Metabolic profiling of CSF from people suffering from Sporadic and LRRK2 Parkinson's disease: a pilot study.

Ali Yilmaz (✉ ali.yilmaz@beaumont.org)

Oakland University William Beaumont School of Medicine

Zafer Ugur

Beaumont Health

Ilyas Ustun

DePaul University

Ray O. Bahado-Singh

Oakland University William Beaumont School of Medicine

Michael Maddens

Oakland University William Beaumont School of Medicine

Jan O. Aasly

Norsk Treteknisk Institutt

Stewart F. Graham

Oakland University William Beaumont School of Medicine

Research

Keywords: Parkinson Disease, 1H NMR, Targeted mass spectrometry, metabolomics, Machine learning.

Posted Date: April 24th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-20742/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Cells on October 31st, 2020. See the published version at <https://doi.org/10.3390/cells9112394>.

Abstract

Background: CSF from unique groups of Parkinson's disease (PD) patients were biochemically profiled to identify previously unreported metabolic pathways linked to PD pathogenesis and novel biochemical biomarkers of the disease were characterized.

Methods: Utilizing both ^1H NMR and DI-LC-MS/MS we quantitatively profiled CSF from patients with sporadic PD (n=20), those who are genetically predisposed (LRRK2) to the disease (n=20) and compared them with age,- and gender-matched controls (n=20). Further, we systematically evaluated the utility of several machine learning techniques for the diagnosis of PD.

Results: ^1H NMR and mass spectrometry-based metabolomics, in combination with bioinformatic analyses, provided useful information highlighting previously unreported biochemical pathways and CSF based biomarkers associated with both sporadic PD (sPD) and LRRK2 PD. Results of this metabolomics study further support our group's previous findings identifying Bile Acid metabolism as one of the major aberrant biochemical pathways in PD patients.

Conclusion: This study demonstrates that a combination of two complimentary techniques can provide a much more holistic view of the CSF metabolome and by association, the brain metabolome. Future studies for the prediction of those at risk of developing PD should investigate the clinical utility of these CSF based biomarkers in more accessible biomatrices. Further, it is essential that we determine if the biochemical pathways highlighted here are recapitulated in the brain of PD patients with the aim of identifying potential therapeutic targets.

Introduction

Parkinson's Disease (PD) is a progressive, adult-onset neurodegenerative disorder associated with the degeneration of dopaminergic (DAergic) neurons and the presence of proteinaceous inclusions such as α -synuclein (1). PD is the second most common neurodegenerative disorder after Alzheimer's Disease with a prevalence rate of 1% in people over 60 years of age (2). The etiology of PD is thought to be multifunctional in that genetic factors, environmental exposures, and aging contribute to disease development (3). Genetic studies revealed several causative monogenetic mutations that account for a small portion (10%) of PD cases, while the remaining cases are sporadic (4). The exact underlying mechanism for selective DAergic cell loss in PD remains elusive (5).

Clinically, the diagnostic criteria for PD are overwhelmingly based on impaired motor and non-motor functions and confirmed following improvements with dopamine treatment (6). However, the accuracy of such diagnostic criteria are low, as 40–50% of DAergic neurons have already degenerated by the time a clinical diagnosis is made (7). Moreover, premotor symptoms, including olfactory deficiency, obstipation, sleep disorders, and depression are very unspecific to PD, further complicating early diagnosis (8).

The development of molecular biomarkers, specific to PD, in body fluids such as urine, saliva, blood, and CSF would prove invaluable, allowing detection of the disease at the preclinical stage, monitoring the progression of the disease, and evaluating the therapeutic impact of disease modifiers. Metabolomics, or metabolic profiling, is the quantitative measurement of metabolites in cells, biofluids, tissues or organisms and accurately represents the biochemical phenotype of the organism in diseased and healthy state (9). In recent years, metabolomics has been widely applied for the study of PD and has shown great potential for early diagnosis(10, 11). CSF is a widely used biomatrix used for developing biomarkers of CNS disorders such as brain injury (12), Alzheimer's disease (13), and PD (7) due to its close proximity with the extracellular space in the brain. To the authors' knowledge, this is the first targeted metabolomics study combining two complimentary metabolomics platforms to provide a holistic view of the CSF metabolome and inversely, the brain metabolome to gain insight into PD pathophysiology and preclinical diagnosis.

