THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Validating differential volatilome profiles in Parkinson's disease

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published In:
ACS Central Science

OPEN ACCESS

Research Article

Validating Differential Volatilome Profiles in Parkinson's Disease

Eleanor Sinclair, Caitlin Walton-Doyle, Depanjan Sarkar, Katherine A. Hollywood, Joy Milne,
Sze Hway Lim, Tilo Kunath, Anouk M. Rijs, Rob M. A. de Bie, Monty Silverdale, Drupad K. Trivedi,
and Perdita Barran*

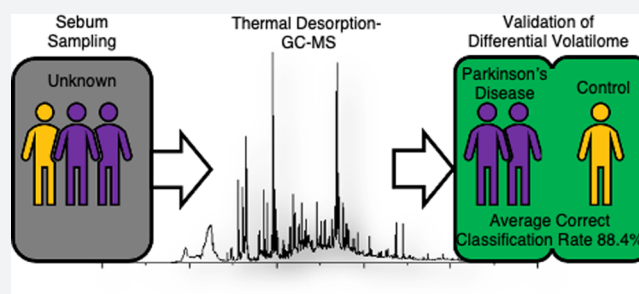Cite This: ACS Cent. Sci. 2021, 7, 300−306

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Parkinson's disease (PD) is a progressive neurodegenerative disorder that does not currently have a robust clinical diagnostic test. Nonmotor symptoms such as skin disorders have long since been associated with the disease, and more recently a characteristic odor emanating from the skin of people with Parkinson's has been identified. Here, dynamic head space (DHS) thermal desorption (TD) gas chromatography−mass spectrometry (GC-MS) is implemented to directly measure the volatile components of sebum on swabs sampled from people with Parkinson's—both drug naive and those on PD medications ($n = 100$) and control subjects ($n = 29$). Supervised multivariate analyses of data showed 84.4% correct classification of PD cases using all detected volatile compounds. Variable importance in projection (VIP) scores were generated from these data, which revealed eight features with VIP > 1 and $p < 0.05$ which all presented a downregulation within the control cohorts. Purified standards based on previously annotated analytes of interest eicosane and octadecanal did not match to patient sample data, although multiple metabolite features are annotated with these compounds all with high spectral matches indicating the presence of a series of similar structured species. DHS-TD-GC-MS analysis of a range of lipid standards has revealed the presence of common hydrocarbon species rather than differentiated intact compounds which are hypothesized to be breakdown products of lipids. This replication study validates that a differential volatile profile between control and PD cohorts can be measured using an analytical method that measures volatile compounds directly from skin swabs.

INTRODUCTION

The use of scent as an indicator for physical (and mental) biological functions has been associated with many disease states, such as diabetes *mellitus*, tuberculosis, liver and kidney disease, and cancer.[1] Although odor can be linked to many diseases, it is not a commonly applied approach to define biomarkers in clinical diagnostic tests. The application of canine detection within the medical field has progressed since a pioneering report of melanoma detection by Williams and Pembroke in 1989.[2−4] Some characteristic odors released in bodily secretions and excretions can occur before many other symptoms of the disease have developed as previously reported in Parkinson's disease (PD).[5,6]

Odorous skin secretions are most frequently related to areas of high sweat excretion and its subsequent bacterial degradation. However, sebaceous gland excretion is an alternative process that produces a biological matrix with a signature due to volatile organic compound(s) (VOC(s)).[7] The production of sebum from these glands occurs everywhere on the body, with the exception of the palms of hands and soles of feet, and is most evident on the face and upper trunk of the body which are categorized as sebum-rich locations.[8]

A noteworthy nonmotor symptom of PD is the development of skin-related disorders such as seborrheic dermatitis (SD),

which occurs in up to 60% of PD patients and manifests as an increase in sebum production.[9,10] The analysis of volatile species from skin secretions such as sebum and sweat is an under-developed methodology for clinical applications, and hence, there is no recognized standard procedure for its sampling.[11,12] Gauze swabs,[13] polydimethylsiloxane (PDMS) patches,[14] cigarette papers,[15] glass rollers or beads,[16,17] and more advanced methods such as wearable iontophoresis biosensors[18] have been previously reported, all of which have associated advantages and disadvantages in application.[19] These methods have primarily been directed to the collection of sweat rather than sebum, the waxy lipid-rich component.

The chemicals associated with these odorous profiles belong to the VOC class. There has been a rise in the development of analytical techniques to measure VOCs for clinical applications, most notably in the advancement of "electronic noses" and
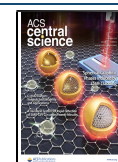
within the field of breathomics.[20−24] Analysis of VOCs using gas chromatography−mass spectrometry (GC-MS) is a popular nonselective analytical technique in which complex mixtures are separated prior to m/z detection of molecular ions and corresponding fragment ions to aid in structural elucidation. Headspace analysis, either dynamic or static, is a technique which promotes the volatile components of a sample to enter the gaseous "headspace" of an enclosed vessel by the partition of a concentration gradient, often using a temperature incubation prior to sampling to enhance the process.[25,26] While both techniques are effective, dynamic headspace (DHS) analysis has the benefit of concentrating volatile species through the continuous collection of the headspace which is an obvious benefit for untargeted metabolomics.[25]

This study leads on from earlier work which demonstrated that there was a differential VOC signature in Parkinson's disease patients.[5] This subsequent study provides a completely independent validation set of data using a different instrument alongside new patient samples.

