Sebum: a window into dysregulation of mitochondrial metabolism in Parkinson's disease

Eleanor Sinclair[1], Drupad K Trivedi[1], Depanjan Sarkar[1], Caitlin Walton-Doyle[1], Joy Milne[1], Tilo Kunath[2], Anouk M. Rijs[3], Rob de Bie[4], Royston Goodacre[5] Monty Silverdale[6] and Perdita Barran*[1]

[1]Manchester Institute of Biotechnology, School of Chemistry, The University of Manchester, Princess Street, Manchester, UK, M1 7DN
[2]Institute for Stem Cell Research, School of Biological Sciences, The University of Edinburgh, Edinburgh UK, EH16 4UU
[3]Radboud University, Institute for Molecules and Materials, FELIX Laboratory, Toernooiveld 7c, 6525 ED, Nijmegen, The Netherlands
[4]Department of Neurology, Amsterdam University Medical Centers, University of Amsterdam, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands
[5]Institute of Integrative Biology, Department of Biochemistry, University of Liverpool, Crown Street, Liverpool L69 7ZB
[6]Department of Neurology, Salford Royal Foundation Trust, Manchester Academic Health Science Centre, University of Manchester, UK

Abstract

Parkinson's disease (PD) is a progressive neurodegenerative disorder, which is characterised by degeneration of distinct neuronal populations, including dopaminergic neurons of the *substantia nigra*. This causes complex non-motor symptoms in early PD, and subsequent impairment of motor function. A metabolomics profiling approach was conducted to identify diagnostic biomarkers of PD from sebum, a non-invasively available biofluid. In this study, we used liquid chromatography-mass spectrometry (LC-MS) to analyse 274 samples from participants (80 drug naïve PD, 138 medicated PD and 56 well matched control subjects) and detected metabolites that could predict PD phenotype. Partial least squares-discriminant analysis (PLS-DA) models based on this sebum metabolome had correct classification rates of 70.4% and 69.7% to distinguish between drug naïve PD and medicated PD from control, respectively. Variable importance in projection (VIP) scores indicate compounds with significance belonged to sphingolipid, triacylglycerol and fatty acid/ester lipid classes. Pathway enrichment analysis showed alterations in lipid metabolism and mitochondrial dysfunction *viz.* the carnitine shuttle, sphingolipid

metabolism and arachidonic acid metabolism. This study unveiled novel diagnostic sebum-based biomarkers for PD, and provides insight towards our current understanding of the pathogenesis of PD.

Introduction

Parkinson's disease (PD) is a neurodegenerative disorder affecting over 6 million globally, second only in prevalence to Alzheimer's disease.[1] The principal pathological hallmark of PD is the formation of aggregated α-synuclein deposits in the brainstem, which are the major components of Lewy bodies.[2,3] The disease is also characterised by the loss of dopaminergic neurons in the *substantia nigra pars compacta* producing a decline in striatal dopamine levels and subsequent loss of motor function.[4] There is no conclusive preclinical diagnostic test for PD. Clinical diagnosis is achieved primarily through observations by a physician, of the decline in motor functions.[5,6] These clinical manifestations normally present as a combination of one or more of the four cardinal signs of PD, namely; bradykinesia, resting tremor, rigidity and postural instability.[7,8] A formal diagnosis often occurs following the depletion of 60-80% of the brains dopaminergic neurons.[9] Non-motor symptoms are thought to precede motor symptoms by up to 20 years, some of these include; mood disorders, sleep disorders and olfactory deficits.[10,11] Seborrheic dermatitis is a common non-motor symptom reported in up to 60% of PD sufferers.[12,13] This condition presents as 'oily skin' that correlates to an excess of sebum, produced and secreted by the sebaceous glands in the dermis of the skin. Sebum is a complex lipid-rich substance that is predominantly composed of triglycerides, fatty acids, wax esters, squalene and cholesterol.[14] It serves as a protective agent to the skin providing waterproofing, thermoregulation and photoprotection, alongside suggested antimicrobial and antioxidant activities.[15,16] Studies of sebum are commonplace in dermatological conditions such as acne, however sebum as a biofluid has rarely been used in disease diagnostics. In our recent study, we have reported the presence and differential regulation of volatile organic compounds in the sebum of PD sufferers.[17]

The analysis of complex mixtures of metabolites present in a lipid-rich biofluid such as sebum, calls for a sensitive and robust analytical platform. Mass spectrometry (MS) is a

leading analytical technique for clinical metabolomics analyses and when hyphenated to chromatography, benefits from increased resolution and sensitivity.[18,19] Liquid chromatography-mass spectrometry (LC-MS) facilitates the qualitative and quantitative analysis of the wide range of molecular species found within complex mixtures such as sebum. LC-MS has been used to study a number of biofluids in relation to PD prognosis and diagnosis, such as blood, saliva and cerebrospinal fluid (CSF).[20–26] Alterations in the expression of metabolites and the downstream effects on their corresponding metabolic pathways have also been extensively studied for PD diagnostics within the blood and CSF metabolome, including: catecholamines, dopamine metabolites, amino acids and urate alongside fatty acid metabolism, energy metabolism and kynurenine metabolism.[23,27–31] The use of sebum as a diagnostic tool for PD provides an exciting prospect from which a novel non-invasive and inexpensive test could be developed to detect the onset of the disease. In this study, we have used LC-MS to separate and detect lipid-like species and small molecules present in sebum. We have used data-driven approaches, with suitable statistical validation, to discover biomarkers of Parkinson's disease present in sebum. This will inform the development of future PD biomarkers alongside the understanding of metabolic pathways altered in PD. Additionally, we also investigate whether variations in the measured sebum metabolome between early drug naïve PD and later medicated PD were observed, suggesting changes in the metabolic pathways during disease progression.

Results and Discussion

Analysis of Patient Metadata

The study population comprised of 274 participants which included 138 medicated PD, 80 drug naïve PD and 56 control subjects. An overview of important patient demographics is summarised in Table 1. The results of significance tests between cohort group metadata are reported in Table S1. Two-tailed Mann-Whitney U-test showed age is significantly different ($p < 0.05$) between control and PD cohorts (both drug naïve and medicated), however, BMI was not statistically significantly between these groups. There were more male participants in both PD cohorts (M/F > 1.5) compared to a higher proportion of female participants within the control group (M/F < 1). This was perhaps expected as the higher incidence and

prevalence rates of PD in the male population is recognised and studies show a 1.4-1.5 fold increase in the number of male PD cases, although the reason for this is not yet understood.[1,32] A similar comparison of the number of participants who smoke (yes/no) or consume alcohol (yes/no) showed no significant differences between drug naïve PD and control cohorts, with p-values of 0.837 and 0.192, respectively. However, the number of participants who consume alcohol was found to be 2.5 times higher in the control group compared to medicated PD. There were no smokers in the medicated PD cohort and 7 % within the control group which was deemed significant by a Fisher's exact test (p-value 0.006). The discovery of significant differences of these metadata parameters between PD and control cohorts has led us to test their impact on classification accuracy, which are described within the following results sections.