Methods

CSF Samples

The study was approved by the Institutional Review Board at the Beaumont Research Institute and at St. Olav's Hospital (IRB# 201 2016/541, PD and markers for progression). Written informed consent was obtained from all participants who presented in the morning following an overnight fast. All participants were sequenced for the main pathogenic LRRK2 mutations prior to sample collection. A total of 80 subjects were enrolled in this study; CSF specimens obtained via lumbar puncture from participants suffering from sPD (n = 20), those genetically predisposed to the disease (LRRK2 acquired PD) (n = 20), and age- and gender-matched controls for each group to include patients with the G2019S mutation in the LRRK2 gene (n = 20) and those without (n = 20). Samples were immediately processed according to the Parkinson's Progression Markers Initiative (PPMI) recommendations (14).

¹H NMR Analysis

Samples were thawed at room temperature. CSF samples were filtered through 3-kDa cut-off centrifuge filter units (Amicon Micoron YM-3; Sigma-Aldrich, St. Louis, MO) to remove any protein. Centrifugal filters were washed 7 times prior to filtering CSF specimens to remove excess glycerol (15). An aliquot of 350 µl of each CSF sample was filtered at 13,000 *g* for 30 minutes to remove large macromolecules. The samples were prepared by transferring a 300 µL aliquot of filtered CSF fluid to a 1.5 mL Eppendorf tube followed by the addition of 35 µL of D₂O and 15 µL of a standard solution (3.73 mM disodium-2,2-dimethyl-2-silapentane5-sulphonate [DSS]), 233 mM imidazole, and 0.47% NaN₃ in H₂O, Sigma-Aldrich, St. Louis, MO). The CSF sample (200 µL) was then transferred to a standard 3 mm NMR tube. In total, 80 CSF samples were prepared in this manner, each containing 0.16 mM DSS, 10 mM imidazole, and 0.02% NaN₃ at a pH of 7.3 to 7.7.

All ^1H NMR spectral data were acquired at 300.0 (± 0.05) K using a Bruker Ascend III HD NMR spectrometer (Bruker-Biospin, USA) operating at 600 MHz equipped with a 5 mm TCI cryo-probe and z-gradient system; a Bruker SampleJet sample changer was used to transfer samples; samples were kept at 4°C while queued for analysis. Prior to analysis, samples were heated to room temperature over 3 minutes before being transferred to the magnet. ^1H NMR spectra were acquired at 300 K using the first transient of a standard NOESY-presaturation pulse sequence, chosen for its high degree of high quantitative accuracy (16). The duration of the 90-degree pulses was automatically calibrated for each individual sample using a homonuclear gated nutation experiment on the locked and shimmed samples after automatic tuning and matching of the probe head. Spectra were collected with 256 transients using a 5.1 s acquisition time and a 5 s recycle delay. Prior to spectral analysis, all FIDs were zero-filled to 64 k data points, and a line broadening of 0.5 Hz was applied. The methyl singlet of DSS served as an internal standard for chemical shift referencing (set to 0 ppm) and for quantification. All ^1H NMR spectra were processed and analyzed using the Chenomx NMR Suite Professional software package version 8.1 (Chenomx Inc., Edmonton, AB). Prior to statistical analysis, all NMR spectra were manually inspected for technical faults.

Targeted Mass Spectrometry Analysis

Using the commercially available AbsoluteIDQ p180 Kit (Biocrates Life Sciences AG) [19] we biochemically profiled the CSF specimens. The samples were processed according to the manufacturer's instructions. In brief, 30 μL of each CSF sample was mixed with isotopically labeled internal standards in a 96-well plate. Amino acids and biogenic amines were derivatized using 5% phenylisothiocyanate (PITC) and subsequent separation was performed on a BEH C18 column (2.1 mm $\times$ 75 mm, particle size of 1.7 μm) on a Waters iClass coupled to a Waters Xevo TQ-S (Waters, Milford, MA) operating in the multiple reaction mode. The subsequent lipid fraction was analyzed using direct flow, operating in multiple reaction monitoring mode. Metabolite concentrations were calculated and expressed as μM . Any metabolites below the limit of detection (LOD) were excluded from our analysis.

Bile-acid Analysis

Bile acids were analyzed using the commercially available Biocrates® Bile Acids Kit (Biocrates Life Science AG, Innsbruck, Austria) as described by Marksteiner et al. (17). This assay allows identification and quantification of primary and secondary bile acids as well as their taurine and glycine conjugated derivatives. The metabolite panel includes 17 individual bile acids, corresponding internal standards, and calibration ranges. Data analysis was performed with TargetLynx (Waters) and Biocrates® MetIDQ™ software.

