

Paper Spray Ionisation Ion Mobility Mass Spectrometry of Sebum Classifies Biomarker Classes for the Diagnosis of Parkinson's Disease

Depanjan Sarkar¹, Eleanor Sinclair¹, Sze Hway Lim², Caitlin Walton-Doyle¹, Kaneez Jafri¹, Joy Milne¹, Johannes P.C. Vissers³, Keith Richardson³, Drupad K. Trivedi¹, Monty Silverdale², and Perdita Barran^{1*}

¹Manchester Institute of Biotechnology, School of Chemistry, the University of Manchester, Princess Street, Manchester, UK, M1 7DN.

²Department of Neurology, Salford Royal Foundation Trust, Manchester Academic Health Science Centre, University of Manchester, UK.

³Waters Corporation, Stamford Avenue | Altrincham Road, Wilmslow, SK9 4AX UK.

Abstract

Parkinson's disease (PD) is the second most common neurodegenerative disorder and identification of robust biomarkers to complement clinical diagnosis will accelerate treatment options. Here we demonstrate the use of direct infusion of sebum from skin swabs using paper spray ionisation coupled with ion mobility mass spectrometry (PS-IM-MS) to determine the regulation of molecular classes of lipids in sebum that are diagnostic of PD. A PS-IM-MS method for sebum samples that takes three minutes per swab was developed and optimised. The method was applied to skin swabs collected from 150 people and elucidates ~ 4200 features from each subject which were independently analysed. The data included high molecular weight lipids (>600 Da.) that differ significantly in the sebum of people with PD. Putative metabolite annotations of several lipid classes, predominantly triglycerides and larger acyl glycerides, were obtained using accurate mass, tandem mass spectrometry and collision cross section measurements.

Introduction

Neurodegenerative diseases are the leading source of disability globally.¹ According to the 2015 Global Burden of Disease, Injuries, and Risk Factors Study (GBD), Parkinson's disease (PD) is globally the fastest growing neurological disorder.¹ PD is also the second most common age-related neurodegenerative disorder with a prevalence of approximately 2% among people aged 65 and over, with motor symptoms including bradykinesia, tremor, rigidity, and postural instability as well as several non-motor symptoms.^{2,3} Whilst some of this increase tracks the demographics of greater life expectancies, this does not account for it entirely. The increased occurrence may also be due to better diagnosis and to environmental factors, particularly in mid social security disability insurance (SSDI) countries. Currently, it is predicted that these numbers are projected to increase globally to over 20 million by 2050.¹ The increase in PD globally, and its commensurate prevalence in younger people,⁴ compounds the need to identify biomarkers and methods to detect them; to provide a diagnostic pathway that may be applied prior to the onset of motor symptoms.

Increased oiliness and flaky skin, especially on the face and scalp, is a common symptom of PD, first noted by Krestin in 1927.⁵⁻⁷ The light yellow, oily substance present on all human skin, is known as sebum, and increased sebum production is a hallmark of PD. The sebaceous glands produce sebum in the skin, which help keep the skin and hair moisturized, and prevents sweat from evaporating, thus assisting with the body's temperature regulation. Sebum is an underexplored biofluid, which is readily obtained from non-invasive skin swabs, that primarily consists of a mixture of triglycerides, cholesterol, free fatty acids, waxy esters, and squalene.^{8,9} We have previously shown that sebum contains volatile biomarkers of PD,¹⁰ and that it can reveal mitochondrial dysregulation as PD progresses.¹¹ Here we set out to develop a method to analyse sebum in its native state to facilitate rapid assessment of PD status. Paper spray ionisation mass spectrometry (PS-MS) which allows the direct analysis of compounds from paper, has previously been demonstrated to detect of small molecules (50-800 Da.) from unprocessed biofluids such as blood, urine, and CSF,¹²⁻¹⁵ but not to date with sebum.

Combining ambient ionisation with ion mobility mass spectrometry has merit as a method to ‘clean up crude samples’ and provides reproducible ion selected drift time data in the place of a retention time for identification along with m/z .^{16–19} PS-MS has already been mooted as a diagnostic method with clear advantages for inexpensive sampling and rapid analysis. Here we couple paper spray ionisation with ion mobility mass spectrometry (PS-IM-MS) and demonstrate for the first time its application for biomarker discovery workflows which could lead to clinical methods. PS-IM-MS allows quick sample analysis with minimal sample processing compared to LC-MS, which here reveals far larger lipid moieties and provides enhanced separation for analytes with overlapping m/z ratios.

Results and Discussion

Developing an Optimised Workflow for Sebum Analysis

The procedure developed for sample collection from sites across the UK and subsequent PS-IM-MS analysis is shown in Figure 1, (see Supplementary information, Table 1 for further details on the sites).