Table 1. Demographics of participants included in classification modelling and statistical analysis

Parameters	Independent Control	Drug Naïve PD	Medicated PD
N	56	80	138
Age (years)[b]	54.3 ± 14.4	69.8 ± 9.4	70.3 ± 8.2
BMI (kg/m²)[a,b]	26.1 ± 4.4	25.8 ± 4.9	26.3 ± 5.4
Gender (Male:Female)[c]	0.87	1.76	1.65
Alcohol Intake (Yes:No)[d]	4.60	1.76	1.81
Smoker (Yes:No)[d]	0.08	0.07	0.00

[a] Body Mass Index (BMI)
[b] BMI and Age values are expressed as mean ± standard deviation
[c] Expressed as a ratio (Male:Female)
[d] Expressed as a ratio (Yes:No)

<u>Data Driven Prediction of PD</u>

In order to assess variation between the measured metabolome by phenotype, partial least squares-discriminant analysis (PLS-DA) was used. Two PLS-DA models were constructed, each using a two-class input: (1) drug naïve PD *vs.* control and (2) medicated PD *vs.* control. It is well known that unbalanced numbers within classification groups may bias prediction accuracy towards the majority class and to overcome this here, Synthetic Minority Over Sampling Technique (SMOTE) was applied.[33] PLS-DA models were built and validated using bootstrap resampling with replacement (*n*=250). Figure 1 reports the classification sensitivity and specificity rates of each PLS-DA model alongside the observed and null distributions (from permutation testing).
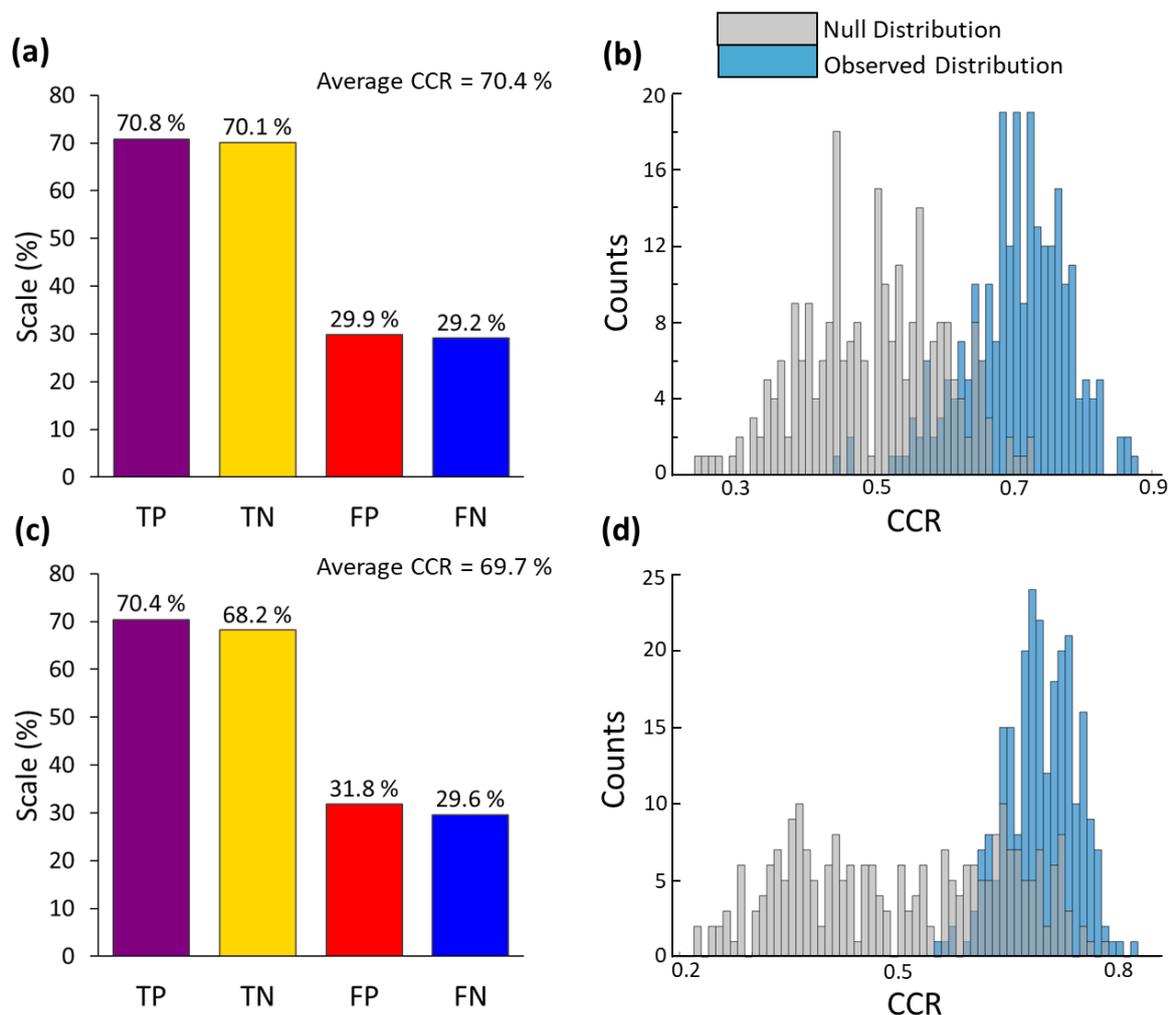

Figure 1: PLS-DA classification models for (a-b) drug naïve PD *vs.* control and (c-d) medicated PD *vs.* control. (a) and (c) report the classification rates for each model

including, true positive (TP, sensitivity), true negative (TN, specificity), false positive (FP) and false negative (FN). (b) and (d) display the null distribution (grey bars) and observed distribution (blue bars) for each PLS-DA bootstrap model. The correct classification rate (CCR) were calculated from the test sets only.