Statistical Analysis

Univariate Data Analysis

Raw data were analyzed using MetaboAnalyst (v4.0). Normally distributed data were analyzed using a 2-tailed Student t-test and non-normally distributed data were analyzed using the Mann-Whitney U test.

Multivariate Data Analysis

To ensure no violation of the normality assumption data were standardized by sum normalization, followed by scaling to mean zero and unit variance, and log transformed. Before performing pattern recognition, metabolomics data from each individual group were subjected to explorative multivariate statistical data analysis by PCA to check for potential outliers or systematic variation ($p < 0.05$). This was further investigated by plotting T^2 -values (distance to model) against residual Q^2 (the variance not covered by the model) for each group. Following outlier detection, orthogonal projections to latent structures discriminant analysis (O-PLS-DA) was employed to determine if it would be possible to differentiate LRRK2 PD from the LRRK2 control, and sPD from the sPD control, and to identify features separating each group. For each model, the optimum number of components was assessed by a single 10-fold venetian blind cross-validation. The validity of the group separation for each model was evaluated through permutation testing with 2000 repeats. The raw metabolomics data were also subjected to unsupervised hierarchical clustering modelling to generate heat map representation between each group.

Machine LearningBased Regression Analysis

Before starting an intensive investigation on the predictive performance of various machine learning algorithms, missing values in the dataset were imputed by each group's median values for certain metabolites. Sum normalized data were subjected to log transformation and auto-scaling was applied for each metabolite. Following the data preprocessing steps, any metabolite concentration for a particular sample showing significant deviation within each group was replaced by the median value. This avoided violation of the normality assumption for any given metabolite. The data was then split into training (60% of data) and testing (40% of data) sets. The best set of predictor values were the combination of the most important variables obtained from the xgboost and random forest algorithms applied on the training set. The 4 most important variables were obtained from the 2 algorithms, which were then combined. Two of the variables were overlapping between the algorithms resulting in 6 features selected. Once the best set of metabolites were selected, some model hyper-parameters were optimized using 3-fold cross validation. The details of the optimization and the function of the important parameters can be found in the literature (18–20). To avoid overfitting, only 6 features were used, and the trained models were tested using a rather large hold-out set comprising 40% of the data.

Metabolite Pathway Enrichment Analysis

Metabolite pathway enrichment analysis (MSEA) was completed using MetaboAnalyst (v4.0) (21). Metabolite names were converted to Human Metabolite Database (HMDB) identifiers, and the raw data were imported in rows, normalized to the sum, and subjected to log transformation and auto scaling. The pathway-associated metabolite set was the chosen metabolite library, and all compounds in this library were used. Pathways with a Holm corrected pvalue and a q value < 0.1 were considered altered due to parkinsonism.

Data availability: Metabolomics data have been deposited to the EMBL-EBI MetaboLights database (22) with the identifier MTBLS863. The complete dataset can be accessed here <https://www.ebi.ac.uk/metabolights/MTBLS863>

Results

Statistical and Metabolite Pathway Enrichment Analysis

Using both ¹H NMR and targeted mass spectrometry we accurately identified and quantified 173 metabolites in CSF to include an additional 12 bile acids. Of the recorded metabolites, 17 were measured across both platforms and an average concentration was used. Supplementary **Figure S1** displays the box plot representation of the univariate analysis results comparing the mean concentrations of metabolites from each group. Of the 173 recorded metabolites, 15, and 28 metabolites reached statistical significance for each comparison, respectively ($p < 0.05$; $q < 0.1$). Table 1 reports the results of multi-group comparisons of important demographic factors such as age and gender. The results of the ANOVA revealed that both gender and age were not statistically significantly different between the groups.

Table 1
Multi-group comparison of available demographic information.

	sPD Control	sPD	LRRK2 Control	LRRK2 PD	p-value
n	20	20	20	20	
Age, mean (SD)	57.65 (9.56)	58.85 (8.95)	60.05(9.39)	59.43 (9.11)	0.45 ^a
Gender					
Male	10	11		34	0.26 ^b
Female	10	9		43	
^a One way ANOVA, ^b Chi-square Test					

For each group, the PCA score plots and T^2 vs. Q^2 plots clearly demonstrate that no patient sample was considered to be an outlier (**Figure S2**). Figure 1 illustrates the O-PLS-DA score plots for all the pair-wise comparison models and corresponding loading plots driving the separation. As evidenced from the score plots, classification of each group was successfully achieved. However, following 2000 permutations (**Table S1**), only the scores plot comparing LRRK2 PD with controls was significant ($p = 0.028$).