■ EXPERIMENTAL METHODS

Sample Participants. The participants included within this study were part of a nationwide recruitment process taking place at 25 different NHS clinics across the UK. This study consists of a subset of 129 participants from three subject groups: independent controls ($n = 29$), drug naïve Parkinson's disease participants ($n = 17$), and medicated Parkinson's disease participants ($n = 83$). Patient demographics are reported in Table S1. Ethical approval for this project (IRAS project ID 191917) was obtained by the NHS Research Authority (REC references: 15/SW/0354).

Sample Collection. Each participant was swabbed on the upper back with cotton-based sterile medical gauze (7.5 cm × 7.5 cm) to collect sebum present on the skin. Participants were asked not to wash 24 h before sebum collection. We have not controlled for diet or water intake, but we expect this variation to be more random in subjects than the effect of disease. The patient sampled gauze was sealed in background-inert plastic bags and transported to the central facility at the University of Manchester, where they were stored at −80 °C until analysis.

Chemicals and Materials. The chemicals and materials used in this study were gauze swabs (Arco, UK), sterile sample bags (GE Healthcare Whatman, UK), 20 mL glass vials (GERSTEL, Germany), TENAX TA thermal desorption tubes and liners packed with TENAX TA for CIS 4/6 liner for the thermal desorption unit (TDU) (GERSTEL, Germany), Optima LC-MS grade methanol (Fisher Scientific), and HiPerSolv CHROMANORM absolute ethanol 99.8% purity (VWR Chemicals). The QC sample was composed of a mixture of seven compounds, each sourced from Sigma-Aldrich; L(−)-carvone (27.0 μM), δ-decalactone (96.4 μM), ethyl butyrate (150.5 μM), ethyl hexanoate (30.5 μM), hexadecane (43.7 μM), nonane (83.07 μM), and vanillin (100.8 μM), all in MeOH:EtOH (9:1). Octadecanal (Apollo Scientific) and eicosane (Sigma-Aldrich) standards were both 99% purity.

DHS-TD-GC-MS Analytical Method. Thawed gauze swabs were aseptically transferred into 20 mL glass vials and analyzed by DHS-TD-GC-MS (TD, thermal desorption). During DHS preconcentration, the samples were incubated at 80 °C for 10 min to promote the concentration of VOCs in the vial headspace. Trapping was initiated with dry nitrogen as the purge gas at a flow rate of 70 mL min^{-1} for a total gas volume of 1 L. The volatile compounds were captured on a TENAX TA

adsorbent tube (GERSTEL, Germany) held at 40 °C. The adsorbent tube was transferred from the DHS unit to the TD unit (TDU) using an automated GERSTEL MPS dual head workstation, at which time the captured VOCs were desorbed from the TENAX sorbent. The TDU was operated in splitless mode and was held at 30 °C for 1 min before the application of a 12 °C s^{-1} temperature ramp and held at 280 °C for 5 min. The desorbed analytes were focused in a cooled injection system (CIS) which was operated in solvent vent mode, using a vent flow of 80 mL min^{-1}. The CIS was held at 10 °C for 2 min before a second 12 °C s^{-1} temperature ramp and maintained at 280 °C for 5 min. The CIS vent valve (and flow) was not initiated until 3 min after analyte release on the GC column and can therefore be classed as a splitless method. The GC analysis was performed on an Agilent GC 7890A coupled to an Agilent MSD 5975 system interfaced by an electron impact (EI) source. Separation was induced by an Agilent VF-5MS column (30 m × 0.25 mm × 0.25 μm). The column flow was kept at 1 mL min^{-1}. The oven ramp was programmed as follows: 40 °C held for 1 min, 25 °C min^{-1} to 180 °C, 8 °C min^{-1} to 240 °C and held for 1 min, and 20 °C min^{-1} to 300 °C held for 2.9 min for a total run time of 21 min. The transfer line to the MS was maintained at 300 °C, the EI source at 230 °C, and the Quadrupole at 150 °C. The mass-selective detector (MSD) was operated in scan mode for a mass range between 30 and 800 m/z. The GC-MS system is fitted with a GERSTEL olfactory detection port (ODP3) using Agilent Technologies capillary flow technology (three-way splitter plate equipped with makeup gas). This was controlled as a 1:1 (v:v) split between MSD and ODP, and therefore, the GC eluent was diluted by a factor of 2 with helium makeup gas.

Patient Sample Analysis. Gauze samples from 129 subjects were analyzed across seven stratified, randomized, and blinded analytical batches. Quality control (QC) samples were injected at the beginning ($n = 3$), after every fifth injection, and at the end ($n = 3$) of each batch analysis. QC samples (5 μL) were used to gauge the analytical reproducibility in batch-to-batch analyses, which were analyzed on a shorter DHS-TD-GC-MS method to enable a faster analysis time per analysis batch. An example chromatogram for a PD sample and a control sample is shown in Figure S1A,B, respectively.

Analyte of Interest Standards Analysis. Octadecanal and eicosane standards were run on an identical DHS-TD-GC-MS method to patient samples. Headspace sampling was performed with 10 μL of octadecanal (37.4 μM) and eicosane (35.4 μM) in MeOH which were individually analyzed in 20 mL vials.

