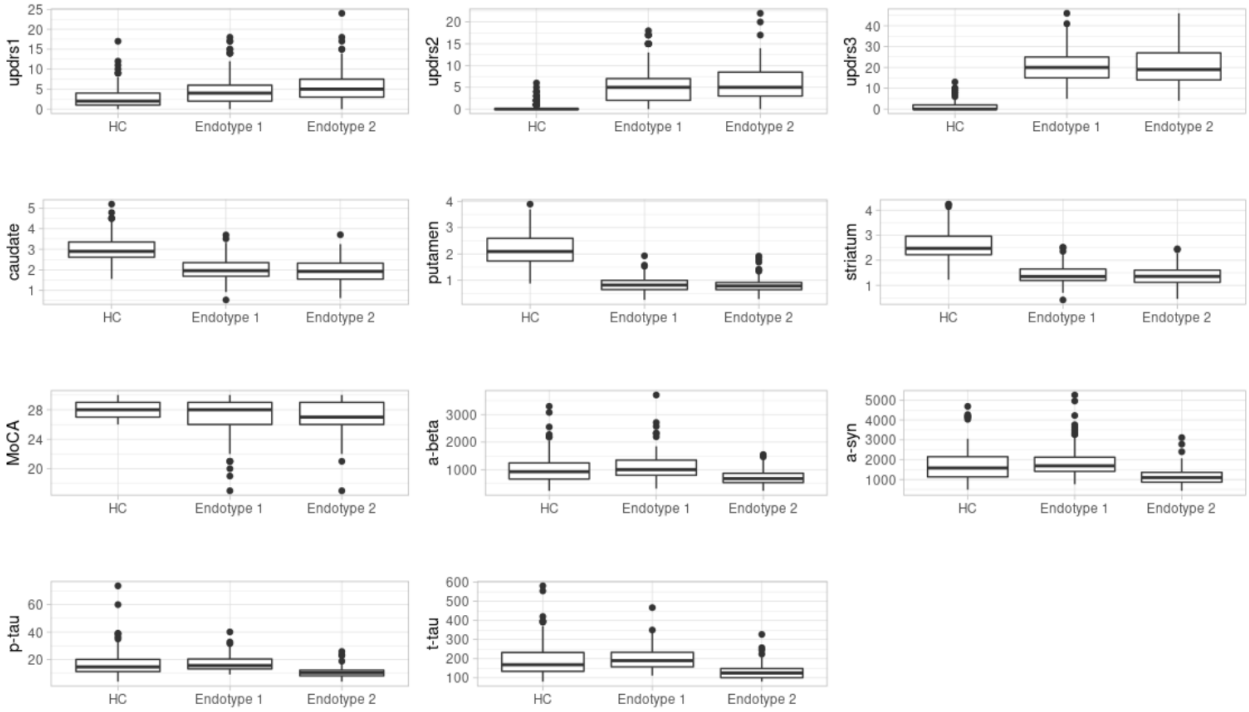

Figure 5. Comparison of clinical variables between endotypes of idiopathic PD. Healthy controls (HC) are shown for reference. The clinical variables are UPDRS part 1, 2 and 3 (scores), caudate, striatum and putamen thickness (mm), MoCA (score), and levels of amyloid beta (a-beta; pg/mL), alpha-synuclein (a-syn; pg/mL), phospho-tau 181P (p-tau; pg/mL) and total tau (t-tau; pg/mL). Differences between endotype 1 and 2 are only significant for the four biomarkers: amyloid beta, alpha-synuclein, phospho-tau and total tau.