Following the multivariate analysis, pathway enrichment analysis of the mean concentration data revealed that a total of 5 metabolic pathways were statistically significantly ($P < 0.05$) disturbed in sporadic PD. These include fatty acid biosynthesis, ethanol degradation, ketone body metabolism, bile acid synthesis, and propionate metabolism (**Table S2**). We also identified 7 biochemical pathways to be significantly ($P < 0.05$) perturbed in LRRK2 PD CSF as compared with controls ($p < 0.05$; $q < 0.1$; **Table S3**)

and these include: spermidine and spermine biosynthesis, fatty acid metabolism, mitochondrial beta oxidation of long chain fatty acid (LCFA), methionine metabolism, bile-acid metabolism, inositol metabolism, and beta oxidation of short chain fatty acid (SCFA).

Machine Learning based Classification Analysis

The utility of different machine learning approaches for diagnosing parkinsonism was intensively evaluated. Performance metrics, model specific tuning parameters and the panel of metabolites used for each approach are listed in **Table S4**.

In the case of sPD diagnosis, the best models distinguishing sPD from controls were found to be logistic regression (AUC = 0.88), support vector machine (svm) linear (AUC = 0.88), and Smv kernel (AUC = 0.88). For distinguishing LRRK2 PD from corresponding controls, logistic regression (AUC = 0.94), and linear discriminant analysis (AUC = 0.94) yield the best performance (Fig. 2).

Discussion

To the authors' knowledge, this is the first reported study to identify sPD and LRRK2 PD specific metabolic signatures in CSF using a targeted metabolomics approach employing both ¹H NMR spectroscopy and mass spectrometry. Using complementary techniques allows us to detect and quantify many more metabolites and as such we require much more robust and accurate analytical data tools to help us discriminate between sample types. In this study, we used several machine learning algorithms to systematically evaluate our data. All the metrics have been reported in the supplementary information (**Table S4**). When comparing the multivariate approach, which is the traditional classification approach in the field of metabolomics, machine learning algorithms demonstrated superior performance for the diagnoses of PD. Particularly, logistic regression, random forest and support vector machine algorithms resulted in the best diagnostic accuracy (Fig. 2). This is because these algorithms work well with high-dimensional data and can deal with unbalanced and missing values. Multivariate modelling is all about getting a simple formulation of a frontier in a classification model problem that potentially tend to fail if there are non-linear boundaries between the groups aimed to be classified. On the other hand, Machine Learning algorithm seems to capture all patterns beyond any boundaries of linearity or even continuity of the boundaries between any groups implying superior classification performance. In a recent study, Stoessel et al. (2017) investigated PD-specific metabolic changes in CSF using non-targeted mass spectrometry and identified a panel of biomarkers for the diagnosis of PD. Using Random forest and PLS-DA, they reported models with an AUC = 0.74 and 0.73, respectively (23). Goldstein et al. (2012) reported models using biomarker candidates in CSF with 0.89 sensitivity and 0.80 specificity (24). In another study Hong et al. (2010) reported predictive model using DJ-1 and α -synuclein levels in CSF provided predictive accuracies equal to 0.77 with sensitivity of 0.94 and specificity of 0.50 (25). In a recent study by Mondello et al. (2014) α -Synuclein and ubiquitin carboxy-terminal hydrolase L1 (UCH-L1) levels in CSF were used to discriminate PD from controls. Both biomarkers discriminated PD from controls well with AUC of 0.82 with sensitivity of 0.87 and specificity of 0.79, respectively (26). In our study we go one step

further. We profiled sporadic and LRRK2 positive patients and produced powerful models with AUC = 0.88 (sensitivity = 1; specificity = 0.75) for sporadic and AUC = 0.94 (sensitivity = 0.88; specificity = 1) for LRRK2 PD patients. To the authors knowledge these are one of the most highly sensitive and specific models to be reported in the literature.

The primary objective of this study was to identify a panel of biomarkers which could accurately diagnose PD using CSF and to gain an insight into the biochemistry behind PD. To ensure we got the most accurate snapshot of the disease metabolome only samples from unmedicated participants were utilized to minimize any confounding factors.