Lipid Standards Analysis. Lipid standards were purchased from Avanti Polar Lipids [L-α-phosphatidylethanolamine (Brain, Porcine) (PE), L-α-phosphatidylserine (brain, porcine) (sodium salt) (PS), and 1′,3′-bis[1,2-dioleoyl-sn-glycero-3-phospho]-glycerol (sodium salt) (18:1 cardiolipin) (CL)] and were diluted in CHCl$_3$ while L-α-phosphatidylcholine (brain, porcine) (PC) was diluted in MeOH. Glucosylsphingosine (GlcSph) (Sigma-Aldrich) was diluted in MeOH. All standards were analyzed using a slightly varied analytical method to that of the patient samples; the changes were as follows: a 5 min solvent vent was applied in addition to a further 4 min high-temperature hold at the end of the GC temperature gradient. The analytical method was 25 min in total. From a standard solution of 100 μM total, a 20 μL portion was analyzed in separate headspace vials. Lipid standards were measured in triplicate alongside blank headspace vials and blank solvent analyses (both MeOH and CHCl$_3$) which were measured both prior to and after lipid standard analysis. A comparative example of a lipid standard (L-

α-phosphatidylethanolamine (Brain, Porcine) (PE)), blank headspace vial, and solvent-only analysis is reported as overlaid chromatograms in Figure S1C.

Data Preprocessing and Deconvolution. All TD-GC-MS patient sample and lipid standard data were converted to open source mzML format using ProteoWizard.[27] The data set was deconvolved using an in-house script using the eRah package for R.[28] The deconvolved analytes were assigned putative identifications by matching fragment spectra with compound spectra using the Golm database. The resulting matrices comprised variables and their corresponding peak areas per sample. The patient sample data set comprised 671 features, and these were further refined by the removal of features absent in more that 5% of all samples, which generated a reduced data set of 520 features.

Features present in lipid standard data were filtered based upon a set of further criteria: (a) match factor (MF) > 75, (b) detection only in lipid sample analysis and not present in either the blank headspace vial or solvent-only analyses, (c) abundance > 1000 counts, and (d) each feature present in at least two of three triplicate analyses.

Statistical Analysis. We used a two-pronged approach with the statistical investigation of data. A data-driven multivariate approach was used to validate if there was a differential VOC signature associated with PD. Further, a targeted approach was used to verify selective analyte of interest peaks. Data were autoscaled and all missing values replaced using spline interpolation prior to statistical analysis. To account for variances in sebum production between participants, all samples were normalized to their respective TIC. There was less than an order of magnitude difference between samples which had the highest and lowest summed total ion count (TIC) (1.23×10^8 and 2.73×10^7). Of the five samples with the highest summed TIC, one was the control, and four were PD; of the five samples with the lowest summed TIC, one was the control, and four were PD. To further investigate that effect of biomass variations, we considered the relative intensities of common ions in high- and low-response samples in both cohorts. The four highest and four lowest summed TIC samples for both PD and the control were chosen, totalling 16 reference samples. Features not present in all of these 16 samples were removed, and all features were normalized to their respective TIC. There were no trends in the features as a proportion of summed TIC, demonstrating that the amount of sebum sampled does not necessarily correlate to the composition. Examples of plots of normalized intensities in four representative features are reported in Figure S2A−D.

The synthetic minority oversampling technique (SMOTE) was implemented to balance the sample numbers within each class and hence remove possible bias to the majority class.[29] Partial least-squares-discriminant analysis (PLS-DA) was executed in MATLAB (2019a) for classification and prediction of classes within data.[30,31] Models were validated by resampling bootstrapping (n = 250). Multivariate receiver operating characteristic (ROC) analysis was performed using Metab-oAnalyst Biomarker Analysis (Version 4.0);[32] PLS-DA was implemented for classification and feature ranking with a two latent variable input. PLS-DA score plots for PD vs control and drug naïve PD vs medicated PD classification models are reported for reference in Figure S3A,B, respectively. ROC curves were generated by balanced Monte Carlo cross validations (MCCVs) in which two-thirds of the samples were used to evaluate feature performance, and the remaining one-third were used to validate the classification. Iterations of this process (n =

30) were performed to calculate model performance and calculate confidence intervals for area under the curve (AUC) for each model.[33] Like any metabolomics data, sebum produces a highly complex set of features. This is mostly due to high-resolution mass spectrometry and very sensitive detectors. Principal component analyses (PCAs) are often successful for classification purposes when the data come from a really clean system such as microbiome data where metabolism is well-defined, and perturbation responses are very unique, i.e., KO vs WT or growth condition differences. With human biological samples, complexity is very high, and often PCA or MDS are not the best approaches to classify data due to subject-to-subject variations.[5] At best, PCA is a good tool for dimensionality reduction in such cases.

PLS-DA models without appropriate validations certainly can be prone to overfitting or just picking up on noise. This study was blinded, performed in batches, and randomized, so if any analytical artifact was introduced, it would not affect only one class in our supervised approach. Any random effect would make the model worse than what we have shown. We believe this is encouraging because without highly controlled data set, the volatilome between PD and the control is so different that we have high classification accuracy by modeling this data. With future work where we can have better libraries to annotate metabolites of interest, control for diet, and/or exposome under our new ethics, lesser noise will further improve supervised models.