To evaluate if gender influenced classification accuracy, two PLS-DA models were built for each gender separately, for drug naïve PD *vs*. control and medicated PD *vs.* control. If the compounds accounting for variance between disease and control were gender specific, we could expect consistent and significantly higher sensitivity and specificity values for one gender, which we did not find to be true (see Table S3). Combined gender models (Figure 1) were used for subsequent analysis owing to the heightened power attributed to statistical models with larger input groups. PLS-DA was also used to determine if geographical location or variances between clinician sampling could impact classification using an independent control cohort. Samples (n=40) were chosen from four recruitment clinics, located in the north (n=2) and south (n=2) of the UK. Confounding factors were controlled so that age and BMI were not statistically significant between groups (one-way ANOVA p-value > 0.05) and the male-to-female ratio was identical. The average CCR for this model was 21% which therefore indicates that our data is not biased by recruitment site or the clinician who collected samples.

Selection of Significant Features Which Classify PD

To define the features responsible for the measured variance in PLS-DA prediction models, variable importance in projection (VIP) scores were calculated. Receiver operating characteristic (ROC) analysis was performed on variables with VIP score > 1 (Figure 2). The number of variables that met this threshold were 15 in Drug naïve PD analysis 26 in medicated PD analyses. The area under the curve (AUC) and 95% confidence intervals (CI) for each individual variable obtained from univariate ROC curve analysis are reported in the Supplementary Information Figure S2. A limitation in ROC analysis of individual features is the failure to consider relationships between the features that account for the observed variance. The outcome of a multivariate analysis is reduced to a univariate one, in which each

individual feature is treated as the sole biomarker accounting for 100% of the variation between the classes. Therefore, in combination with assessing individual metabolite ROC curves, a multivariate ROC analysis approach was also implemented based on the PLS-DA method (Figure 2a and 2b).
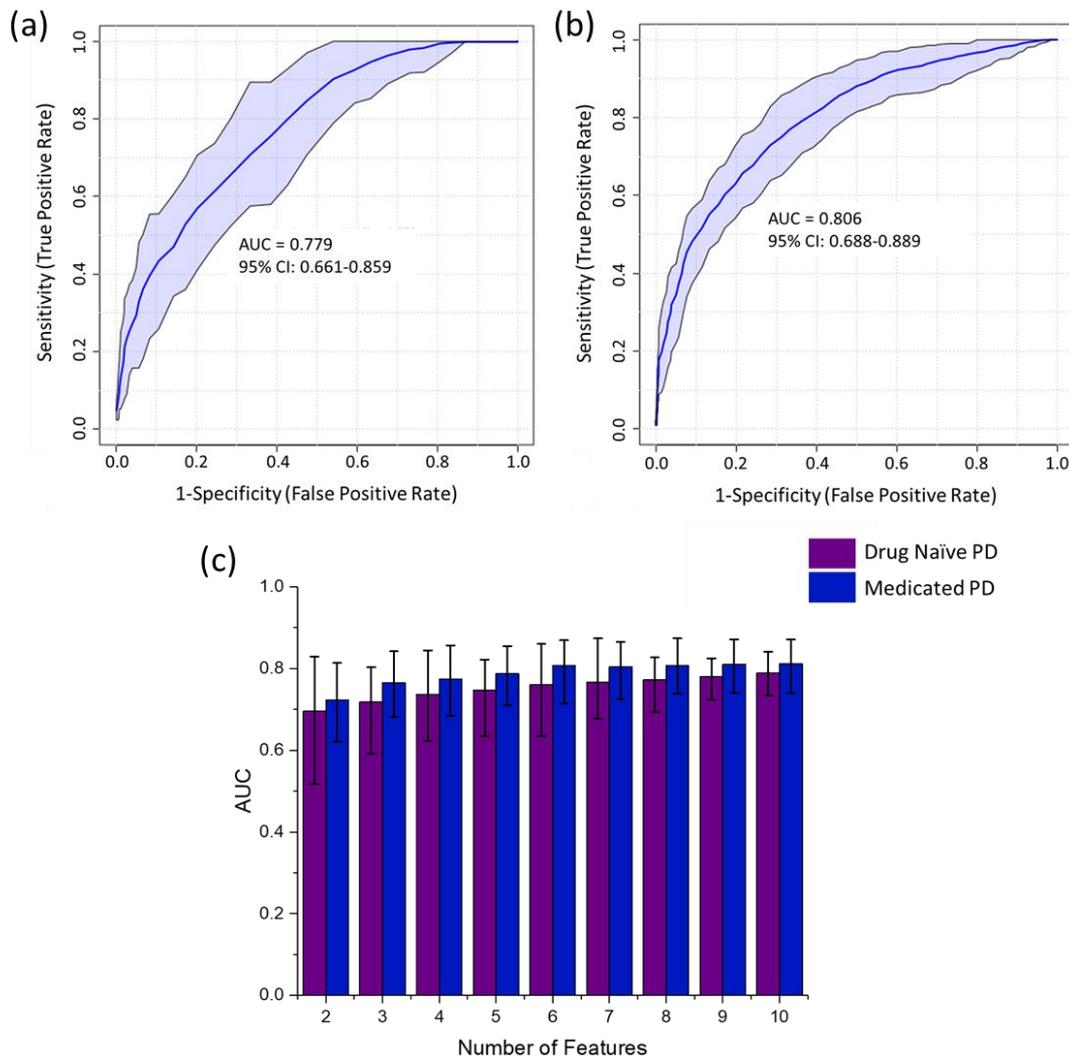

Figure 2: ROC curve analyses based on a multivariate PLS-DA algorithm with a two latent variable input, AUC and 95% confidence intervals (CI) were calculated by Monte Carlo cross validation (MCCV) using balanced subsampling with multiple repeats. (a) ROC curve analysis of metabolite features (*n*=15) in drug naïve PD *vs.* control PLS-DA with VIP >1, (b) ROC curve analysis of metabolite features (*n*=26) in medicated PD *vs.* control PLS-DA with VIP > 1, (c) a bar chart displaying the comparison of AUCs for drug naïve PD (purple) and medicated PD (blue) using common VIPs between models (*n*=10), error bars show the 95% CI range.