Evident from the heat map, we can see that there are dysregulations in several metabolite groups due to parkinsonism (Fig. 3). Characteristics of the metabolites indicate perturbations in the carnitine, glycerophospholipid, sphingolipid, and amino acid metabolism in PD. MSEA identified a marked alteration in bile acid metabolism in sPD (Fig. 4a). Bile acids are form of cholinic acids synthesized from cholesterol in the liver (27). They are involved in many essential biological and metabolic cascades including glucose, lipid, cholesterol, drug metabolism and closely associated with intestinal hormones, microbiotas and energy balance (28). However, very little is known about the molecular mechanisms of bile acids in the central nervous system (29). In a study by Chun et al. (2012), they reported that UDCA (Ursodeoxycholic acid) reduced reactive oxygen species (ROS), reactive nitrogen species (peroxynitrite and nitric oxide), and helped to maintain intracellular glutathione (GSH) levels in a cell model of PD (30). Correspondingly, significant reduction in apoptosis markers such as nuclear fragmentation, caspase activation, and cytochrome were detected. Further, they reported that inhibiting phosphatidylinositol-3-kinase (PI3K) and Akt/PKB blocked the favorable effects of UDCA on SNP-induced cytotoxic cell death (30). The beneficial effect of UDCA on impairment of mitochondrial function has also been reported by Mortiboys et al.(2015) (31). Moreover, the naturally occurring taurine conjugate of UDCA (TUDCA) was tested for its neuroprotective effect in motor neuron disease (32).

Another important biochemical pathway found to be aberrant in the CSF of PD patients was taurine and hypotaurine metabolism. Taurine is a major intracellular free β -amino acid in mammalian tissues and intervenes in many physiological functions, such as neuromodulation, maintenance of calcium homeostasis, antioxidant and anti-inflammatory processes. The level of taurine has been reported to be elevated in the region of brain controlling the dopamine release and dopaminergic neuron activity (33). Moreover, taurine has been reported to reduce dopaminergic neurodegeneration and α -synuclein oligomerization through suppression of microglial M1 polarization via NOX2-NF- κ B pathway in a pesticide-induced PD model (34). Thus, one may hypothesize that the perturbation of this particular metabolic pathway could be a neuroprotective reaction by the brain.

It is also noteworthy that this is the first metabolomics study to report ethanol degradation to be significantly perturbed in the CSF of PD patients. Interestingly, one of the key enzymes involved in the ethanol degradation pathway is alcohol dehydrogenase (ADH). Mutations in ADH genes could play a role in the etiology of Parkinson's disease (PD) because of the important function they undertake, particularly

in retinoid and dopamine metabolism and/or aldehyde detoxification (35). In support of our hypothesis, Tan et al. (2001) reported that a polymorphism at allele A1 for ADH was correlated with an increased risk of PD (36).

Interestingly, propionate metabolism was found to be significantly disturbed between in sPD and corresponding controls. Propionate, the end-product of the microbial digestion of carbohydrates, presents together with other SCFA in the gastro-intestinal tract. Gut microbiota and their metabolic products are among potential candidates that could ignite a process that eventually leads to Lewy body formation in the enteric nervous system. In recent studies, differences in the abundance of certain gut microbiota and their metabolic products such as SCFA and propionate was reported (37).

MSEA also highlighted several metabolic pathways perturbed in the CSF harvested from LRRK2 PD patients (Fig. 4b). One such pathway was fatty acid metabolism. Perturbation in fatty acid metabolism in PD has also been previously reported (38). An enzymatic deficiency in either fatty acid breakdown or disturbance of fatty acid transport across the mitochondrial membrane due to defects in the carnitine transport system result in dysregulation of fatty acid (39). Supported by the significant change in the level of carnitines in CSF, we hypothesize that lipid metabolism is directly perturbed as a result of the change in carnitine levels. Brain acylcarnitines support lipid biosynthesis and the activity of antioxidants; they also enhance cholinergic neurotransmission (40). Further, when oxygen levels are low, the brain transitions from glucose metabolism to anaerobic respiration. Alternatively, it can also shift from glucose metabolism entirely, using fatty acids or ketones during pathological conditions such as neurodegeneration, hypoxia/ischemia, or post-traumatic brain injury (41). Therefore, the perturbation in fatty acid metabolism we observed could be a potential attempt at attenuating neuronal cell death, further supported by a significant change in the level of 1-methylhistamine which is a metabolite in histamine metabolism. Histamine, a neurotransmitter which is widely distributed throughout the human brain and an increase in it has been reported to be involved in the histaminergic system in PD (42). Notably, both Mitochondrial Beta oxidation of LCFA and Mitochondrial Beta oxidation of SCFA were significantly disturbed in LRRK2 PD. Taken together, these results suggest a profound change in energy metabolism.