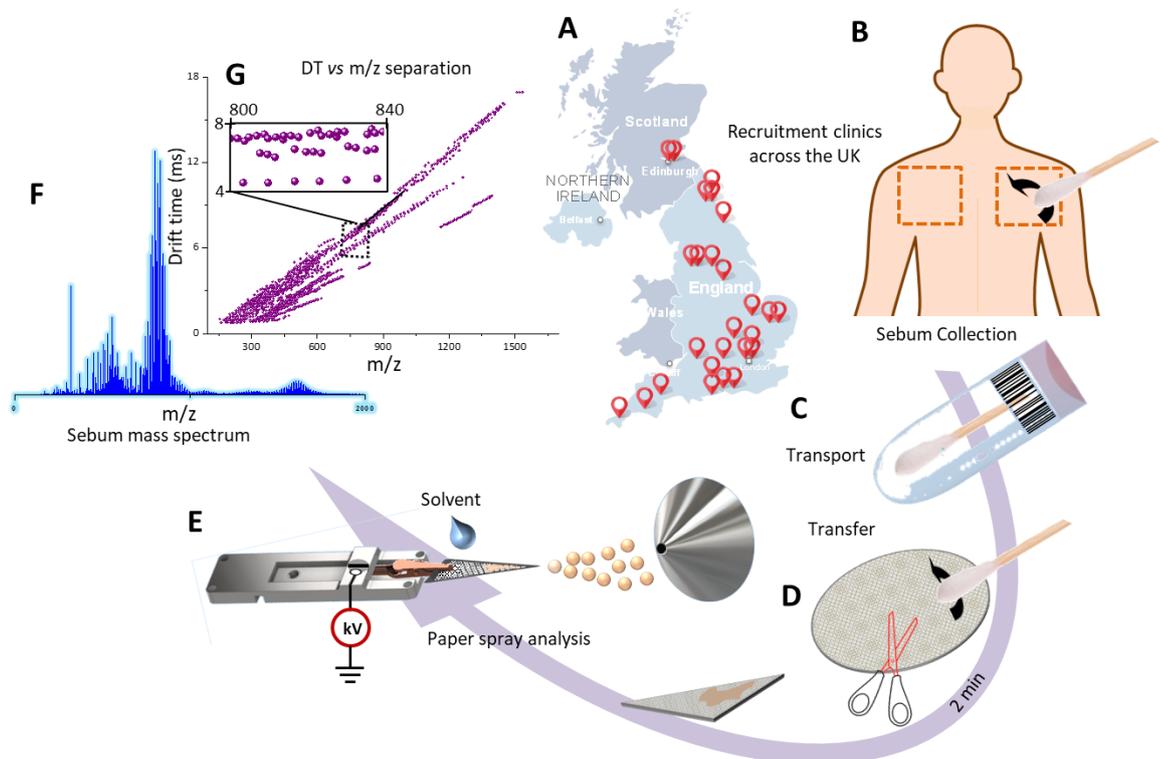


Figure 1. Workflow for paper spray analysis from clinic to raw data A) locations of collecting clinics in the UK, B) sebum collection from the mid-back of participants, C) medical Q-tips containing sebum samples transported under ambient conditions, D) schematic of the sample transfer from Q-tip onto the paper, E) the paper spray process, and F) representative mass spectrum collected from sebum with a distinctive multimodal distribution, and G) drift time vs. m/z distribution with zoom region exemplifying multiple drift time peaks associated with a specific m/z value.

Critical parameters of the PS-IM-MS experiments were optimised, which include type, size and shape of the paper, distance from the MS inlet, and eluting solvent composition (see methods for further details). Crucial to reproducibility was how the sebum is transferred from the sampled Q-tip swab to the analytical/filter paper. Two methods were tested, firstly, direct transfer from the Q-tip to the paper triangle in a ‘touch and roll’ approach. Alternatively, a rapid solvent extraction via vortex-mixing of the sampled Q-tip in ethanol (800 μ L) for 5 s, followed by spotting on to paper for analysis. Figure S1 shows mass spectra collected using these two approaches. Both are rich in features, and there is a notable increase in transmission of high molecular weight molecules (between m/z 1200-2000) in the touch and roll transfer mass spectrum (Figure S1A) versus the solvent extraction method (Figure S1B). Hence, this was chosen for all further sebum analyses (see methods). Different eluting solvent mixtures were tested and optimal for sebum found to be 4.5 μ L H₂O:EtOH (v:v, 4:1). Following these refinements, the mass spectra of human sebum consistently showed the presence of four distinct envelopes of predominantly singly charged species in the higher mass region (m/z 700-1800), with evidence of doubly charged species from isotopic distributions in highly congested feature-rich mass spectra (Figure 1F, and 2A).

Drift Time Separation Elucidates PD from Control

Ion mobility coupled to mass spectrometry is used here to resolve the high molecular weight species. Its application to resolve (and identify via collision cross section (CCS) matching) conformational isomers and isobaric structural isomers have been previously reported, albeit typically for lower molecular weight lipids.²⁰ Figure 1G is a drift time (DT) vs. m/z distribution of the ions generated and selected via Progenesis Q1 (Waters Corporation, UK), using default peak picking parameters, from a sebum sample of a PD patient. The

richness of information from the IM-MS datasets is illustrated in Figure 1G (inset), in which 41 m/z -DT (4-8 ms) resolved ions can be detected within a 40 Da window (m/z 800-840). A mixture of features with very similar and identical m/z values is present within this range. By analysing data from the full datasets from 150 patient and control samples we observe ~4200 m/z -DT features where 500 have a statistically significant relative abundance difference between PD and control samples (p -value < 0.05). Figure S2 displays the extracted arrival time distributions (ATDs) for ions measured at m/z 689.1 and 1394.8, showing the additional species resolved by IM. Enhanced separation between DT features for higher m/z values was achieved using a higher resolution SELECT SERIES cyclic IMS geometry (Waters Corporation) (data shown for m/z = 1394.8 (Figure S2B, C). A subset of the statistically significant molecules has drift time resolved features that are only observed in PD samples, and we, therefore focused on their elucidation.

Figures 2A-D show three-dimensional DT vs. m/z distributions from m/z 700-900 for PD (blue) and control (magenta) samples. Five dominant regions of m/z and drift time aligned species with respect to drift/IM separation can be observed: 2.5-3.5 ms (Cluster 1), 3.5-5 ms (Cluster 2), 6-7 ms (Cluster 3), 7.5-9 ms (Cluster 4), and 9-10 ms (Cluster 5). Within these regions, species can be detected that differ in relative abundance between PD and control samples. These rich data sets require substantial interpretation as demonstrated below. Based on the isotopic spacing between the m/z detected species, we assign Cluster 1 as triply charged, Clusters 2 and 3 as doubly charged and Clusters 4 and 5 as singly charged. These assignments are further supported when we convert DT to CCS values (Supplementary Table 2), since the earlier arriving Clusters (1-3), would result in CCS values that are too small for this particular m/z range if they would have been singly charged (see below). Within the m/z spectrum, there are repeating units in Cluster 3 that are separated by m/z 7.01, and within Clusters 4 and 5 the equivalent spacing is m/z 14.02, as revealed by inspection of drift time selected mass spectra (Figure S3A, B). These data taken together, supports assignment to a class of lipids that contains repeating units of CH_2 . Further to this, the mass spectrum shows that for every lipid there are at least four ions that differ by two mass units, which is due to different degrees of saturation within the hydrocarbon chains (Figure S4). Moreover, given that two singly charged Clusters (4 and 5) can be detected, afforded by the IM-MS shape

differentiation, suggests possible conformational differences for a number of the detected species. Tandem mass spectrometry experiments can aid to distinguish the isotopic and saturation features (Figure S5).

Significantly, Cluster 3 (hollow red arrow, Figure 2A and D) contains species that are abundant in samples from PD and not detected in controls (Figure 2A, 2C vs. 2B, 2D). Clusters 4 and 5 also show substantial differences in relative intensity between PD and control samples (discussed in detail later). These features, resolved by DT that distinguish between PD and control samples are observed throughout ions detected in the m/z range 700-900 (Figure S6). To validate the results from Progenesis QI analysis, we also used an independent peak detection algorithm (Apex3D) and Bayesian statistics for complementary quantitative analysis. Figures 2E and 2F show the probability of-up-regulation vs. log-ratio distribution for Clusters 5 and 3, respectively. The results are summarised in Supplementary Tables 3 and 4, including the output from a Bayesian analysis²¹ of the components detected within the separate clusters (log fold change, confidence interval, and probability of up-regulation), as well as the outcome (p -value) of two-tailed T-test analyses of the same peak detected PS-IM-MS data. Supplementary Table 3 reports the quantitative analysis of the components within Clusters 4 and 5, and Supplementary Table 4 of Cluster 3. This data shows that there is statistically significant upregulation of these DT features in samples from PD patients.