We note that PLS-DA could not accurately differentiate medicated PD and drug naïve PD. Sensitivity and specificity values of 59.7% and 50.3% were returned for PLS-DA models in which medicated PD was the 'positive' predicting class (data shown in Figure S1). Figure 2a and 2b report ROC curves for drug naïve PD and medicated PD models, respectively, which each use all VIP compounds > 1 for each respective model. VIP score examination of drug naïve PD *vs.* control and medicated PD *vs.* control models confirms that 10 variables (VIP > 1) are common between the two PD groups. To investigate biomarkers associated with the diagnosis of PD rather than disease stage stratification and to avoid possible effect of medication, the common metabolites (VIP > 1) between drug naïve and medicated PD analyses were evaluated further. Figure 2c presents a multivariate ROC analysis for each common variable and this analysis reports increased sensitivity and specificity rates as a function of the number of variables included in each model as demonstrated by higher AUC values. In addition, the 95% confidence interval range decreases as the number of variables in each model increases.

Pearson correlation coefficients were calculated for each significant variable (VIP > 1) to investigate association of alcohol and significant variables. None of the significant compounds are associated to an increase in alcohol consumption (Figure S3). To exclude the possible contribution of age to disease classification, each participant's age was included as a variable in additional PLS-DA classification models. The difference in CCR between models with and without the inclusion of age were negligible (< 0.5 %), and VIP scores for the age variable were 1.17×10^{-11} and 2.11×10^{-11}, respectively. In perspective, the variables were ranked at 6492 and 6498 out of a possible 6505 ranks, which strongly indicates that age is not a contributing factor for the separation presented.

Annotation of Metabolites Associated with PD Diagnosis

Metabolomics Standards Initiative (MSI) guidelines were adhered to,[34] for the annotation of common significant metabolites (*n*=10) (Table 2). Ceramides, triacylglycerol, glycosphingolipid, fatty ester and fatty acid class of metabolites were detected. Notably, metabolites belonging to ceramide, glycosphingolipid and fatty acid/ester classes were downregulated whereas triacylglycerol and fatty acid/ester metabolites were upregulated in

PD. Box plots comparing control, drug naïve PD and medicated PD cohorts for these compounds are displayed in Figure 3. Compound annotations for all metabolites with VIP score > 1 in drug naïve PD and medicated PD analyses are found in Tables S5 and S6, respectively.

Table 2. Putative identification of 10 compounds common between drug naïve PD and medicated PD analyses with VIP > 1.

Putative Compound Class Identification	MSI Level of Identification	m/z	Expression Drug Naïve PD (Fold Change)	Expression Medicated PD (Fold Change)
Ceramide	3	666.6370	↓ (0.60)	↓ (0.47)
Ceramide	3	638.6067	↓ (0.61)	↓ (0.47)
Ceramide	3	610.5763	↓ (0.63)	↓ (0.48)
Triacylglycerol	3	825.6939	↓ (0.77)	↓ (0.64)
Glycosphingolipid	3	764.5681	↑ (1.15)	↑ (1.10)
Fatty acid/Fatty ester	3	414.4308	↑ (1.23)	↓ (0.84)
Fatty acid/Fatty ester	3	358.3677	↓ (0.81)	↓ (0.78)
Fatty acid	3	194.1396	↑ (1.74)	↑ (1.78)
N/A[a]	4	550.6277	↑ (1.33)	↑ (1.10)
N/A[a]	4	368.4242	↓ (0.15)	↓ (0.14)

Fold change calculated with mean intensity values prior to log scaling. Denominator = Control

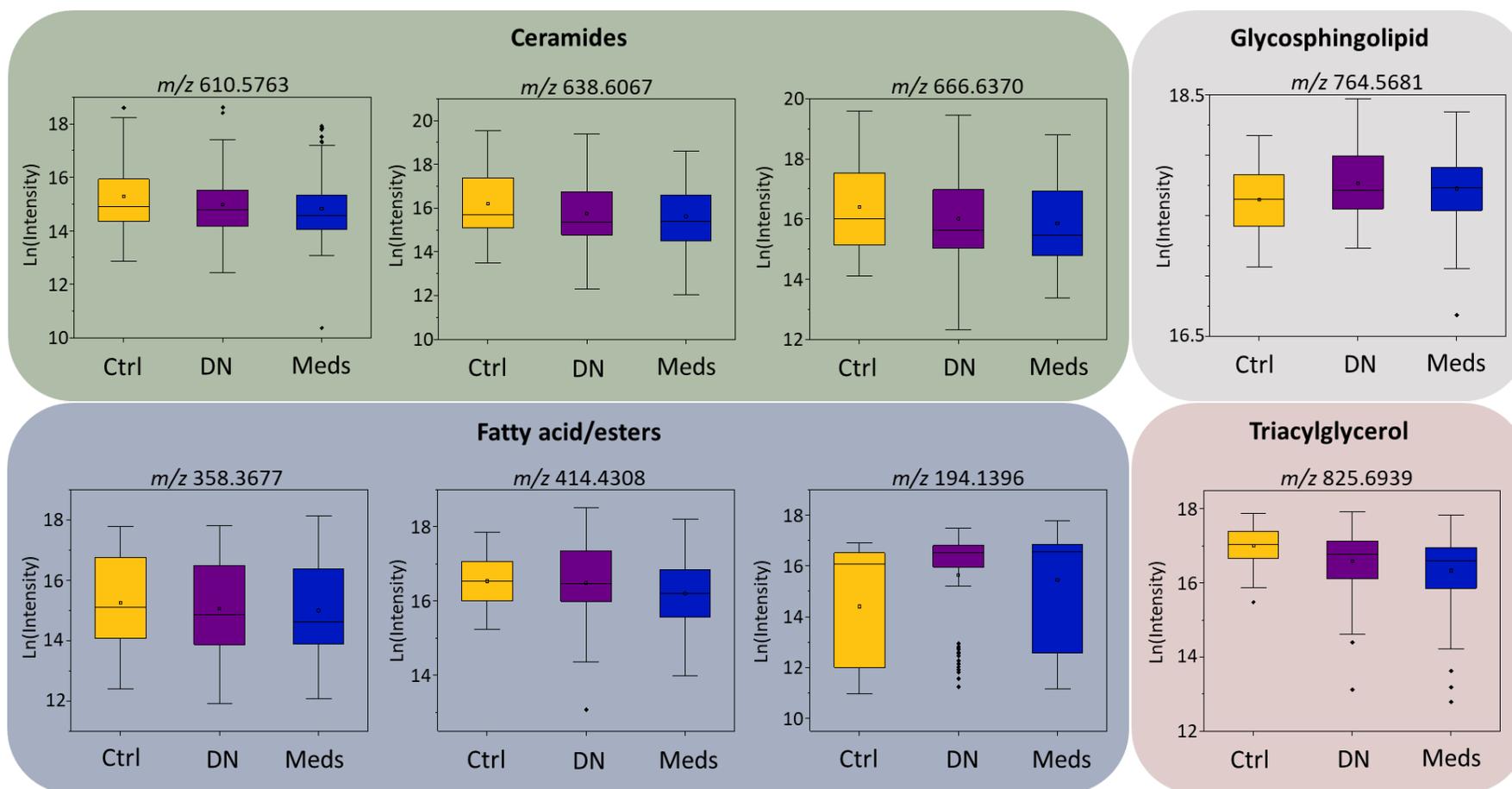[a] Level 4 annotation (i.e. unknown assignment)