■ RESULTS AND DISCUSSION

Validating Changes within VOC Profiles with the Onset of PD. To assess and validate if there are measurable discriminant VOCs in sebum, PLS-DA models were generated from new independent data. Models were generated to analyze (i) the classification accuracy between PD (drug naïve and medicated) and control samples and (ii) drug naïve PD against medicated PD participants to investigate differences in the disease phenotype alongside any possible effects of medication on this measured phenotype. An average correct classification rate (CCR) of 84.4% was obtained by combining the drug naïve and medicated PD cohorts in which the CCR was validated via bootstrapping (n = 250) (Figure 1). This prediction rate is similar to the average CCR (86%) achieved by PLS-DA with 5-fold cross validation reported in our pilot study.[5] Patient demographics were assessed to determine if they impacted the classification accuracy within the model. This is discussed in detail within the Supporting Information alongside significance tests for metadata parameters (Table S2). In summary, no significant confounding effects were observed on PLS-DA classification due to gender (Figure S4), BMI, alcohol intake or smoking (Figure S5), or age.

This classification validates findings from our initial TD-GC-MS study which revealed a unique VOC signature that can be associated with PD patients and be measured by the headspace sampling of sebum collected noninvasively. It is noted that the measured VOC profiles between drug naïve PD and medicated PD cohorts did not classify well, with a CCR of 66.7%, and had poor validation by permutation testing (Figure 2). This indicates that the sampled VOCs do not have sufficient discriminatory power to classify between drug naïve and medicated PD but can distinguish PD from the control, indicating that the associated odor profile is fundamental to PD. It can therefore be hypothesized that there is minimal change within the VOC profile after the onset of PD, with treatment, and/or that our
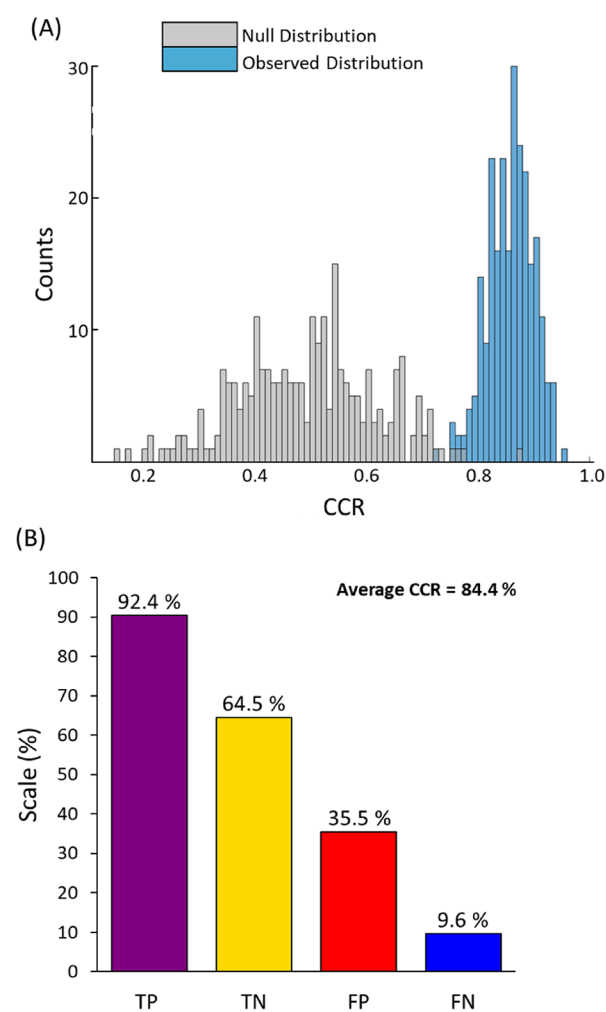
Figure 1. PLS-DA classification model for a two-class input using combined drug naïve and medicated PD cohorts vs controls. (A) Histogram reporting the distribution of the correct classification rate for the null (gray) and observed (blue) distributions obtained from bootstrap validation ($n = 250$) of the PLS-DA classification model. (B) Chart displaying the true positive (TP), true negative (TN), false positive (FP), and false negative (FN) classification rates.

Figure 2. PLS-DA classification of drug naïve PD vs medicated PD in which medicated PD was the positive predictive class. (A) Histogram reporting the distribution of correct classification rates (CCRs) for the null (gray) and observed (blue) distributions obtained from bootstrap validation ($n = 250$) of the PLS-DA classification model. (B) Chart reporting the classification rates from the PLS-DA model.

classification of the disease based on drug naïve vs medication does not translate well to clinical PD staging.

Discriminatory Features in PLS-DA Classification Modeling. Variable importance in projection (VIP) scores were calculated for the two-class PLS-DA model: combined PD vs controls. A multivariate ROC analysis was performed for compounds ($n = 12$) with VIP score >1 (Figure 3A). The area under the curve (AUC) values increase as the number of features included in the model increases, in conjunction with a reduction in the confidence interval (CI) range. The predictive power of the model increases going from a two-variable classifier up to seven-variable classifier; however, it begins to plateau following this threshold. The compounds that had high VIP scores (top 10%) were further selected ($n = 50$), and ROC analysis was performed, showing the same trend of increasing AUC values until a plateau between the 5-variable and 10-variable models, which is consistent with the previous ROC analysis (VIP > 1). This is presented in Figure S6.