Consistent with an essential role in cellular function, lack of inositol in cells leads to a rapid loss of viability (43). As inositol metabolic pathway was found to be perturbed, the association between LRRK2 gene and inositol metabolism needs to be further elucidated.

Our study is not without its limitations. Firstly, our small sample size limits what we deduce from the results. Secondly, the available clinical information is lacking such as UPDSR. This precluded us from considering key clinical variables in the models to optimize performance and to determine whether there are other confounding factors that could affect the metabolic profile.

Conclusion

This study introduces a novel approach to identify potential metabolic pathways and unique biochemical profiles associated with PD pathogenesis using CSF samples. Whereas conventional approaches utilize only one multivariate statistical method, here, the utility and predictive performance of several machine-learning methods was investigated. The machine-learning predictive model provided highly discriminating models for distinguishing LRRK2 PD from LRRK2 controls, and sPD from sPD controls, respectively. Results suggest, for the first time, that inositol-related metabolic pathways might be the key for better understanding the LRRK2 PD. Future studies for the prediction and diagnosis of those at risk of developing PD should investigate whether the biochemical pathways highlighted here are recapitulated in the brain of PD patients and if the CSF panel of biomarkers identified are also as effective for diagnosing or predicting those at greatest risk of developing PD in less invasively harvested biomatrices.

Abbreviations

sPD: Sporadic Parkinson; LRRK2: Leucine-rich repeat kinase 2; MS: mass spectrometry; ^1H NMR: proton nuclear magnetic resonance; CSF: Cerebrospinal Fluid; LCFA: long chain fatty acid; SCFA: short chain fatty acid; AUC: Area under curve

Declarations

Funding:

There is no funding to declare.

Acknowledgement:

We authors thank for generous contribution made by the Fred A. & Barbara M. Erb Foundation.

Authors' contributions:

SFG, AY and MM JOA conceived and designed the experiments; AY and ZU performed the experiments; AY and IU analyzed the data; AY, RBS, MM, IU and SFG. wrote the paper.

Availability of data and materials:

Supporting tables and figures can be reached in supplementary materials. Metabolomics data have been deposited to the EMBL-EBI MetaboLights database with the identifier MTBLS863. The complete dataset can be accessed here <https://www.ebi.ac.uk/metabolights/MTBLS863>

Consent for publication

Not applicable.

Competing interest

The authors declare no conflict of interest

Author details

Oakland University-William Beaumont School of Medicine, Rochester, MI, USA

Department of Obstetrics and Gynecology, William Beaumont Health, Royal Oak, MI, USA

Ethics approval and consent to participant

The study was approved by the Institutional Review Board at the Beaumont Research Institute and at St. Olav's Hospital (IRB# 201 2016/541, PD and markers for progression). Written informed consent was obtained from all participants

References

1. Graham S, Rey N, Ugur Z, Yilmaz A, Sherman E, Maddens M, et al. Metabolomic Profiling of Bile Acids in an Experimental Model of Prodromal Parkinson's Disease 2018.
2. de Lau LM, Breteler MM. Epidemiology of Parkinson's disease. *The Lancet Neurology*. 2006;5(6):525-35.
3. Klein C, Westenberger A. Genetics of Parkinson's disease. *Cold Spring Harbor perspectives in medicine*. 2012;2(1):a008888.
4. Li J-Q, Tan L, Yu J-T. The role of the LRRK2 gene in Parkinsonism. *Molecular Neurodegeneration*. 2014;9:47.
5. Chang KH, Cheng ML, Tang HY, Huang CY, Wu YR, Chen CM. Alternations of Metabolic Profile and Kynurenine Metabolism in the Plasma of Parkinson's Disease. 2018.
6. Jenner P. Molecular mechanisms of L-DOPA-induced dyskinesia. *Nature reviews Neuroscience*. 2008;9(9):665-77.
7. Trezzi JP, Galozzi S, Jaeger C, Barkovits K, Brockmann K, Maetzler W, et al. Distinct metabolomic signature in cerebrospinal fluid in early parkinson's disease. *Movement disorders : official journal of the Movement Disorder Society*. 2017;32(10):1401-8.
8. Goldman JG, Postuma R. Premotor and nonmotor features of Parkinson's disease. *Current opinion in neurology*. 2014;27(4):434-41.
9. Patti GJ, Yanes O, Siuzdak G. Innovation: Metabolomics: the apogee of the omics trilogy. *Nature reviews Molecular cell biology*. 2012;13(4):263-9.
10. Hatano T, Saiki S, Okuzumi A, Mohny RP, Hattori N. Identification of novel biomarkers for Parkinson's disease by metabolomic technologies. *Journal of neurology, neurosurgery, and psychiatry*. 2016;87(3):295-301.
11. Graham SF, Kumar P, Bahado-Singh RO, Robinson A, Mann D, Green BD. Novel Metabolite Biomarkers of Huntington's Disease As Detected by High-Resolution Mass Spectrometry. *Journal of proteome*