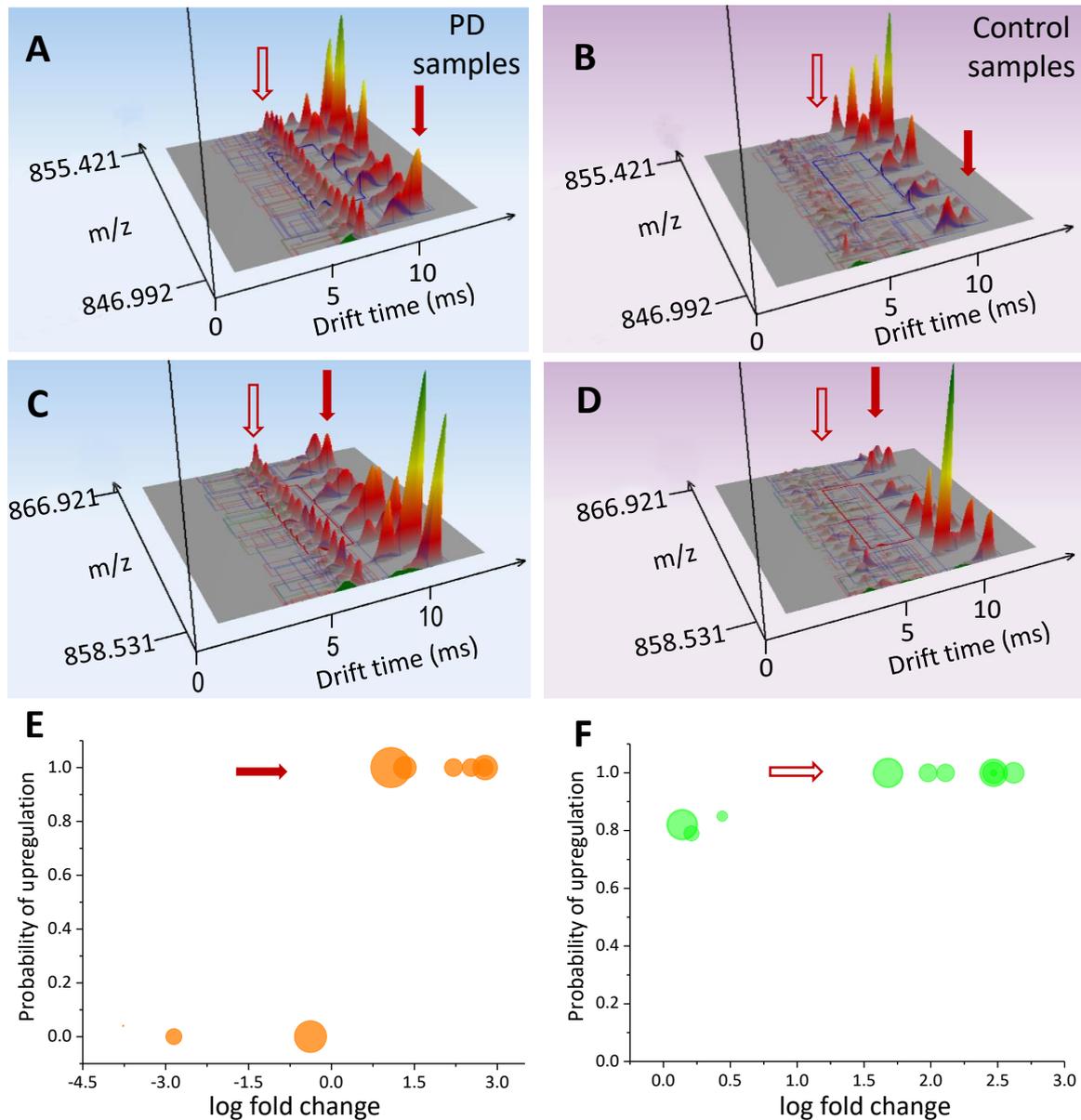


Figure 2. Three-dimensional DT vs. m/z and intensity distributions for PD (A and C) and control (B and D) samples from ~ 8 Da m/z windows exemplify the difference in the molecular composition of sebum produced by people with PD. 2.5-3.5 ms (Cluster 1), 3.5-5 ms (Cluster 2), 6-7 ms (Cluster 3), 7.5-9 ms (Cluster 4), and 9-10 ms (Cluster 5). A-D are normalised averages of samples from PD ($n=79$) and Control ($n=71$) participants from six recruiting sites. The red arrows indicate the Clusters (3 open and 5 closed) that alter substantially between PD and control. (Supplementary Tables 3 and 4). E and F show probability of up-regulation vs. log-ratio for Cluster 5 (orange) and Cluster 3 (green); The area of the circles is proportional to $1/\text{confidence interval}$.

Annotation of PD Lipid Biomarkers from IM-MS-MS datasets

To annotate these statistically important features in the DT vs. m/z spectra we employed accurate mass searching of available databases (HMDB and LipidMaps).^{22,23} These resources do not include experimental validation for lipids above mass 1400 Da, and provide only scant information for those above 700 Da. Focussing first on the singly charged ions in the range m/z 700-950 (Clusters 4 and 5) enabled tentative identification of multiple classes of lipids (Supplementary data set 1) with the highest confidence assignment as triacylglycerides, $C_nH_mO_p$, where $n=45-57$ and $m=84-104$ and $p=6$ over a mass range of 700 to 950 Da.

The region from m/z 500-650 is also rich in features, but fewer of these are statistically different in relative abundance between PD and control. We have found three regions (m/z 569.52, 597.55, 611.57), where there are statistical differences and where the DT vs. m/z distributions are also different by manual interpretation (Figure S7). Accurate mass searching of this region indicates that the lipid class most represented is diglycerides (DG), again with evidence of saturation in the hydrocarbon tails.

Further annotation of these data used tandem MS as well as information from the ion mobility separation, *i.e.* CCS, of the detected species. A commercially available lipid standards mixture was examined with PS-IM-MS using the same method. CCS values were obtained for the lipid species in this mixture with reference to the Major Mix calibrant (Waters Corporation, UK) and tandem mass spectra were recorded for the lipid standards. Both datasets were used to support identification of lipid classes in the sebum data. Figure 3A shows a CCS vs. m/z distribution for species measured from the standard mix (See Supplementary Figure S8 for full mass spectrum). CCS values for adducted ions are reported for best comparison with the sebum measurements (Supplementary Table 2). Figure 3B shows the CCS values obtained from the sebum lipids together with those from the standard mix.