Mann−Whitney nonparametric U tests between PD and control cohorts were performed for selected compounds (VIP > 1), and eight of these compounds were found to be statistically
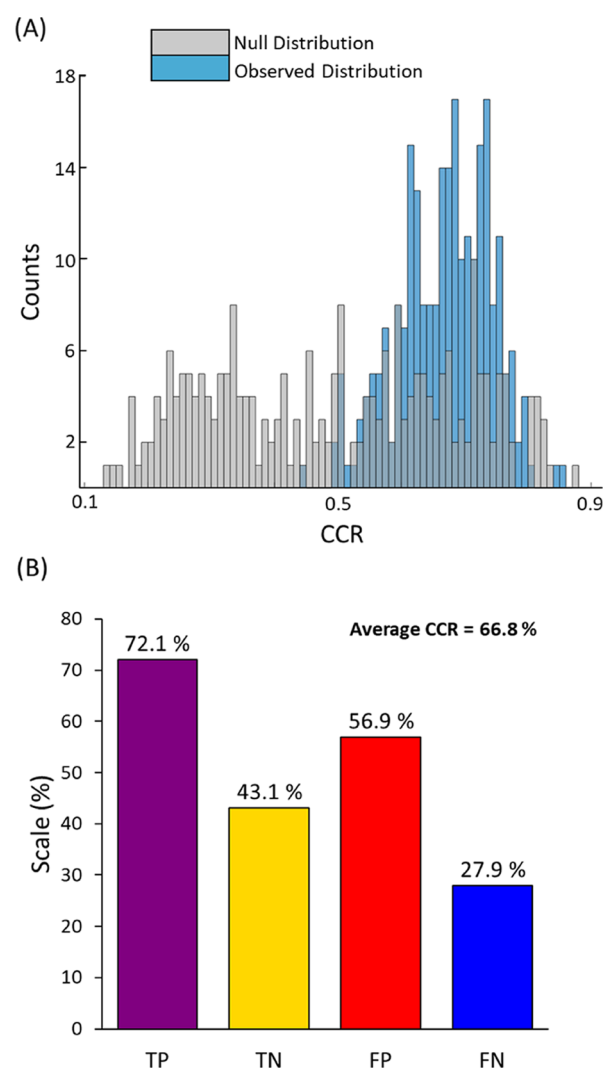
significant with $p < 0.05$ (reported in Table S3). ROC analysis of these compounds provides an improved classification model (Figure 3B) in which the addition of all eight features improves the AUC to 0.872 and the lowest CI range (0.801−0.942). Figure S7 displays the individual ROC curve for each added compound in this model. Due to limitations of volatile compound identification within this study, species are simply referred to as "VIP features" (1−8). Box and whiskers plots for these eight compounds are reported in Figure S8 alongside calculated fold changes (nonscaled). Each feature is downregulated within the PD cohort with fold changes ranging from 0.2 to 0.73 between PD and control cohorts.

Targeted Analysis of Previously Identified Analytes of Interest. The following stage of the study concerned investigating the presence and significance of a panel of analytes of interest obtained from the results of our previous study.[34] A targeted data analysis approach was performed for the four target compounds: octadecanal, eicosane, hippuric acid, and perillaldehyde. The data matrix revealed that multiple features were annotated with the same compound identification for each
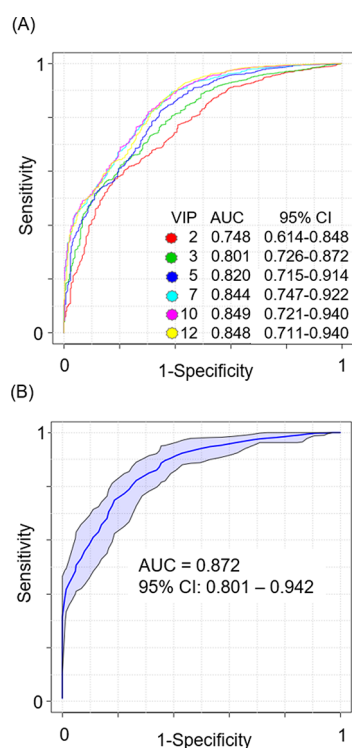
Figure 3. Multivariate ROC curve analyses to evaluate the performance of VIP compounds in biomarker models. Resampling was used to calculate 95% confidence intervals (CIs) using a Monte Carlo cross validation (MCCV) approach. (A) Each colored line represents the ROC curve using a specific number of variables; these are listed in the bottom right-hand corner from red (2-variable) to yellow (12-variable), generated from all compounds with VIP > 1. (B) ROC curve displaying the sensitivity and specificity of all eight-variables (VIP > 1, $p < 0.05$) combined; the shaded bands represent the 95% CI for this model.

of these targets based upon high similarities of their fragmentation pattern signatures. Database spectral matching typically returns a match factor (MF), in this case between 1 and 100, which indicates the similarity of an experimentally recorded spectrum to that of a database spectrum based on the presence of molecular and fragment ion peaks and their relative intensities. Eicosane and octadecanal each returned three identification matches, separately, in which these annotations were the primary ID, falling within a 2−3 min retention time (RT) window. Standards of eicosane and octadecanal were analyzed using an identical analytical method, and the corresponding reference spectra were matched against these highlighted features within the patient sample sebum data. Although retention times could not be matched, there was a high similarity between mass spectra (MF > 90). It can be inferred that these putative assignments based on spectra matching result from closely related molecular species that have large hydrocarbon chains, for example, lipids. Sebum is composed of a large number of lipid species such as triglycerides, fatty acids, and wax esters.[35,36] Thus, it is not surprising to observe multiple closely related lipid peaks.