Tables

Table 1

PD subcohorts

Subcohort	Parkinson's Disease	Controls
GBA+	65	162*
LRRK2+	154	196*
Idiopathic	350	176#

* "unaffected controls" and # "healthy controls" in the text

Table 2

PD causal proteins from CSF

Protein	Number of IVs	Method	β	FDR	Colocalization PP.H4.abf	Horizontal Pleiotropy Test (MRPRESSO p-value)
GPNMB[1]	11	IVW	0.14	1.17×10^{-17}	0.94	0.65
FCGR2B	17	IVW	0.07	5.11×10^{-14}	0.96	0.49
FCGR2A	18	IVW	0.07	2.70×10^{-12}	0.95	0.26
CTSB	12	IVW	−0.11	2.63×10^{-10}	0.22	0.47
HLA-DQA2[2]	33	IVW	−0.14	1.43×10^{-9}	0.03	0.00
CD38	1	Wald ratio	−0.53	2.45×10^{-9}	0.45	NA
HP	19	IVW	0.06	2.01×10^{-6}	0.01	0.60
LTF	23	IVW	0.05	2.16×10^{-6}	0.04	0.65
HAVCR2	13	IVW	−0.10	2.87×10^{-5}	0.30	0.67
BST1[2]	10	IVW	0.12	4.00×10^{-5}	0.24	0.00
CLEC3B	4	IVW	−0.16	1.02×10^{-4}	0.29	0.97
HAPLN1	15	IVW	0.06	3.21×10^{-4}	0.05	0.79
MANEA	17	IVW	0.05	3.21×10^{-4}	0.10	0.94
NQO2	11	IVW	0.05	3.21×10^{-4}	0.05	0.98
ARSA[1]	5	IVW	0.13	6.82×10^{-4}	0.58	0.62
LGALS3	8	IVW	0.07	7.11×10^{-4}	0.02	0.50
IL1RL1	26	IVW	0.03	9.63×10^{-4}	0.01	0.90
PAM	12	IVW	0.09	9.63×10^{-4}	0.08	0.47
TPSAB1	8	IVW	−0.06	0.0011	0.15	0.80
HSP90B1	22	IVW	0.04	0.0011	0.03	0.19

GLCE	14	IVW	0.05	0.0015	0.10	0.96
MANSC4	18	IVW	0.05	0.0016	0.02	0.85
RABEPK	10	IVW	−0.05	0.0021	0.10	0.86
ICAM1	8	IVW	−0.07	0.0021	0.32	0.74
C4B[2,3]	101	IVW	0.02	0.0021	0.00	0.00
LCT[2]	25	IVW	0.04	0.0021	0.00	0.00
SIRPB1	25	IVW	−0.03	0.0023	0.01	0.60
C4A[2,3]	101	IVW	0.02	0.0023	0.00	0.00
PCSK7	18	IVW	−0.04	0.0028	0.02	0.98
IL9	12	IVW	0.06	0.0028	0.05	0.96
CLN5[1]	2	IVW	−0.15	0.0028	0.38	NA
AGT	4	IVW	0.12	0.0039	0.06	0.97
CD274	16	IVW	0.08	0.0042	0.02	0.46
RBP7	3	IVW	0.20	0.0047	0.09	NA
PLA2G7	14	IVW	0.04	0.0061	0.05	0.57
EGF	3	IVW	−0.13	0.010	0.11	NA
ASIP	3	IVW	−0.13	0.010	0.17	NA
TPSB2	10	IVW	−0.05	0.011	0.15	0.27
ACP1	15	IVW	0.04	0.011	0.04	0.88
RNASE3	6	IVW	0.09	0.011	0.01	0.26
A4GALT	2	IVW	−0.21	0.012	0.11	NA
DSCAM	3	IVW	−0.30	0.015	0.19	NA
COLEC11	8	IVW	−0.05	0.016	0.03	0.61