Figure 3: Box whisker plots for each of the eight putatively annotated compounds for control (Ctrl, yellow), drug naïve PD (DN, purple) and medicated PD (Meds, blue) cohorts. The y-axis of each plot corresponds to the natural log of intensity values. Below each box plot is the structure of each chemical class and above the plot is the measured *m/z* value which correlates to Table 2

Sebum Metabolome Measurements: Context to Current Understanding of PD

Pathway enrichment analysis was performed to explore changes in metabolic pathways with respect to disease onset and progression. A prerequisite for traditional pathway analysis methods is the annotation of all analytically detected features *via* spectral and compound database matching. This is a major bottleneck in untargeted metabolomics workflows and due to the large number of features detected in this study, Mummichog analysis was employed.[35] The analysis was performed independently for the two PD cohorts using a Student's t-test (*p*-value < 0.05) between control subjects *vs.* (1) drug naïve PD and (2) medicated PD. There were 1378 and 504 features for drug naïve PD and medicated PD, respectively, which were significant between disease and control groups. Further details of significantly enriched pathways associated with PD can be found in Supplementary Information Table S5 and S6 for drug naïve PD and medicated PD, respectively.
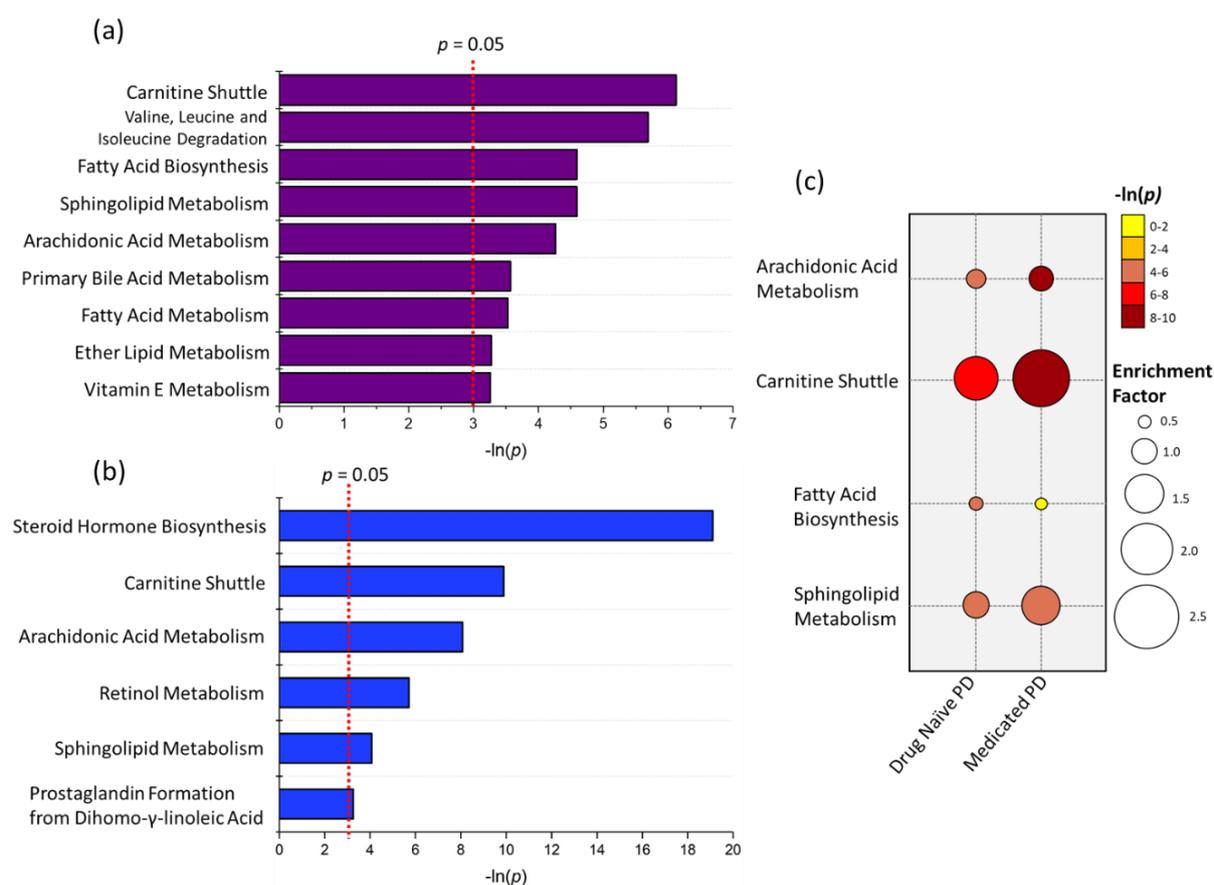

Figure 4. Bar charts showing the results of Mummichog pathway analysis of pathways with *p* < 0.05 for (a) drug naïve PD *vs.* control and (b) medicated PD *vs.* control, (c) a bubble chart

displaying the common significant pathways between drug naïve PD and medicated PD compared against controls; the bubble size refers to the enrichment factor of the pathway and the colour represents the natural log of the pathway *p*-value.

Mummichog analysis reveals the carnitine shuttle to be the most important pathway linked to drug naïve PD patients (*p*=0.002) (Figure 4a), which was found to increase in significance (*p*=5.09x10^{-5}) and enrichment within the medicated PD cohort, this can be visualised in Figure 4c. The carnitine shuttle is highly involved in energy metabolism through the facilitation of long chain fatty acid (LCFA) β-oxidation *via* assisted transportation into the mitochondria by acyl-carnitine substrates.[36] Decreased long-chain acyl-carnitines, associated with insufficient β-oxidation, have previously been reported as potential diagnostic markers for PD,[30] and the dysregulation of carnitine shuttle and vitamin E pathways have also been observed in frail elderly cohorts (between 56 and 84 years old) compared to resilient age-matched individuals.[37] The mapped *m/z* features correspond to a series of differing length fatty acid chains of acyl-carnitine conjugates. As the carnitine shuttle is a mediation pathway for fatty acid oxidation, it follows on that the perturbation of fatty acid biosynthesis and fatty acid metabolism pathways could be linked, which is further supported by the putative assignment of associated compound classes to VIP compounds.