To corroborate this inference, lipid standards were purchased and analyzed to investigate the resultant chromatographic spectra from known standards. Evidence from two parallel studies analyzing PD and control sebum, using liquid chromatography-MS[37] and paper spray-MS,[38] have shown that the primary compounds of interest can be putatively assigned to a series of lipids. Lipid mixtures were investigated to distinguish

the type of species detected in (nonderivatized) lipid DHS-TD-GC-MS experiments and to then evaluate their compound annotation in the same libraries. Each lipid standard mixture yielded a large number of resolvable compound features across each chromatogram that had high database spectral MF scores and good reproducibility across replicates. Between 68% and 87% of these features were annotated as hydrocarbon species assigned to at least two detected variables. A list of the hydrocarbon species detected within the lipids standards data is reported in Table 1, alongside the number of discrete features

Table 1. Putative ID and Chemical Formula of the Hydrocarbon Species Present within the Lipid Standards Data[a]

		number of features				
putative ID	chemical formula	PE	PC	PS	CL	GlcSph
decane	$C_{10}H_{22}$	2	0	2	2	0
dodecane	$C_{12}H_{26}$	9	4	6	7	6
tridecane	$C_{13}H_{28}$	1	0	0	1	0
tetradecane	$C_{14}H_{30}$	2	0	2	0	0
pentadecane	$C_{15}H_{32}$	2	0	2	2	2
hexadecane	$C_{16}H_{34}$	6	1	2	6	1
heptadecane	$C_{17}H_{36}$	2	2	4	3	1
nonadecane	$C_{19}H_{40}$	0	0	1	2	0
eicosane	$C_{20}H_{42}$	2	2	4	3	0
heneicosane	$C_{21}H_{44}$	0	0	0	0	1
docosane	$C_{22}H_{46}$	0	1	0	1	0
tricosane	$C_{23}H_{48}$	3	0	3	2	0
tetracosane	$C_{24}H_{50}$	0	0	0	0	2

[a]The number of distinct features that are annotated with each given compound ID is listed beside. The lipid abbreviations are as follows: PE, phosphatidylethanolamine; PC, phosphatidylcholine; PS, phosphatidylserine; CL, cardiolipin (18:1); and GlcSph, glucosylsphingosine.

provided with this annotation. Definitive annotation of large, chemically similar hydrocarbons, like lipids, is challenging because of the obvious similarities expected in their fragment products. This is further confounded by the large number of conceivable structural isomers for each hydrocarbon; for example, dodecane, the most frequently detected species in the data, has 355 possible isomers. These experiments were repeated using a traditional liquid injection GC-MS experimental setup for the analysis of some of these lipid standards (Table S4), and a similar trend in recurring hydrocarbon features was noted. We hypothesize that the overlap of compound annotations across distinct deconvolved features in our patient sample sebum data arises from the decomposition of lipid species.

For example, lipid oxidation is a well-known mechanism for lipid degradation and can proceed via an assortment of pathways dependent on the lipid and enzyme in question and the type of oxidation. These reactions produce a variety of product species including hydrocarbons, alongside aldehydes, ketones, alcohols, esters, and acids.[39,40] Therefore, the multiple instances of hydrocarbon species that are reproducibly detected throughout lipid-based chromatographic analyses (both patient sample and from standards) are due to lipid decomposition. Such degradation can lead to an accumulation of large concentrations of common hydrocarbon chains of different lengths. Similar products due to the loss of more distinctive structural moieties, such a lipid head groups, are readily observed upon collisional

activation and in source dissociation of these labile compounds. These results provide a vital insight into the expected profile upon analysis of the hydrocarbon-rich lipid-based biological matrix that is sebum. The identification of pure, large-molecular-weight lipid standards by GC-MS is challenging, and data shown here yield an array of product species across the chromatographic domain; unquestionably, the myriad of endogenous and exogenous compounds in sebum will offer a greater challenge. Prior to the generation of comprehensive annotated databases that reflect this chemical complexity, it is likely that biomarker detection will consist of Metabolomics Standards Initiative (MSI) level 2 feature detection rather than the verification of compounds.

The remaining two analytes of interest—hippuric acid and perillaldehyde—were only annotated as a derivatized form with trimethylsilyl and methoxime adducts, respectively. Again, multiple features were assigned these annotations, and they spanned a wide chromatographic time range; therefore, annotation of neither of these analytes of interest could be reliably verified within this data set. A direct limitation of analyzing a biological matrix without derivatization is the limited compound annotation using commercial databases, which are composed majorly of derivatized compounds. The dominating method in GC-MS analysis for metabolomics approaches uses chemically derivatized sample extracts, and therefore, biologically relevant compound spectra are often of derivatized forms. Further work is needed in the area of database curation for nonderivatized, biologically relevant, volatile species present in a sebum as a biological matrix, to enable confident compound identification using DHS-GC-TD-MS.

CONCLUSION

Volatile organic compounds from sebum measured using DHS-TD-GC-MS can accurately discriminate between PD and control samples. Our results validate previous findings that VOCs measured from skin have a differential profile in Parkinson's disease that can be modeled using PLS-DA analysis. Putatively annotated analytes of interest from our previous study could not be validated within this data set. However, multiple features were putatively annotated as two of these candidates (octadecanal and eicosane), and derivatized compound annotations were associated to the remaining two analytes of interest (hippuric acid and perillaldehyde). Spectral matching to an in-house database of standards, viz. octadecanal and eicosane, yielded high spectral similarity (MF > 90), although retention times did not match. This leads us to conclude that these compounds have very similar structures and/or could be common breakdown products of larger species in which the discriminatory structural moiety was lost as a neutral fragment. This hypothesis was strengthened with the analysis of lipid standard mixtures, which reveal the detection of common hydrocarbon species. Compounds selected using VIP score-based ranking from PLS-DA modeling performed well in multivariate ROC analysis, and each metabolite feature showed a fold change relating to a lower expression in PD samples. These compounds have not been annotated with putative identifications due to poor performance of database matching to available GC-MS mass spectra libraries. Future studies will create an in-house GC-MS database for both spectral and RT matching that will address the current bottleneck in the field for annotation of nonderivatized, biologically relevant VOCs.