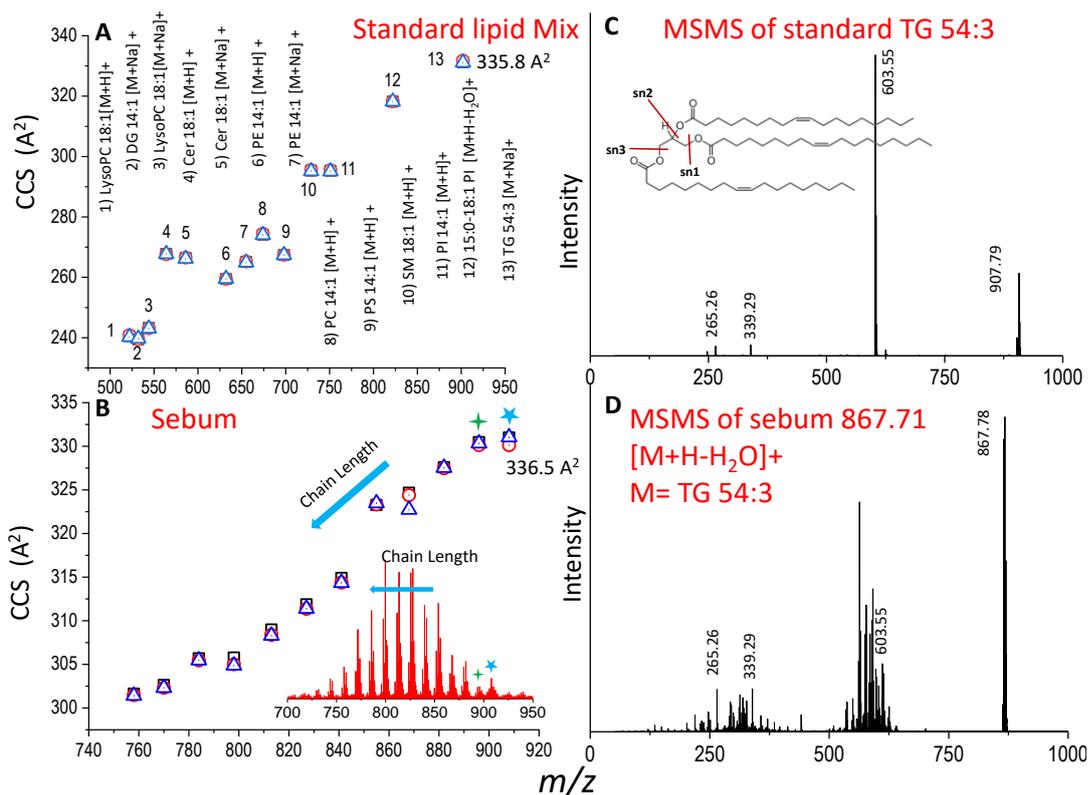


Figure 3. A) CCS vs. *m/z* distribution for a mixture of lipid standards 500-900 Da mass range. Data points numbered 1-13 show CCS values for different classes of lipids. Triplicate measurements were acquired under identical conditions ($\Delta \square \diamond$ symbols represent each measurement). B) CCS vs. *m/z* distribution for sebum. The inset shows the corresponding mass spectrum (*m/z* 700-950). C) MSMS spectrum of triolein (TG 54:3), which is one of the components in the standard lipid mixture. The predominant fragments are labelled in the figure, corresponding to *m/z* 603.55, 339.29, and 265.26. D) An MSMS spectrum of a feature at *m/z* 867.78 in sebum showing similar fragmentation compared to that observed for the TG 54:3 lipid standard (C). All data presented in this figure were collected using PS-IM-MS using higher resolution cyclic IM-MS travelling wave instrument (Waters Corporation, UK).

Two lipids from the standards mixture fall within the mass region of interest in the sebum data, namely those from the Triacylglycerol (TG) and Phosphatidylinositol (PI) lipid classes. The measured CCS values for the TG 54:3 [M+Na]⁺ (*m/z* 907.79) and PI 33:1 [M+H-H₂O]⁺ (*m/z* 822.62) lipid components present in the standard mixture were 335.8 Å² and 318.2 Å² respectively (numbers 12 and 13 in Figure 3A). One of the features that falls within the higher mass region of interest in the sebum data is a species at *m/z* 907.79 (marked with * in

Figure 3B). The mass (error: 4.8 ppm) and CCS value 336.5 \AA^2 (error: 0.2%) of this sebum feature closely match with TG 54:3. In addition to the $[M+Na]^+$ adduct measured in the standards data, TG 54:3 is also known to generate a precursor ion at m/z 867.7800 which corresponds to $[M-H_2O+H]^+$.²⁴ This species was also detected in the sebum data (m/z 867.7819, mass error 2.2 ppm).

McLean *et al.* have previously reported CCS values for major lipid classes, and demonstrate that the correlation of CCS with mass (as the chain lengths increase) differs due to head group chemistry and within each class on saturation of the hydrocarbon tails.²⁰ A linear correlation in CCS values was observed in the sebum data for the series of peaks at 14 Da intervals (inset of Figure 3B), with a small decrease in CCS value going down the series. For example, the peak at m/z 895.77 (marked with + in Figure 3B) has a CCS value of 335.5 \AA^2 . Hence, there is evidence that the ion mobility data presented in Figure 3B correlates to a mixture of TG lipids with varying fatty acid chain lengths. It is also known from literature that the most significant components of human sebum are TG and fatty acids, comprising approximately 57.5%.^{9,25} Additionally, accurate mass and CCS matching were performed for the DG class of lipid features. For example, the accurate mass search of m/z 527.47 (one the features seen in m/z 500-650 region), suggests that it to be either DG(12:0/17:0/0:0) or DG(13:0/16:0/0:0) (mass error 9.6 ppm in both cases). The measured CCS value for this species (249.2 \AA^2) is most comparable to the CCS value (243.3 \AA^2) of DG (14:1/14:1/0:0), present in the standard mixture (2.4% difference). Figure S9 shows the CCS vs. m/z distribution in the region of m/z 500-650. In this case a linear correlation was also observed for increases in the chain length of the attached hydrocarbons.

To increase confidence in our assignment and complement accurate mass and CCS measurements, tandem MS was utilised. Figure 3C shows the MSMS spectrum of m/z 907.79 (TG 54:3 $[M+Na]^+$) measured from the mixture of standard lipids. Three predominant fragment peaks were measured at m/z 603.55, 339.29, and 265.26 corresponding to fragmentation in sn1, sn2 and sn3 positions. The sodium adduct of the TG 54:3 species is not a dominant feature in the sebum spectrum. Therefore the more abundant $[M-H_2O+H]^+$ species was selected for fragmentation. The MSMS spectrum shows a high similarity in fragmentation to that obtained from the TG 54:3 standard (Figure 3D). This behaviour was

also observed following tandem MS of other sebum ions (Figure S10) which yielded the same fragment ion series as the TG standard. This indicates that these lipids have two fatty acid chains in common and that variation may be solely in the third fatty acid chain.

The combination of three independent parameters, accurate mass, fragmentation data and CCS value of the m/z 907.78 feature compared to the TG 54:3 standard enables lipid annotation at MSI level 1 (labelled with * in Figure 3B).²⁶ This annotation to the TG class can be applied to every other significant singly charged ion in the series between 700-950 m/z which supports their assignment as TG lipids with different fatty acid chain lengths and saturation. Linear correlation with incremental changes in CCS values also supports this conclusion. The IM-MS data indicates the presence of various extents of unsaturation in the fatty acid chains (discussed in detail below). As a negative control, we demonstrate the elimination of an alternative lipid class annotation of this series of peaks present/measured in sebum in the m/z region 700-950 using MSMS data. A species at m/z 822.61 was detected in sebum which matches closely in mass to PI 33:1, a component of the standards mixture (m/z 822.62, 11 ppm). MSMS spectra of the standard PI 33:1 (Figure S11) shows a loss of 259 Da which is characteristic for this class of lipid via loss of the lipid head group. Fragmentation data from sebum for this species does not contain this loss and therefore lacks support for a PI lipid identification.