Additional compounds putatively annotated from PLS-DA models (VIP > 1) belong to the sphingolipid class of compounds (Table 2). The sphingolipid metabolism pathway was enriched in both drug naïve and medicated PD. Sphingolipids are a major lipid class that are abundant in lipid-rich structures of the body (such as skin) and have central roles in cell signalling and regulation. Interestingly, disruption to the sphingolipid metabolism has been reported as a downstream effect of increased α-synuclein[38,39] and α-synuclein is disrupted in PD skin.[40] Alteration in the expression of lipids within the sphingolipid pathway consequently relate to lysosomal and mitochondrial degradation which are often implicated in the pathogenesis of neurodegenerative diseases such as PD and Gaucher's disease.[39,41,42] In particular, recent studies have found the dysregulation of ceramide levels, which are common structural units of all sphingolipids, in numerous diseases including PD, Alzheimer's disease and depression, although the general consensus from studies of sphingolipids in PD is an increase in ceramide levels.[43–45] Due to their bioactive role within cell membranes

sphingolipids are strongly linked to sterol metabolism pathways and have an established role in the modulation of steroidogenesis. There is a direct link between ceramides and the biosynthesis of cholesterol which is then the feed in substrate for steroid hormone biosynthesis, the most significantly altered pathway shown for medicated PD patients.[46,47]

Conclusions and Final Remarks

In conclusion, an untargeted LC-MS analysis of sebum obtained non-invasively from a simple skin swab from PD sufferers reveals a difference in the composition of sebum compared to control subjects. The overlap of ten metabolites from separate statistical analyses for drug naïve PD and medicated PD, strengthens the evidence that these compounds are associated with PD and not associated with dopaminergic medication. This is further supported by the identification of common pathways between the two PD classes that are significantly enriched. Insufficient clinical data is available for these patients to hypothesise on the ability of a sebum analysis to help stratify disease progression, although it should be included in further studies. Future work will also focus on targeting the putatively identified lipid classes, with the inclusion of ion mobility to enhance separation and increase the confidence in metabolite identification.

<u>Materials and Methods</u>

<u>Sample Participants:</u> The participants included in this study were part of a nationwide recruitment process taking place at 25 different NHS clinics, in addition to subjects (n=4) that participated in a clinical trial in the Netherlands.[48] A total of 274 participants were recruited from four subject groups: control (n=56), drug naïve PD (n=80), medicated PD (n=138). The participants included in this study were selected at random from these sites. Ethical approval for this project (IRAS project ID 191917) was obtained by the NHS Health Research Authority (REC reference: 15/SW/0354).

<u>Chemicals and Materials:</u> The chemicals and materials utilised in this study were: gauze swabs (Arco, UK), sample bags (GE Healthcare Whatman™, UK), 15 mL and 50 mL centrifuge tubes (Greiner Bio-One, UK), microcentrifuge tubes 2 mL (Eppendorf, UK), Ministart® 0.2

μm syringe filter (Sartorius, UK), Optima® LC-MS grade solvents 2-propanol, acetonitrile, methanol and formic acid (Fisher Scientific), HPLC grade HiPerSolv CHROMANORM® ethanol absolute (99.8%), CHROMASOLV^{TM} LC-MS grade water (Honeywell) and Leucine Enkephalin (Waters, Wilmslow, UK).

Sample Collection: Using a standard sampling procedure, each participant was swabbed by a clinician on the upper back with cotton-based medical gauze (7.5 cm x 7.5 cm) to collect sebum present on the skin. The sampled gauze swabs were sealed in background-inert plastic bags and transported to the central facility at the University of Manchester, where they were stored at - 80 ⁰C until end of recruitment.

Sample Extraction: Gauze swabs were removed from - 80 °C storage and allowed to equilibrate to room temperature. A solvent extraction method was used to prepare the samples for LC-MS analysis. Each gauze swab was transferred to an inert glass bottle. Methanol (9 mL) was added to each glass bottle and followed by vortex-mixing (10 s) and sonication (30 min) at ambient temperature, to extract sebum metabolites from gauze. The extracted metabolite-rich methanol was decanted from the gauze swab bottle and this solution was passed through a filter (0.2 μm). A recovery rate of approximately 7 mL per sample was achieved, which was aliquoted into three 2 mL fractions and one 1 mL fraction. Each 2 mL fraction was vacuum concentrated (Eppendorf) at ambient temperature for 12 h to remove methanol which resulted in three identical sebum extracts per patient sample. These dried pellets were stored at - 80 °C until required for analysis. A portion of the remaining 1 mL liquid fraction of each sample (100 μL) was used to create a biological pooled quality control (QC) sample. The mixture was vacuum centrifuged (Eppendorf) for 12 h at ambient temperature and the dried extract stored at - 80 °C until analysis.

Sample Reconstitution: Prior to LC-MS analyses dried sebum extracts were equilibrated to ambient temperature before reconstitution. Extracts were resuspended in 200 μL of MeOH:EtOH (*v/v*, 50:50). Samples were vortex-mixed (20 s), sonicated (5 min) and centrifuged (Eppendorf) at 12,000 *g* for 10 min. The recovered supernatant (160 μL) was then submitted for LC-MS analysis.

LC-MS Analysis: LC-MS analysis was performed on an Ultimate 3000 UHPLC (Thermo Scientific) coupled to a Synapt G2-S*i* QToF mass spectrometer (Waters). An ACQUITY UPLC

BEH C18 column (1.7 μm, 2.1 mm x 100 mm) heated at 55 °C was utilised for chromatographic separation. The mobile phases were as follows; mobile phase A was acetonitrile:water (*v/v* 60:40) with 0.1% formic acid, mobile phase B was isopropanol:acetonitrile (*v/v*, 90:10) with 0.1% formic acid. An injection volume of 5 μL was used. The flow rate was set at 0.6 mL/min and the gradient elution began at 40% B and increased to 50% B over 30 s, then to 69% B at 1.8 min, with a final ramp to 88% B at 6 min. The gradient was reduced back to 40% B and held for 1 min to equilibrate column. Full MS spectra were obtained for the mass range *m/z* 50-2000, whilst infusing Leucine-Enkephalin (*m/z* 556.2766) as an online mass calibrant to retain mass accuracy. MS settings were as follows: Synapt G2-*Si* MS was operated in Q-ToF mode. Capillary voltage was set to 3.0 kV, sampling cone voltage was set to 40 V, source temperature was kept at 120 °C, desolvation temperature was set to 550 °C and desolvation gas flow was 900 L/h.