ASSOCIATED CONTENT

Ⓢ Supporting Information

AUTHOR INFORMATION

Corresponding Author

Perdita Barran − *Manchester Institute of Biotechnology, School of Chemistry, The University of Manchester, Manchester M1 7DN, United Kingdom;* ⓘ orcid.org/0000-0002-7720-586X; Email: perdita.barran@manchester.ac.uk

Authors

Eleanor Sinclair − *Manchester Institute of Biotechnology, School of Chemistry, The University of Manchester, Manchester M1 7DN, United Kingdom*

Caitlin Walton-Doyle − *Manchester Institute of Biotechnology, School of Chemistry, The University of Manchester, Manchester M1 7DN, United Kingdom*

Depanjan Sarkar − *Manchester Institute of Biotechnology, School of Chemistry, The University of Manchester, Manchester M1 7DN, United Kingdom*

Katherine A. Hollywood − *Manchester Institute of Biotechnology, School of Chemistry and Manchester Centre for Synthetic Biology of Fine and Speciality Chemicals (SYNBIOCHEM), Manchester Institute of Biotechnology, School of Chemistry, The University of Manchester, Manchester M1 7DN, United Kingdom;* ⓘ orcid.org/0000-0002-7028-047X

Joy Milne − *Manchester Institute of Biotechnology, School of Chemistry, The University of Manchester, Manchester M1 7DN, United Kingdom*

Sze Hway Lim − *Department of Neurology, Salford Royal Foundation Trust, Manchester Academic Health Science Centre, The University of Manchester, Manchester M13 9PL, United Kingdom*

Tilo Kunath − *Institute for Stem Cell Research, School of Biological Sciences, The University of Edinburgh, Edinburgh EH16 4UU, United Kingdom*

Anouk M. Rijs − *Division of BioAnalytical Chemistry, AIMMS Amsterdam Institute of Molecular and Life Sciences, Vrije Universiteit Amsterdam, 1081 HZ Amsterdam, The Netherlands;* ⓘ orcid.org/0000-0002-7446-9907

Rob M. A. de Bie − *Department of Neurology, Amsterdam Neuroscience, Amsterdam University Medical Centers, University of Amsterdam, 1105 AZ Amsterdam, The Netherlands*

Monty Silverdale − *Department of Neurology, Salford Royal Foundation Trust, Manchester Academic Health Science Centre, The University of Manchester, Manchester M13 9PL, United Kingdom*

Drupad K. Trivedi − *Manchester Institute of Biotechnology, School of Chemistry, The University of Manchester, Manchester M1 7DN, United Kingdom*

Complete contact information is available at:

https://pubs.acs.org/10.1021/acscentsci.0c01028

Notes

The authors declare no competing financial interest.

REFERENCES

(1) Shirasu, M.; Touhara, K. The scent of disease: volatile organic compounds of the human body related to disease and disorder. *J. Biochem.* **2011**, *150*, 257−266.

(2) Williams, H.; Pembroke, A. Sniffer dogs in the Melanoma Clinic? *Lancet* **1989**, *333*, 734.

(3) Willis, C. M.; et al. Olfactory Detection of Human Bladder Cancer by Dogs: Proof of Principle Study. *BMJ.* **2004**, *329*, 712.

(4) Pirrone, F.; Albertini, M. Olfactory detection of cancer by trained sniffer dogs: A systematic review of the literature. *J. Vet. Behav. Clin. Appl. Res.* **2017**, *19*, 105−117.

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

(6) Morgan, J. Joy of super smeller: sebum clues for PD diagnostics. *Lancet Neurol.* **2016**, *15*, 138−139.

(7) Gallagher, M.; et al. Analyses of volatile organic compounds from human skin. *Br. J. Dermatol.* **2008**, *159*, 780−791.

(8) Borda, L. J.; Wikramanayake, T. C. Seborrheic Dermatitis and Dandruff: A Comprehensive Review. *J. Clin. Investig. Dermatology* **2015**, *3*, 1−22.

(9) Arsic Arsenijevic, V. S.; et al. A laboratory-based study on patients with Parkinson's disease and seborrheic dermatitis: the presence and density of Malassezia yeasts, their different species and enzymes production. *BMC Dermatol.* **2014**, *14*, 5.

(10) Krestin, D. The seborrhoeic facies as a manifestation of post-encephalitic parkinsonism and allied disorders. *QJM* **1927**, *21*, 177−186.

(11) Jadoon, S.; et al. Recent developments in sweat analysis and its applications. *Int. J. Anal. Chem.* **2015**, *2015*, 164974.

(12) Hussain, J. N.; Mantri, N.; Cohen, M. M. Working up a good sweat - The challenges of standardising sweat collection for metabolomics analysis. *Clin. Biochem. Rev.* **2017**, *38*, 13−34.

(13) Curran, A. M.; Rabin, S. I.; Prada, P. A.; Furton, K. G. Comparison of the volatile organic compounds present in human odor using SPME-GC/MS. *J. Chem. Ecol.* **2005**, *31*, 1607−1619.