SPOCK2	8	IVW	0.08	0.017	0.03	0.98
VWA2	5	IVW	−0.11	0.018	0.04	0.76
RPN1	10	IVW	0.08	0.020	0.02	0.57
ADAMTS4	8	IVW	−0.06	0.020	0.19	0.36
PDCD1LG2	12	IVW	−0.09	0.020	0.04	0.07
PRTN3	2	IVW	0.17	0.020	0.06	NA
ADGRE2	10	IVW	−0.05	0.020	0.01	0.87
RNASE2	10	IVW	0.06	0.020	0.02	0.61
MPIG6B	18	IVW	−0.05	0.020	0.00	0.18
SIGLEC9	19	IVW	0.03	0.020	0.02	0.99
TAPBPL	10	IVW	−0.03	0.020	0.03	0.50
LRP12[1]	1	Wald ratio	0.73	0.023	0.66	NA
DNAJC30	3	IVW	−0.09	0.024	0.03	NA
CCL15	3	IVW	−0.08	0.024	0.12	NA
VTN	22	IVW	−0.03	0.028	0.01	0.75
NUCB1	3	IVW	−0.12	0.029	0.05	NA
TRH	4	IVW	−0.08	0.029	0.07	0.75
POSTN	8	IVW	−0.07	0.034	0.02	0.76
PLXNB2[1]	36	IVW	−0.02	0.034	0.08	0.09
IL18R1	23	IVW	0.03	0.034	0.01	0.79
IDUA	3	IVW	0.10	0.038	0.00	NA
CFD	10	IVW	−0.05	0.039	0.01	0.58
GGH[1]	4	IVW	0.08	0.040	0.02	0.49

FGFRL1	5	IVW	0.13	0.046	0.00	0.07
LMAN2L	3	IVW	−0.15	0.049	0.03	NA

[1] PD protein marker in Supplementary Table 3
[2] corrected by removing outliers by MRPRESSO
[3] has homolog detected by the same SOMAmer

744 proteins have a significant cis-pQTL

Table 3

List of SOMAmers with FDR<0.05 for the comparison PD patient vs. controls divided by subcohort (*GBA+*, *LRRK2+* and idiopathic PD patients).

Subcohort	Gene Symbol	PD change in direction	p-value (FDR adj)	Reference linking to PD
GBA+	CALCA	+	0.022	Svenningsson et al. 2017
	CD2	+	0.025	–
	DLK1	–	0.025	Jacobs et al. 2009
	GCH1	+	0.025	Puschmann 2013 Yoshino et al. 2018 Rudakou et al. 2019
	IL17A	+	0.025	Liu et al. 2017 Pengfei et al. 2018 Sommer et al. 2018 Green et al. 2019 Liu et al. 2019
	SEMG2	–	0.022	–
LRRK2+	ARSA	+	0.023	Lee et al. 2019 Angelopolou et al. 2020 Paciotti et al. 2020 Yoo et al. 2020
	ACP7	+	0.043	–
	CA10	+	0.037	–
	CTSB	+	0.037 0.027(#)	Blauwendraat et al. 2020
	SIAE	+	0.037	–
	SMPD1	+	0.037	Blauwendraat et al. 2020
	TENM4	–	0.037	Pu et al. 2020 Liang et al. 2021
Idiopathic	AK1	+	0.049	Garcia-Esparcia et al. 2015
	CCL14	+	0.023	–
	DLK1	–	0.032	Jacobs et al. 2009
	FRZB	+	0.036	–
	GPI	+	0.006	Knight et al. 2014
	HAMP	+	0.001	–
	LPO	–	1.5e-6	Agil et al. 2006 Kahn & Ali 2018 Muhammad et al. 2019
	LRRTM4	–	0.023	–
	NETO1	–	0.049	–
	NFH	+	0.014	–
	PTHLH	–	0.029	–
	PTPRR	–	0.032	–
	RAB31	–	0.044	Kumar et al. 2020

RELT	–	0.046	–
RIPK2	–	4.6e-5	Yan & Liu 2017
ROBO3	–	0.049	–
RSPO4	–	0.029	–
SEMG2	–	0.029	–
SHANK1	–	0.023	Hernandez et al. 2019
SPINK9	–	0.040	–
TXN	–	0.023	Domenico et al. 2010 (*)
VEGFA	–	0.006	Wu et al. 2016 Axelsen & Woldbye 2018
VIP	–	0.023	Gonzalez-Rey et al. 2005 Korkmaz et al. 2012

The change in PD represents (+) increased and (-) decreased in PD vs controls, respectively.
* Indirect evidence, # Interaction term