Sample Sequence and Quality Control: Pooled QC samples were used to check analytical reproducibility both during analysis and during the data processing stages.[49] QC samples were injected at the beginning of each analytical batch (*n=3*), every 5th injection, and at the end of each analytical batch (*n=2*). Samples from 274 participants were stratified and randomised into 15 equal analytical batches. Each batch was reconstituted on the day of analysis to maintain sample integrity.

Data Pre-Processing and Deconvolution: LC-MS raw data were deconvolved using Progenesis QI (Waters, Wilmslow, UK). Peak picking, alignment and area normalisation were carried out with reference to a pooled QC. Features extracted from raw data were annotated using accurate mass match with METLIN, Human Metabolome Database (HMDB) and LipidMaps. The resulting peak table had 8765 metabolite features. Features that were absent in more than 10% of pooled QC injections throughout analysis were removed. From the remaining features those with more than 20% relative standard deviation (RSD) in peak intensity across pooled QC injections were also removed. The remaining peak set of 6205 metabolite features were robust features detected reproducibly throughout analysis within QC samples. The data were mean centred and auto-scaled and missing values were replaced with cubic spline interpolation in MATLAB 2019a (MathWorks) prior to statistical analysis.

Statistical Analysis: PLS-DA was performed for classification and prediction of data; resampling with replacement (bootstrapping) was used for model validation where the

correct classification rates (CCRs) from the Y-variable were computed for the (n=250) test data sets only. An in-house script was used in MATLAB (2019a) to performed PLS-DA. Univariate ROC analysis was performed in Origin (Version 2017, OriginLab Corporation, Northampton, MA, USA) and multivariate ROC curve based exploratory analysis was executed using MetaboAnalyst Biomarker Analysis (Version 4.0) in which the data matrix was auto-scaled and PLS-DA was used for the classification method and feature ranking method with a two latent variable input.

Pathway Analysis: Mummichog analysis was performed using MetaboAnalyst (Version 4.0). During mummichog analysis a list of all m/z features (L_{ref}) and a refined list of significant m/z features (L_{sig}) were generated using Student's t-test as the discriminatory test (p-value < 0.05). Significant m/z features were mapped onto a combination of metabolic models: Kyoto Encyclopedia of Genes and Genomes (KEGG), Biochemical Genetic and Genomic knowledgebase (BiGG) and the Edinburgh Model. Feature hits on known metabolite networks were tested against a null distribution produced from permutations of random m/z features from L_{ref} to yield significance values of metabolites enriched within any given network.[35]

Conflicts of interest
The authors declare no conflicts of interest.

Author Contributions
E.S.: Conception and design of work, data collection, data analysis and interpretation, drafting and editing of article. D.K.T.: Conception and design of work, data analysis and interpretation and editing of article. D.S. and C.W.D.: Conception and design of work and data collection. R.B.: Conception of work and sample collection. J.M., M.S., T.K., A.M.R and R.G.: Conception and design of work, interpretation of results and editing of article. P.E.B: Conception and design of work, supervision of data analysis, interpretation of results and editing of article.

Acknowledgements

References

1. Ray Dorsey, E. *et al.* Global, regional, and national burden of Parkinson's disease, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Neurol.* **17,** 939–953 (2018).

2. Dauer, W. & Przedborski, S. Parkinson's Disease: Mechanisms and Models. *Neuron* **39,** 889–909 (2003).

3. Kalia, L. V. & Lang, A. E. Parkinson's Disease. *Lancet* **386,** 896–912 (2015).

4. Obeso, J. A. Past, present, and future of Parkinson's disease. *Mov. Disord.* **32,** 1263–1263 (2017).

5. Marsili, L., Rizzo, G. & Colosimo, C. Diagnostic criteria for Parkinson's disease: From James Parkinson to the concept of prodromal disease. *Front. Neurol.* **9,** 1–10 (2018).

6. Goetz, C. G. *et al.* Movement disorder society-sponsored revision of the unified Parkinson's disease rating scale (MDS-UPDRS): Process, format, and clinimetric testing plan. *Mov. Disord.* **22,** 41–47 (2007).

7. Gelb, D. J., Oliver, E. & Gilman, S. Diagnostic criteria for Parkinson disease. *Arch. Neurol.* **56,** 33–39 (1999).

8. Jankovic, J. Parkinson's disease clinical features and diagnosis. *J Neurol Neurosurg Psychiatry* **79,** 386–376 (2008).

9. Przedborski, W. D. and S., Dauer, W. & Przedborski, S. Parkinson's Disease: Mechanisms and Models. *Neuron* **39,** 889–909 (2003).

10. Hawkes, C. H. The prodromal phase of sporadic Parkinson's disease: Does it exist and if so how long is it? *Mov. Disord.* **23,** 1799–1807 (2008).

11. Hawkes, C. H., Del Tredici, K. & Braak, H. A timeline for Parkinson's disease. *Park.*

Relat. Disord. **16,** 79–84 (2010).

12. Ravn, A.-H., Thyssen, J. P. & Egeberg, A. Skin disorders in Parkinson's disease: potential biomarkers and risk factors. *Clin. Cosmet. Investig. Dermatol.* **10,** 87–92 (2017).

13. Mastrolonardo, M., Diaferio, A. & Logroscino, G. Seborrheic dermatitis , increased sebum excretion, and Parkinson's disease : a survey of (im) possible links. *Med. Hypotheses* **60,** 907–911 (2003).

14. Stewart, M. E. & Downing, D. T. *Chemistry and Function of Mammalian Sebaceous Lipids. Advances in Lipid Research* **24,** (Academic Press, 1991).

15. Lovászi, M., Szegedi, A., Zouboulis, C. C. & Törőcsik, D. Sebaceous-immunobiology is orchestrated by sebum lipids. *Dermatoendocrinol.* **9,** 1–10 (2018).

16. Picardo, M. & et al. Sebaceous gland lipids. *Dermato-Endocronology* **1,** 68–71 (2009).

17. Trivedi, D. K. *et al.* Discovery of Volatile Biomarkers of Parkinson's Disease from Sebum. *ACS Cent. Sci.* **5,** 599–606 (2019).