(14) Thomas, A. N.; et al. Novel noninvasive identification of biomarkers by analytical profiling of chronic wounds using volatile organic compounds. *Wound Repair Regen.* **2010**, *18*, 391−400.

(15) Shetage, S. S.; et al. Effect of ethnicity, gender and age on the amount and composition of residual skin surface components derived from sebum, sweat and epidermal lipids. *Ski. Res. Technol.* **2014**, *20*, 97−107.

(16) Bernier, U. R.; Kline, D. L.; Barnard, D. R.; Schreck, C. E.; Yost, R. A. Analysis of human skin emanations by gas chromatography/mass spectrometry. 2. Identification of volatile compounds that are candidate attractants for the yellow fever mosquito (Aedes aegypti). *Anal. Chem.* **2000**, *72*, 747−756.

(17) Kutyshenko, V. P.; Molchanov, M.; Beskaravayny, P.; Uversky, V. N.; Timchenko, M. A. Analyzing and mapping sweat metabolomics by high-resolution NMR spectroscopy. *PLoS One* **2011**, *6*, e28824.

(18) Emaminejad, S.; et al. Autonomous sweat extraction and analysis applied to cystic fibrosis and glucose monitoring using a fully integrated wearable platform. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, 4625−4630.

(19) Riazanskaia, S.; Blackburn, G.; Harker, M.; Taylor, D.; Thomas, C. L. P. The analytical utility of thermally desorbed polydimethylsilicone membranes for in-vivo sampling of volatile organic compounds in and on human skin. *Analyst* **2008**, *133*, 1020−1027.

(20) Turner, A. P.; Magan, N. Electronic noses and disease diagnostics. *Nat. Rev. Microbiol.* **2004**, *2*, 161−166.

(21) Bach, J. P.; et al. Measuring compounds in exhaled air to detect Alzheimer's disease and Parkinson's disease. *PLoS One* **2015**, *10*, e0132227.

(22) D'amico, A.; et al. Identification of melanoma with a gas sensor array. *Ski. Res. Technol.* **2008**, *14*, 226−236.

(23) Rattray, N. J. W.; Hamrang, Z.; Trivedi, D. K.; Goodacre, R.; Fowler, S. J. Taking your breath away: Metabolomics breathes life in to personalized medicine. *Trends Biotechnol.* **2014**, *32*, 538−548.

(24) Beale, D. J.; et al. A review of analytical techniques and their application in disease diagnosis in breathomics and salivaomics research. *Int. J. Mol. Sci.* **2017**, *18*, 24.

(25) Wang, Y.; McCaffrey, J.; Norwood, D. L. Recent advances in headspace gas chromatography. *J. Liq. Chromatogr. Relat. Technol.* **2008**, *31*, 1823−1851.

(26) Sithersingh, M. J.; Snow, N. H. Headspace-Gas Chromatography. In *Gas Chromatography*; Poole, C., Ed.; Elsevier Inc., 2012; pp 221−233. DOI: 10.1016/b978-0-12-385540-4.00030-4.

(27) Chambers, M. C.; et al. A cross-platform toolkit for mass spectrometry and proteomics. *Nat. Biotechnol.* **2012**, *30*, 918−920.

(28) Domingo-Almenara, X.; et al. eRah: A Computational Tool Integrating Spectral Deconvolution and Alignment with Quantification and Identification of Metabolites in GC/MS-Based Metabolomics. *Anal. Chem.* **2016**, *88*, 9821−9829.

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

(30) Xu, Y.; Muhamadali, H.; Sayqal, A.; Dixon, N.; Goodacre, R. Partial least squares with structured output for modelling the metabolomics data obtained from complex experimental designs: A study into the Y-block coding. *Metabolites* **2016**, *6*, 38.

(31) Xu, Y.; Goodacre, R. On Splitting Training and Validation Set: A Comparative Study of Cross-Validation, Bootstrap and Systematic Sampling for Estimating the Generalization Performance of Supervised Learning. *J. Anal. Test.* **2018**, *2*, 249−262.

(32) Chong, J.; et al. Metabo Analyst 4.0: Towards more transparent and integrative metabolomics analysis. *Nucleic Acids Res.* **2018**, *46*, W486−W494.

(33) Xia, J.; Broadhurst, D. I.; Wilson, M.; Wishart, D. S. Translational biomarker discovery in clinical metabolomics: An introductory tutorial. *Metabolomics* **2013**, *9*, 280−299.

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

(35) Picardo, M.; et al. Sebaceous gland lipids. *Derm.-Endocrinol.* **2009**, *1*, 68−71.

(36) Pappas, A. Epidermal surface lipids. *Derm.-Endocrinol.* **2009**, *1*, 72−76.

(37) Sinclair, E. et al. Sebum: A Window into Dysregulation of Mitochondrial Metabolism in Parkinson's Disease. *ChemRxiv*, 2020, DOI: 10.26434/chemrxiv.11603613.

(38) Sarkar, D. et al. Paper Spray of Sebum: A Step Towards Rapid Confirmatory Diagnosis of Parkinson's Disease. Unpublished work, 2020.

(39) Ross, C. F.; Smith, D. M. Use of volatiles as indicators of lipid oxidation in muscle foods. *Compr. Rev. Food Sci. Food Saf.* **2006**, *5*, 18−25.

(40) Domínguez, R.; et al. A comprehensive review on lipid oxidation in meat and meat products. *Antioxidants* **2019**, *8*, 429.