- research. 2016;15(5):1592-601.
12. Glenn TC, Hirt D, Mendez G, McArthur DL, Sturtevant R, Wolahan S, et al. Metabolomic analysis of cerebral spinal fluid from patients with severe brain injury. *Acta neurochirurgica Supplement*. 2013;118:115-9.
 13. Nagata Y, Hirayama A, Ikeda S, Shirahata A, Shoji F, Maruyama M, et al. Comparative analysis of cerebrospinal fluid metabolites in Alzheimer's disease and idiopathic normal pressure hydrocephalus in a Japanese cohort. *Biomarker research*. 2018;6:5-.
 14. Simuni T, Siderowf A, Lasch S, Coffey CS, Caspell-Garcia C, Jennings D, et al. Longitudinal Change of Clinical and Biological Measures in Early Parkinson's Disease: Parkinson's Progression Markers Initiative Cohort. *Movement disorders : official journal of the Movement Disorder Society*. 2018;33(5):771-82.
 15. Mercier P, Lewis MJ, Chang D, Baker D, Wishart DS. Towards automatic metabolomic profiling of high-resolution one-dimensional proton NMR spectra. *Journal of Biomolecular NMR*. 2011;49(3):307-23.
 16. Ravanbakhsh S, Liu P, Bjordahl TC, Mandal R, Grant JR, Wilson M, et al. Accurate, Fully-Automated NMR Spectral Profiling for Metabolomics. *PLOS ONE*. 2015;10(5):e0124219.
 17. Marksteiner J, Blasko I, Kemmler G, Koal T, Humpel C. Bile acid quantification of 20 plasma metabolites identifies lithocholic acid as a putative biomarker in Alzheimer's disease. *Metabolomics : Official journal of the Metabolomic Society*. 2018;14(1):1-.
 18. Liao S, Yang C, Ding L, editors. Approximate Parameter Tuning of Support Vector Machines. 2011 3rd International Workshop on Intelligent Systems and Applications; 2011 28-29 May 2011.
 19. Probst P, Boulesteix A-L, Wright M. Hyperparameters and Tuning Strategies for Random Forest2018.
 20. Anwar H, Qamar U, Qureshi AW. Global Optimization Ensemble Model for Classification Methods2014.
 21. Chong J, Soufan O, Li C, Caraus I, Li S, Bourque G, et al. MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis. *Nucleic Acids Research*. 2018;46(W1):W486-W94.
 22. González-Beltrán A, Maguire E, Hastings J, Griffin JL, Haug K, Rijnbeek M, et al. MetaboLights—an open-access general-purpose repository for metabolomics studies and associated meta-data. *Nucleic Acids Research*. 2012;41(D1):D781-D6.
 23. Stoessel D, Schulte C, Teixeira Dos Santos MC, Scheller D, Rebollo-Mesa I, Deuschle C, et al. Promising Metabolite Profiles in the Plasma and CSF of Early Clinical Parkinson's Disease. *Frontiers in aging neuroscience*. 2018;10:51-.
 24. Goldstein DS, Holmes C, Sharabi Y. Cerebrospinal fluid biomarkers of central catecholamine deficiency in Parkinson's disease and other synucleinopathies. *Brain : a journal of neurology*. 2012;135(Pt 6):1900-13.
 25. Hong Z, Shi M, Chung KA, Quinn JF, Peskind ER, Galasko D, et al. DJ-1 and alpha-synuclein in human cerebrospinal fluid as biomarkers of Parkinson's disease. *Brain*. 2010;133(Pt 3):713-26.