18. Dunn, W. B., Broadhurst, D. I., Atherton, H. J., Goodacre, R. & Griffin, J. L. Systems level studies of mammalian metabolomes: the roles of mass spectrometry and nuclear magnetic resonance spectroscopy. *Chem. Soc. Rev.* **40,** 387–426 (2011).

19. Dunn, W. B. *et al.* Mass appeal: Metabolite identification in mass spectrometry-focused untargeted metabolomics. *Metabolomics* **9,** 44–66 (2013).

20. Shao, Y. & Le, W. Recent advances and perspectives of metabolomics-based investigations in Parkinson's disease. *Mol. Neurodegener.* **14,** 1–12 (2019).

21. Cova, I. & Priori, A. Diagnostic biomarkers for Parkinson's disease at a glance: where are we? *J. Neural Transm.* **125,** 1417–1432 (2018).

22. Havelund, J. F., Heegaard, N. H. H., Færgeman, N. J. K. & Gramsbergen, J. B. Biomarker research in parkinson's disease using metabolite profiling. *Metabolites* **7,** (2017).

23. Goldstein, D. S., Holmes, C., Lopez, G. J., Wu, T. & Sharabi, Y. Cerebrospinal fluid biomarkers of central dopamine deficiency predict Parkinson's disease. *Park. Relat. Disord.* **50,** 108–112 (2018).

24. Mattsson, N. CSF biomarkers in neurodegenerative diseases. *Clin. Chem. Lab. Med.* **49,** 345–352 (2011).

25. Engelborghs, S., Marescau, B. & De Deyn, P. P. Amino acids and biogenic amines in

cerebrospinal fluid of patients with Parkinson's disease. *Neurochem. Res.* **28**, 1145–1150 (2003).

26. Hong, Z. *et al.* DJ-1 and α-synuclein in human cerebrospinal fluid as biomarkers of Parkinson's disease. *Brain* **133**, 713–726 (2010).

27. Shao, Y. & Le, W. Recent advances and perspectives of metabolomics-based investigations in Parkinson's disease. *Mol. Neurodegener.* **14**, 1–12 (2019).

28. D'Andrea, G. *et al.* Different Circulating Trace Amine Profiles in De Novo and Treated Parkinson's Disease Patients. *Sci. Rep.* **9**, 1–11 (2019).

29. LeWitt, P. A. *et al.* Metabolomic biomarkers as strong correlates of Parkinson disease progression. *Neurology* **88**, 862–869 (2017).

30. Saiki, S. *et al.* Decreased long-chain acylcarnitines from insufficient β-oxidation as potential early diagnostic markers for Parkinson's disease. *Sci. Rep.* **7**, 1–15 (2017).

31. Cipriani, S., Chen, X. & Schwarzschild, M. A. Urate: A novel biomarker of Parkinsons disease risk, diagnosis and prognosis. *Biomark. Med.* **4**, 701–712 (2010).

32. Moisan, F. *et al.* Parkinson disease male-to-female ratios increase with age: French nationwide study and meta-analysis. *J. Neurol. Neurosurg. Psychiatry* **87**, 952–957 (2016).

33. Chawla, N. V, Bowyer, K. W., Hall, L. O. & Kegelmeyer, P. W. SMOTE: Synthetic Minority Over-Sampling Technique. *J. Artif. Intell. Res.* **16**, 321–357 (2002).

34. Sumner, L. W. *et al.* Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Inititative (MSI). *Metabolomics* **3**, 211–221 (2007).

35. Li, S. *et al.* Predicting Network Activity from High Throughput Metabolomics. *PLoS Comput. Biol.* **9**, (2013).

36. Longo, N., Frigeni, M., Pasquali, M., Biophys, B. & Author, A. Carnitine Transport and Fatty Acid Oxidation. *Biochim Biophys Acta* **1863**, 2422–2435 (2016).

37. Rattray, N. J. W. *et al.* Metabolic dysregulation in vitamin E and carnitine shuttle energy mechanisms associate with human frailty. *Nat. Commun.* **10**, 1–12 (2019).

38. Alecu, I. & Bennett, S. A. L. Dysregulated Lipid Metabolism and Its Role in α-Synucleinopathy in Parkinson's Disease. *Front. Neurosci.* **13**, 1–22 (2019).

39. Lin, G., Wang, L., Marcogliese, P. C. & Bellen, H. J. Sphingolipids in the Pathogenesis of Parkinson's Disease and Parkinsonism. *Trends Endocrinol. Metab.* **30**, 106–117

(2019).

40.   Donadio, V. *et al.* Skin nerve alpha-synuclein deposits: a biomarker for idiopathic Parkinson disease. *Neurology* **82**, 1362–1369 (2014).

41.   Xicoy, H., Wieringa, B. & Martens, G. J. M. The Role of Lipids in Parkinson's Disease. *Cells* **8**, 27 (2019).

42.   Indellicato, R. & Trinchera, M. The Link between Gaucher Disease and Parkinson's Disease Sheds Light on Old and Novel Disorders of Sphingolipid Metabolism. *Int. J. Mol. Sci.* **20**, 3304 (2019).

43.   Kurz, J., Parnham, M. J., Geisslinger, G. & Schiffmann, S. Ceramides as Novel Disease Biomarkers. *Trends Mol. Med.* **25**, 20–32 (2019).

44.   Plotegher, N., Bubacco, L., Greggio, E. & Civiero, L. Ceramides in Parkinson's Disease: From Recent Evidence to New Hypotheses. *Front. Neurosci.* **13**, 1–7 (2019).

45.   Mielke, M. M. *et al.* Plasma Ceramide and Glucosylceramide Metabolism Is Altered in Sporadic Parkinson's Disease and Associated with Cognitive Impairment: A Pilot Study. *PLoS One* **8**, 1–6 (2013).

46.   Gulati, S., Liu, Y., Munkacsi, A. B., Wilcox, L. & Sturley, S. L. Sterols and Sphingolipids: Dynamic Duo or Partners in Crime? *Prog. Lipid. Res.* **49**, 353–365 (2010).

47.   Liang, J. J. & Rasmusson, A. M. Overview of the Molecular Steps in Steroidogenesis of the GABAergic Neurosteroids Allopregnanolone and Pregnanolone. *Chronic Stress* **2**, 247054701881855 (2018).

48.   Verschuur, C. V. M. *et al.* Randomized delayed-start trial of levodopa in Parkinson's disease. *N. Engl. J. Med.* **380**, 315–324 (2019).

49.   Dunn, W. B. *et al.* Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nat. Protoc.* **6**, 1060–1083 (2011).