We have also observed the presence of diverse unsaturation in sebum lipids. Figure S12 shows a CCS value vs. m/z distribution for the lipids present in the mass range 700-900 Da. The variation in CCS values of features very close in mass (circled with a dotted line) reflects changes in the number of double bonds present in the fatty acid chains as previously reported.²⁰

Putative assignment of higher mass features

One of the remarkable aspects of the PS-IM-MS data is the high molecular weight species, both cluster 3 and the singly charged ions that occur in distributions that centre on m/z 1080, 1334 and 1600. The relative intensity of these features does not change following dilution, ruling out aggregates formed during the ionization process. Focusing first on the significant doubly charged peaks that constitute Cluster 3 (Figure 2F, mass spectrum shown

in Figure S3B), it is evident from the accurate masses and ATD data that these are not doubly charged dimeric clusters of the TG class of lipids that constitute Clusters 4 and 5. To verify this, the Cluster 3 m/z values were considered as solutions to the generalised form $[M+A+B]^{2+}$, where A and B were variable adducts including H, Na, K, etc. A corresponding singly charged candidate peak list was generated of the form $[M+A-H]^+$, $[M+A-Na]^+$, $[M+A-K]^+$, etc. which was queried against the peak detected data, but did not match to singly charged ions in the mass spectrum. Searching COMP_DB of LipidMaps for these exact masses provides the best match identifications in the TG class, although these would require each of the three hydrocarbon tails to possess more than 30 carbon atoms on average.

Now considering the high mass singly charged ions, the envelopes at m/z 1080, 1334 and 1600 are separated by 256 Da which can be evaluated as $C_{16}H_{32}O_2$ (Palmitic acid) which is also the separation between the m/z 850 region (TG) and the m/z 600 region (DG). This suggests that the sebum spectrum also contains tetra, penta and hexa-acyl glycerides connected by ester linking additional $C_{12-18}H_{24-32}O_2$ moieties. Such compounds have previously been identified in plants²⁷ but not to date in humans.

Further experiments were used to elucidate these singly charged features observed in the m/z 1100-1700 region, including tandem MS post-IM. In such experiments the ATD of the precursor ion is identical to that of its fragment. When performed on the entire spectrum, the relationship between the envelopes of features can be seen in breakdown curves (Figure S13) which predominantly breakdown via S_N1 and S_N2 and S_N3 cleavages into the lower order acylglycerides. It is also possible to mass isolate a small group of m/z ions in the quadrupole prior to IM separation, followed by tandem MS (Figure S14). By doing this, we observe that the large ions are comprised of isomers of the general form $n(C_{12-18}H_{24-32}O_2)$ with different chain lengths and of course saturations.

The envelope of peaks in sebum differing by 14 Da has been assigned as a recurrent lipid series each changing by a single $-CH_2$ unit. This would implicate the presence of lipids with an odd number fatty acid carbon chains, which are uncommon in healthy serum but have been previously reported to be present in sebum,²⁵ as well as indicators of a disease state.²⁸ The origin of odd chained fatty acid (OCFA) such as C15:0 and C17:0 have been attributed to diet.²⁹ It was recently suggested that OCFA ends up in propionyl CoA via mitochondrial β -

oxidation of dietary OCFA. In contrast, even chains end up in acetyl CoA, having more physiological significance.³⁰

In bloom-forming alga *Emiliana huxleyi*, unsupervised spatially-aware clustering indicated a systemic metabolic shift towards lipids containing OCFA induced during viral infection.³¹ Similarly, a change in lipid metabolism is associated with disease phenotypes observed in a myriad of metabolomics studies from both plants and humans. Jenkins *et al.* suggested an endogenous pathway in human plasma where the C15:0 to C17:0 is seen as 1:2 instead of the expected 2:1 as seen in dairy fat.³⁰ Studies have shown that the presence of OCFA in serum is associated with reduced risks of metabolic diseases,^{32,33} and recently that OFCA are associated with cancer cell growth³⁴. Venn-Watson *et al.* showed that C15:0 may not be produced as readily endogenously but is derived from dairy, fish and plant-rich diets.²⁹ Lower intake and endogenous C15:0 was associated with higher mortality and poorer physiological state. The study showed that C15:0 might be one of the essential fatty acids that have a metabolic role in repairing mitochondrial function to alleviate inflammation, anaemia, dyslipidaemia and fibrosis *in vivo*.

A recent meta-analysis of serum lipids and their association with PD pathogenesis revealed lower amounts of triglycerides in Parkinson's and others have suggested the role of perturbed fatty acids in PD.^{35,36} Certain glycerophospholipids within plasma interact with the residues 1-102 of alpha-synuclein, enhancing aggregation. Increased plasma PA (18:2/15:0) levels are shown to be potential markers of PD.³⁷ Many studies have highlighted the role of short-chain fatty acids in the manifestation of PD,^{36,38} but very little is known about the role of OCFA in Parkinson's disease (PD). We have previously shown the dysregulation of lipids and mitochondrial function as critical perturbations observed in sebum from PD participants.¹¹ The new results presented here indicate that sebum, which acts to sink excessive metabolites from the lymphatic system, may well be the key to understanding the changes in regulation of these essential fatty acids as PD progresses.

Conclusion:

In conclusion, a new and accessible method to analyse sebum samples has been developed and shown to be able to readily distinguish between samples taken from people

with PD versus matched controls. PS-IM-MS of each sebum sample is performed in ~ 2-3 min which is noticeably faster than current clinical mass spectrometry approaches. Previous studies have demonstrated the use of PS-MS to detect metabolites present in the blood, urine and other biofluids,¹³⁻¹⁵ it has not previously been applied to sebum. These results, coupled with the lower signal from high mass species following solvent extraction, also demonstrate the benefit of an ambient ionisation approach direct from the native biofluid. Liquid chromatography-mass spectrometry (LC-MS) was not able to detect such large molecular weight species (> 1200 Da) from sebum extracts.¹¹

Mass spectra of sebum samples acquired using PS-MS demonstrated the technique's utility measuring both low and high molecular weight species (m/z 50-2000) that may be lost in sample preparation, namely solvent extraction and or stick to LC columns, in more traditional analytical methods. PS-MS combined with IM separation reveals specific compounds unique to PD sebum samples when compared to healthy controls. Furthermore, we have identified two classes of lipids namely triacylglyceride and diglyceride, as components of human sebum that are significantly differentially expressed in PD. Non-invasive sampling, followed by PS-IM-MS analysis targeting these compounds, could provide an inexpensive assay to support clinical phenotyping for the confirmatory diagnosis of Parkinson's Disease.