26. Mondello S, Constantinescu R, Zetterberg H, Andreasson U, Holmberg B, Jeromin A. CSF alpha-synuclein and UCH-L1 levels in Parkinson's disease and atypical parkinsonian disorders. *Parkinsonism & related disorders*. 2014;20(4):382-7.
27. Li T, Apte U. Bile Acid Metabolism and Signaling in Cholestasis, Inflammation, and Cancer. *Advances in pharmacology* (San Diego, Calif). 2015;74:263-302.
28. Liu Y, Rong Z, Xiang D, Zhang C, Liu D. Detection technologies and metabolic profiling of bile acids: a comprehensive review. *Lipids in Health and Disease*. 2018;17(1):121.
29. Lieu T, Jayaweera G, Bunnett NW. GPBA: a GPCR for bile acids and an emerging therapeutic target for disorders of digestion and sensation. *British journal of pharmacology*. 2014;171(5):1156-66.
30. Chun HS, Low WC. Ursodeoxycholic acid suppresses mitochondria-dependent programmed cell death induced by sodium nitroprusside in SH-SY5Y cells. *Toxicology*. 2012;292(2-3):105-12.
31. Mortiboys H, Furnston R, Bronstad G, Aasly J, Elliott C, Bandmann O. UDCA exerts beneficial effect on mitochondrial dysfunction in LRRK2(G2019S) carriers and in vivo. *Neurology*. 2015;85(10):846-52.
32. McMillin M, DeMorrow S. Effects of bile acids on neurological function and disease. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*. 2016;30(11):3658-68.
33. Ruotsalainen M, Ahtee L. Intrastriatal taurine increases striatal extracellular dopamine in a tetrodotoxin-sensitive manner in rats. *Neuroscience Letters*. 1996;212(3):175-8.
34. Che Y, Hou L, Sun F, Zhang C, Liu X, Piao F, et al. Taurine protects dopaminergic neurons in a mouse Parkinson's disease model through inhibition of microglial M1 polarization. *Cell Death & Disease*. 2018;9(4):435.
35. Buervenich S, Sydow O, Carmine A, Zhang Z, Anvret M, Olson L. Alcohol dehydrogenase alleles in Parkinson's disease. *Movement disorders : official journal of the Movement Disorder Society*. 2000;15(5):813-8.
36. Tan EK, Nagamitsu S, Matsuura T, Khajavi M, Jankovic J, Ondo W, et al. Alcohol dehydrogenase polymorphism and Parkinson's disease. *Neuroscience letters*. 2001;305(1):70-2.
37. Scheperjans F, Aho V, Pereira PA, Koskinen K, Paulin L, Pekkonen E, et al. Gut microbiota are related to Parkinson's disease and clinical phenotype. *Movement disorders : official journal of the Movement Disorder Society*. 2015;30(3):350-8.
38. Saiki S, Hatano T, Fujimaki M, Ishikawa K-I, Mori A, Oji Y, et al. Decreased long-chain acylcarnitines from insufficient β -oxidation as potential early diagnostic markers for Parkinson's disease. *Scientific reports*. 2017;7(1):7328-.
39. Hajri T, Abumrad NA. Fatty acid transport across membranes: relevance to nutrition and metabolic pathology. *Annual review of nutrition*. 2002;22:383-415.
40. Jones LL, McDonald DA, Borum PR. Acylcarnitines: role in brain. *Progress in lipid research*. 2010;49(1):61-75.

41. Mergenthaler P, Lindauer U, Dienel GA, Meisel A. Sugar for the brain: the role of glucose in physiological and pathological brain function. Trends in neurosciences. 2013;36(10):587-97.
42. Bulte F, Houghton D, Lowes H, Pyle A, Nesbitt S, Yarnall A, et al. metabolic profiling of Parkinson's disease and mild cognitive impairment. Movement disorders : official journal of the Movement Disorder Society. 2017;32(6):927-32.
43. Ishmayana S, Kennedy UJ, Learmonth RP. Preliminary Evidence of Inositol Supplementation Effect on Cell Growth, Viability and Plasma Membrane Fluidity of the Yeast *Saccharomyces Cerevisiae*. Procedia Chemistry. 2015;17:162-9.

Figures

Figure 1

(a)The O-PLS-DA scores plot of SPD control vs. SPD patients, (b) 10 most important metabolites on loadings for separation; (c) O-PLS-DA scores plot of LRRK2 controls vs. LRRK2 PD patients (c) 10 most important metabolites on loadings for separation (d).

Figure 3

Performance metrics of various machine learning algorithms evaluated for the prediction of i) LRRK2 PD as compared LRRK2 controls, and ii) sPD as compared to sPD controls.

Figure 7

Results of metabolic enrichment analysis showing metabolic pathways that are significantly disturbed when a) SPD were compared to SPD controls, and b) LRRK2 PD were compared to LRRK2 controls.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementarymaterials.docx](#)
- [Supplementarymaterials.docx](#)
- [Supplementarymaterials.docx](#)