Acknowledgements:

We thank The Michael J Fox Foundation (grant ref: 12921) and Parkinson's UK (grant ref: K-1504) for funding this study. DS thanks the Royal Society and SERB for the Newton International fellowship. The Engineering and Physical Sciences and Biological and Biotechnological Research Councils UK are also acknowledged for funding equipment used in this study (EP/T019328/1 and BB/L015048/1). We are grateful to Tilo Kunath and Richard Weller (University of Edinburgh) for many helpful discussions on sebum and skin. We also thank our recruitment centres for their enthusiasm and rigour during the recruitment process. We are very grateful to all the participants who took part in this study as well as PIs and nurses across all the recruiting centres.

Materials and Methods

For paper spray measurements, grade 1 and 42 filter papers were used (Whatman International Ltd., UK). The paper was cut into isosceles triangles 5 mm (base) x 10 mm (height). LC-MS grade solvents were used for the study. This included acetonitrile (98% purity), methanol (99% purity) and deionised water (Fisher Scientific, UK) and ethanol (99% purity) (VWR Chemicals, UK). Solvents were used without any further purification. An ESI-L low concentration tuning mix (TM) (Agilent Technologies, UK), L-glutamine, and L-proline (Sigma-Aldrich, UK) were used as standards for optimisation of the process. For tandem mass spectrometric experiments, a range of commercially available standard mixtures of different lipid classes (Differential Ion Mobility System Suitability Synthetic Standard Mixture, and LightSPLASH™ LIPIDOMIX® Quantitative Mass Spec Primary Standard (Avanti Polar Lipids, Inc., USA) were purchased. For direct infusion mass spectrometry, the salt content is an important factor as presence of excess salt in samples is not favourable for better ionisation. Hence, desalination (using Ziptips C18) and dilution of the standards were performed prior to PSI IM MS measurement using a travelling wave instrument (Synapt G2-Si and SELECT SERIES cyclic IMS, Waters Corporation, UK). The instruments were mass and CCS calibrated prior to the sample measurements. Tandem mass (MSMS) spectra were recorded for these lipids using PS-MS. A source was designed in-house (using Autodesk Inventor 2018) and 3D-printed (Ultimaker 3 Extended, GoPrint3D, Ripon, UK) for paper spray analysis on a the Synapt G2-Si instrument. Copper micro-alligator clips (Premier Farnell UK Ltd., UK) were used to hold the paper triangles at a high potential, followed by positioning them close to the MS inlet. Medical Q-tips swabs (Fisher Scientific, UK and Copan Diagnostics, USA) were used for sample collection.

Study Participants

For the initial method development of PS-MS using sebum, samples from healthy controls were used. The developed method was further tested using samples from participants with PD. The participants for this study were part of a recruitment process at 27 NHS clinics all over the UK (Supplementary Table 1). A subset comprising, 34 people with PD and 30 matched control subjects, was measured as a training set and a further 86 samples from 5 different collection sites were used as a validation set. An overview of important patient demographics is summarised in Supplementary Table 5. The results of significance

tests between cohort group metadata are reported in Supplementary Table 6. Ethical approval for this project (IRAS project ID 191917) was obtained from the NHS Health Research Authority (REC reference: 15/SW/0354).

Sample Collection

Sebum samples were non-invasively swabbed from the mid-back of participants with medical Q-tip swabs. Each swab, secured in its holder, was transported under ambient conditions in sealed envelopes to the central facility at the University of Manchester, where they were stored at -80 °C until analysis.

Paper spray ionisation mass spectrometry (PS-MS)

The size and shape of the paper triangles was found to be critical to achieving reproducible data, indicating that the biomass of the sampling surface and the applied sebum can be controlled. Paper triangles were cut manually, and each one was quality controlled for size prior to use. A camera attached to the source housing ensured reproducible positioning of the paper tip for each measurement. An optimised position was marked, and the paper tip placed at that mark for each repeat. Whatman grade 1 and 42 filter papers were tested for PSI-MS experiments using standards (TM, L-proline, and L-glutamine). Following optimisation, the total ion chromatogram (TIC) and the mass spectra acquired using each filter paper were visually similar (Figure S15), although reproducibility was higher with the Whatman grade 42 paper. Relative standard deviation (RSD) in repeated measurements of the standards was 9.6 using Whatman grade 42 paper, and 15.06 using Whatman grade 1 paper (Figure S16); Whatman grade 42 paper was used for all further analysis. Reproducibility of PS-MS for sebum was also tested over multiple runs throughout a day under identical instrumental conditions. Figure S17 represents mass spectra recorded for sebum at 8h intervals. No significant shift in the masses was observed.

Sebum was transferred from the Q-tip swabs onto the paper substrates by gentle touch and roll of the swab onto the sampling area. After sample transfer, the paper triangle was clipped onto the copper alligator clip using tweezers avoiding contamination. Each copper clip was cleaned by ultrasonication in acetone before use. For each sample, a new clip and tweezer were employed to prevent cross-contamination. The clip was connected to a

custom paper spray ion source built in-house, adapted to the Synapt G2 Si HDMS ion mobility mass spectrometer. PS-MS measurements were commenced by positioning the paper tip in front of the MS inlet using a movable xyz nESI stage and subsequently applying a voltage (2.5-3 kV) to the alligator clip by adapting the ESI capillary voltage supply. Upon elution with a polar solvent at that elevated potential, a spray plume of tiny charged-droplets was observable at the tip of the paper simultaneously with the appearance of an ion signal.

All mass spectra were recorded over the range of m/z 50-2000. The critical instrument parameters for each PS-IM-MS experiment were: capillary voltage at 2.5 kV, source temperature at 80 °C, sampling cone at 30 V and source offset of 40 V, IMS wave velocity 650 m/s, IMS wave height 40 V. No desolvation or cone gas was used. Mass spectra were recorded for two minutes at a scan rate of 2 sec per scan. A total of 60 scans were used for further data analysis.

Sebum was measured on the same day as the standards mixture, under the same experimental conditions and using the same CCS calibration to target statistically significant features for CCS values calculation.

Use of internal standards

To check the reproducibility of paper spray across different samples, TM was used. For these experiments, 3.5 μ L of the TM solution was spotted on paper triangles and air-dried. Dried paper triangles were used for PSI-MS measurements of sebum samples following an identical method described in the previous paragraph.

Limit of detection

The amount of biomass present on the Q-tip depends on the following facts. Firstly, the production of sebum between people can be variable. Therefore, if a sample is collected from a participant with low sebum production, the biomass on the Q-tip will be low. Secondly, sample collection efficiency also varies from person to person. A limit of detection experiment was performed to test our method for various amounts of biomass present on Q-tip. In this experiment, an unknown quantity of sebum was extracted from a Q-tip into a known amount of solvent (using the solvent extraction method discussed earlier). This extracted stock was

then repeatedly diluted by 50% until the features of sebum were not visible in the mass spectrum. A known internal standard (a standard lipid) was added for our reference. A relative concentration of sebum was calculated from the absolute intensities in the mass spectrum. In this experiment we could detect as low as 500 pM of sebum. Below this concentration sebum features were not distinguishable from the noise in the mass spectrum.

Data processing with Progenesis QI

After recording the IM-MS data from all the participant samples under identical conditions, the raw data were peak detected and deconvolved using Progenesis QI (Nonlinear Dynamics, UK). Peak picking initially identifies accurate mass m/z values and coincident ion mobility drift times (DT). These correlated m/z -DT features are then aligned, and area normalisation is carried out with reference to the best candidate sample, within the entire data set, chosen by a default set of parameters. Peak picking limits were set to automatic with default noise levels, to balance the signal-to-noise ratio according to the data quality. Signal acquired before 0.1 min of infusion and after 1.4 min of infusion were discarded during processing to only retain reproducible signals. Features were classified as significant if their p -value was less than 0.05, determined by one-way analysis of variance (ANOVA). For annotation, accurate mass features were matched with both the Human Metabolome Database (HMDB) and LipidMaps.^{21,22} For initial quantitative analysis to determine regions of interest, Progenesis QI was applied, which nominates one of the PS-MS raw data files a reference to which the other runs are normalized by determining so-called individual, run-specific scalar factors. Typically, the data file with the largest number of features is chosen as the reference.

Validation of Data Analysis Using Apex3D and Probability-based Quantitation

The use of fold-changes and ANOVA testing, allows for the prompt identification of m/z and drift time resolved regions of interest that afford differentiation between sample groups. The information residing in the regions of interest was further analysed using an established probabilistic (Bayesian) method²¹ which was specifically designed for data of the nature typically associated with direct infusion MS experiments. This method can accommodate uncertainty in feature assignments and can quantify the uncertainty in the

results. At this stage of the analysis of the data, a molecular feature is a mass, drift time pair. To do this IM-MS data were peak detected using Apex3D, embedded in Progenesis Q1 for metabolomics (Nonlinear Dynamics, UK). The processed data were single-point m/z corrected using the singly charged monoisotopic mass of Hexakis(2,2,3,3-tetrafluoropropoxy)phosphazene from Tune Mix (Agilent Technologies, CA) and total ion current (TIC) normalised. Quantitation of the m/z and drift separated components of interest, with analyte CCS cluster indices and ranges listed in Supplementary Table 3 and 4, was conducted using a previously described probability-based framework.²¹ Measurement noise was assessed by nominating drift separated low m/z PS-MS matrix ions (103.05 Th/1.0 ms, 105.07 Th/1.0 ms, and 109.10 Th/1.2 ms) as pseudo internal standards.

CCS calibration

The data were calibrated using a recently described approach, which gives improved performance for multiply-charged analytes in TWIM devices.^{39,40} The reference peaks used to create the calibrations were selected from Tune Mix^{41,42} (Synapt G2-Si data), singly and doubly charged polyalanine⁴³ and MajorMix⁴⁴ (SELECT SERIES cyclic IMS data). CCS reference peaks were identified by accurate mass in peak lists generated using the Apex3D algorithm and the calibrations were created and applied using the IMSCal software.⁴⁰ For both datasets, the velocity relaxation parameter “ a ” was fixed at 1.0 and for the Synapt G2-Si instrument data the radial/RF parameter “ c ” was set at 0.2. Figure S18 shows the uncertainty distributions associated with the CCS calibration for the lipid clusters of interest.

Validation of PS-IM-MS data and extension to other sites

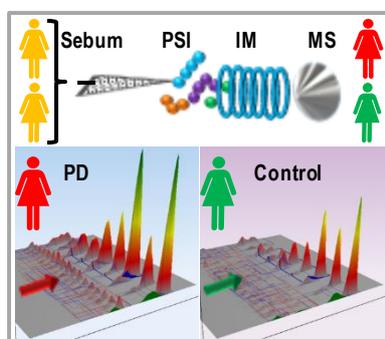
To determine the significance of this approach we repeated the analysis with ten PD samples from five additional collection sites, to investigate the influence of location or the person who collected sebum on the data. Principal component analysis (PCA) (Figure S19), support vector machine (SVM), and random forest (RF) modelling classified samples by collection sites. No apparent separation was possible using PCA, but this is often the case for complex data, since PCA is an unsupervised dimension reduction method. To assess supervised learning, the samples were split 100 times into a training set (75%) and a test set

(25%). For each split, the model under consideration was trained on the training set and then tested using the test set. The prediction output from each test was output as a confusion matrix. Finally, a confusion matrix representing the average of 100 tests was reported (Supplementary Table 7 and 8 for SVM and RF models, respectively). This analysis suggests that this data cannot be classified by collection site, further, since samples from different sites and patients were acquired on different days, we surmise that PSI-IM-MS can be applied to detect differences in the molecular composition of sebum that can diagnose PD without influence from the sampling environment nor batching effects.

Data Availability

The data sets generated during and/or analysed during the current study are available in a supplementary date set which include the data required to generate ROC curves from our analysis. Additional raw data is stored on our data server in accordance with funder requirements and will be available for the next ten years from the corresponding author on request.

TOC



References:

1. Fereshtehnejad, S.-M. *et al.* Burden of neurodegenerative diseases in the Eastern Mediterranean Region, 1990-2016: findings from the Global Burden of Disease Study 2016. *Eur. J. Neurol.* **26**, 1252–1265 (2019).
2. Hassan, Q., Li, S., Ferrag, C. & Kerman, K. Electrochemical biosensors for the detection and study of α -synuclein related to Parkinson's disease - A review. *Anal. Chim. Acta* **1089**, 32–39 (2019).
3. Moustafa, A. A. *et al.* Motor symptoms in Parkinson's disease: A unified framework. *Neurosci. Biobehav. Rev.* **68**, 727–740 (2016).

4. Parkinson's Disease Foundation. Available at: <https://www.parkinson.org/Understanding-Parkinsons/Statistics>.
5. Krestin, D. The Seborrhoeic Facies as a Manifestation of Post-Encephalitic Parkinsonism and Allied Disorders. *QJM An Int. J. Med.* **os-21**, 177–186 (1927).
6. 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).
7. Arsenijevic, V. S. A. *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.* **14**, 5/1-5/9, 9 pp. (2014).
8. Stewart, M. E. & Downing, D. T. Chemistry and function of mammalian sebaceous lipids. *Adv. Lipid Res.* **24**, 263–301 (1991).
9. Lovaszi, M., Szegedi, A., Torocsik, D., Szegedi, A. & Zouboulis, C. C. Sebaceous-immunobiology is orchestrated by sebum lipids. *Dermatoendocrinol.* **9**, e1375636 (2017).
10. Trivedi, D. K. *et al.* Discovery of Volatile Biomarkers of Parkinson's Disease from Sebum. *ACS Cent. Sci.* **5**, (2019).
11. Sinclair, E. *et al.* Metabolomics of sebum reveals lipid dysregulation in Parkinson's disease. *Nat. Commun.* **12**, 1592 (2021).
12. Wleklinski, M. *et al.* Zero Volt Paper Spray Ionization and Its Mechanism. *Anal. Chem.* **87**, (2015).
13. Damon, D. E. *et al.* Direct Biofluid Analysis Using Hydrophobic Paper Spray Mass Spectrometry. *Anal. Chem. (Washington, DC, United States)* **88**, 1878–1884 (2016).
14. Espy, R. D. *et al.* Paper Spray and Extraction Spray Mass Spectrometry for the Direct and Simultaneous Quantification of Eight Drugs of Abuse in Whole Blood. *Anal. Chem. (Washington, DC, United States)* **86**, 7712–7718 (2014).
15. Michely, J. A., Meyer, M. R. & Maurer, H. H. Paper Spray Ionization Coupled to High Resolution Tandem Mass Spectrometry for Comprehensive Urine Drug Testing in Comparison to Liquid Chromatography-Coupled Techniques after Urine Precipitation or Dried Urine Spot Workup. *Anal. Chem. (Washington, DC, United States)* **89**, 11779–11786 (2017).
16. Yan, C. *et al.* Real-Time Screening of Biocatalysts in Live Bacterial Colonies. *J. Am. Chem. Soc.* **139**, 1408–1411 (2017).
17. Myung, S. *et al.* Coupling Desorption Electrospray Ionization with Ion Mobility/Mass Spectrometry for Analysis of Protein Structure: Evidence for Desorption of Folded and Denatured States. *J. Phys. Chem. B* **110**, 5045–5051 (2006).
18. Jackson, A. T. *et al.* Microstructural and conformational studies of polyether copolymers. *Int. J. Mass Spectrom.* **238**, 287–297 (2004).
19. Manicke, N. E. & Belford, M. Separation of Opiate Isomers Using Electrospray Ionization and Paper Spray Coupled to High-Field Asymmetric Waveform Ion Mobility Spectrometry. *J. Am. Soc. Mass Spectrom.* **26**, 701–705 (2015).
20. Leaptrot, K. L., May, J. C., Dodds, J. N. & McLean, J. A. Ion mobility conformational lipid atlas for high confidence lipidomics. *Nat. Commun.* **10**, 1–9 (2019).
21. Richardson, K. *et al.* A Probabilistic Framework for Peptide and Protein Quantification from Data-Dependent and Data-Independent LC-MS Proteomics Experiments. *OMICS*

- 16**, 468–482 (2012).
22. Wishart, D. S. *et al.* HMDB 4.0: the human metabolome database for 2018. *Nucleic Acids Res.* **46**, D608–D617 (2018).
 23. Fahy, E. *et al.* Update of the LIPID MAPS comprehensive classification system for lipids. *J. Lipid Res.* S9–S14 (2009) doi:10.1194/jlr.R800095-JLR200.
 24. Lipid Maps, <www.lipidmaps.org>.
 25. Nicolaides, N. Skin lipids. Their biochemical uniqueness. *Sci. (Washington, DC, United States)* **186**, 19–26 (1974).
 26. Blaženović, I. *et al.* Structure Annotation of All Mass Spectra in Untargeted Metabolomics. *Anal. Chem.* **91**, 2155–2162 (2019).
 27. Matsuzaki, T., Koiwai, A. & Kawashima, N. Isolation of tetra-, penta-, hexa- and heptaacyl glycerides from stigmas of *Nicotiana tabacum*. *Agric. Biol. Chem.* **47**, 77–82 (1983).
 28. Young, R. S. E. *et al.* Apocryphal FADS2 activity promotes fatty acid diversification in cancer. *Cell Rep.* **34**, 108738 (2021).
 29. Venn-Watson, S., Lumpkin, R. & Dennis, E. A. Efficacy of dietary odd-chain saturated fatty acid pentadecanoic acid parallels broad associated health benefits in humans: could it be essential. *Sci. Rep.* **10**, 8161 (2020).
 30. Jenkins, B., West, J. A. & Koulman, A. A review of odd-chain fatty acid metabolism and the role of pentadecanoic acid (C15:0) and heptadecanoic acid (C17:0) in health and disease. *Molecules* **20**, 2425–2444, 20 pp. (2015).
 31. Schleyer, G. *et al.* In plaque-mass spectrometry imaging’ reveals a major metabolic shift towards odd-chain fatty acid lipids induced by host-virus interactions. *bioRxiv, Microbiol.* 1–25 (2018) doi:10.1101/317206.
 32. Smedman, A. E. M., Gustafsson, I.-B., Berglund, L. G. T. & Vessby, B. O. H. Pentadecanoic acid in serum as a marker for intake of milk fat: relations between intake of milk fat and metabolic risk factors. *Am. J. Clin. Nutr.* **69**, 22–29 (1999).
 33. Brevik, A., Veierod, M. B., Drevon, C. A. & Andersen, L. F. Evaluation of the odd fatty acids 15:0 and 17:0 in serum and adipose tissue as markers of intake of milk and dairy fat. *Eur. J. Clin. Nutr.* **59**, 1417–1422 (2005).
 34. Young, R. S. E. *et al.* Apocryphal FADS2 activity promotes fatty acid diversification in cancer. *Cell Rep.* **34**, (2021).
 35. Lu, Y., Jin, X. & Zhao, P. Serum lipids and the pathogenesis of Parkinson’s disease: A systematic review and meta-analysis. *Int. J. Clin. Pract.* **75**, e13865 (2021).
 36. Silva, Y. P., Bernardi, A. & Frozza, R. L. The Role of Short-Chain Fatty Acids From Gut Microbiota in Gut-Brain Communication . *Frontiers in Endocrinology* vol. 11 25 (2020).
 37. Zhao, H. *et al.* Potential biomarkers of Parkinson’s disease revealed by plasma metabolic profiling. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* **1081–1082**, 101–108 (2018).
 38. Baert, F. *et al.* Parkinson’s disease patients’ short chain fatty acids production capacity after in vitro fecal fiber fermentation. *npj Park. Dis.* **7**, 72 (2021).
 39. Richardson, K., Langridge, D., Dixit, S. M. & Ruotolo, B. T. An Improved Calibration Approach for Traveling Wave Ion Mobility Spectrometry: Robust, High-Precision Collision Cross Sections. *Anal. Chem. (Washington, DC, United States)* **93**, 3542–3550 (2021).

40. IMS Calibration. Available at: <https://imscal.on-demand.waters.com/home> .
41. Stow, S. M. *et al.* An Interlaboratory Evaluation of Drift Tube Ion Mobility-Mass Spectrometry Collision Cross Section Measurements. *Anal. Chem. (Washington, DC, United States)* **89**, 9048–9055 (2017).
42. Picache, J. A. *et al.* Collision cross section compendium to annotate and predict multi-omic compound identities †. (2018) doi:10.1039/c8sc04396e.
43. Bush, M. F., Campuzano, I. D. G. & Robinson, C. V. Ion Mobility Mass Spectrometry of Peptide Ions: Effects of Drift Gas and Calibration Strategies. *Anal. Chem. (Washington, DC, United States)* **84**, 7124–7130 (2012).
44. Nye, L. C. *et al.* A comparison of collision cross section values obtained via travelling wave ion mobility-mass spectrometry and ultra high performance liquid chromatography-ion mobility-mass spectrometry: Application to the characterisation of metabolites in rat urine. *J. Chromatogr. A* **1602**, 386–396 (2019).