2023

TD GC-MS and UPLC-MS methods to discover differentially expressed metabolites in sebum

Thomas D. Hoare[#], Caitlin Walton-Doyle[#], Katherine A. Hollywood, Eleanor Sinclair, Drupad K. Trivedi*, Perdita Barran*

#These authors contributed equally to the development of this protocol

*These authors are co-corresponding authors

All authors are affiliated with The Michael Barber Centre for Collaborative Mass Spectrometry, The Manchester Institute of Biotechnology, Department of Chemistry, University of Manchester, M1 7DN, United Kingdom

Corresponding authors email addresses: perdita.barran@manchester.ac.uk and drupad.trivedi@manchester.ac.uk

https://doi.org/10.26434/chemrxiv-2023-5p4tg **ORCID:** https://orcid.org/0000-0002-7720-586X Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

Abstract

Sebum is a lipid-rich biofluid produced by the sebaceous glands which are found in high abundance in areas where hair is present, along the neckline, the 'T'-area of the face and on the upper back. Recent work has shown that sebum can be used in mass spectrometry-based metabolomics experimental workflows to identify metabolic changes that occur in individuals who are infected with the SARS-CoV-2 virus, Parkinson's disease and REM-sleep Behavioural Disorders (RBD) when compared to healthy non-disease controls. We describe here standard operating protocols that can be used to analyse sebum that has been collected on gauze or cotton buds using two common and applicable hyphenated mass spectrometry (MS) methods, namely Thermal Desorption Gas Chromatography-Mass Spectrometry (TD GC-MS) and Ultra High Performance Liquid Chromatography-Mass Spectrometry (UPLC-MS). For TD GC-MS the sebum samples are analysed directly with no sample preparation requirements and for UPLC-MS we detail the extraction process for sebum prior to data acquisition. The methods are demonstrated with results from sebum obtained from in-lab healthy volunteers to permit experiments in laboratories with access to either or both of the utilised chromatography-mass spectrometry methods.

Key words

Sebum, mass spectrometry, LC-MS, GC-MS, mass spectrometry-based diagnostics, clinical mass spectrometry, Parkinson's disease, SARS-CoV-2, lipidomics, metabolomics

Acknowledgements

We thank The Michael J Fox Foundation (grant ref: 12921) and Parkinson's UK (grant ref: K-1504) for funding our work to find biomarkers for PD from sebum and to our collaborators at the Universities of Innsbruck and Surrey. 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

https://doi.org/10.26434/chemrxiv-2023-5p4tg **ORCID:** https://orcid.org/0000-0002-7720-586X Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

recruitment centres for their enthusiasm and rigour during the recruitment process. We are very grateful to all the participants who have taken part in our studies as well as PIs and nurses across all the recruiting centres. Figures 1--3 were created with Biorender.com.

Abbreviations

MS Mass Spectrometry

TD Thermal Desorption

GC-MS Gas Chromatography

UPLC-MS Ultra High Performance Liquid Chromatography-Mass Spectrometry

xC-MS Chromatography Coupled to Mass Spectrometry

SARS-CoV-2 Severe Acute Respiratory Syndrome Coronavirus 2

PD Parkinson’s Disease

RBD REM-sleep Behavioural Disorders

VOCs Volatile Organic Compounds

TIC Total Ion Chromatogram

TDU Thermal Desorption Unit

CIS Cooled Injection System

EI Electron Impact

ESI Electro Spray Ionisation

CID Collision Induced Dissociation

RT Retention Time

QC Quality Control

https://doi.org/10.26434/chemrxiv-2023-5p4tg **ORCID:** https://orcid.org/0000-0002-7720-586X Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

SST System Suitability Test

https://doi.org/10.26434/chemrxiv-2023-5p4tg **ORCID:** https://orcid.org/0000-0002-7720-586X Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

Contents

Abstract 1

Key words 1

Acknowledgements 1

Abbreviations 2

1. Introduction 5

 1.1 TD GC-MS 8

 1.2 UPLC-MS 8

2. Materials 9

 2.1 Sampling and storage materials 9

 2.2 Equipment 10

  2.2.1 TD GC-MS 10

  2.2.1 UPLC-MS 10

 2.3 Buffers and Standards 12

  2.3.1 TD GC-MS 12

  2.3.2 UPLC-MS 12

3. Methods 12

 3.1 xC-MS batch preparation 12

 3.2 TD GC-MS 14

 3.3 UPLC-MS 17

  3.3.1 Sample preparation 17

  3.3.2 Batch run (positive ionisation mode) 18

  3.3.3 Batch run (negative ionisation mode) 20

4. Notes 21

5. Supporting Information 23

 5.1 Exemplar Data 23

  5.1.1 TD GC-MS 23

  5.1.2 UPLC-MS 24

 5.2 xC-MS data processing 25

6. References 25

7. Figure Captions 29

8. Tables 31

https://doi.org/10.26434/chemrxiv-2023-5p4tg **ORCID:** https://orcid.org/0000-0002-7720-586X Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

1. Introduction

Mass spectrometry (MS) is widely used to identify and quantify the components of complex mixtures as it is highly reproducible, sensitive and can be used in high throughput methodologies. MS has applicability in clinical biochemistry settings where it is used to perform analyses on biofluids that contain known disease biomarkers ***(1,2)***. Such assays require the target biomarker(s) to be found in a complex background (matrix) and therefore benefit from chromatographic separation prior to mass spectrometric detection to ensure each species in the mixture is resolved as best as possible. Sufficient separation increases the likelihood that disease specific biomarkers can be detected, identified, and validated. MS in clinical biochemistry is typically applied to find ***(3–5)*** and/or quantify ***(6,7)*** the presence of metabolites and commonly uses: blood (either directly ***(8,9)***, as serum ***(10)*** and/or plasma ***(11)***); urine ***(12)***; or even tears ***(13)*** as rich sources of biomarkers. In this study we apply a similar methodology to the skin secreted biofluid sebum.

Sebum is a lipid-rich, oily biofluid that lubricates skin and assists thermoregulation by mixing with and slowing down the evaporation of sweat in hot weather. It forms a lipid-rich coating on skin which shields against heat loss in cold weather and provides a trapping reservoir for excreted metabolites that can be readily accessed ***(14,15)***. Sebum is released onto the surface of the skin following the holocrine secretion of sebocyte cells within sebaceous glands in the integumentary system ***(16)***. The major components of sebum are free fatty acids, wax esters, acyl glycerols, cholesterol, and squalene ***(17,18)***; the first three of these groups represent highly diverse and complex chemical classes where many differential processes contribute to what may be found in sebum ***(17)***. These constituent compounds may also be subjected to chemical modification on the surface of the skin by exogenous microflora, creating different species to add to those endogenously synthesized and released. Oxidation products of linoleic acid and squalene have been observed in human sebum ***(19)***. Many bacterial species have been shown to metabolise sebaceous secretions, for example *Malassezia restricta* has been reported to metabolise endogenously synthesised triacylglycerols, releasing diacylglycerols, monoacylglycerols and free fatty acids onto the skins surface ***(20)***. Other chemical

https://doi.org/10.26434/chemrxiv-2023-5p4tg **ORCID:** https://orcid.org/0000-0002-7720-586X Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

species from non-biological exogenous sources including cosmetics, washing agents and air pollution are also highly likely to be present on the skin's surface ***(21)***.

Sebum is an attractive prospective biofluid for disease diagnostics as sample collection is both facile and non-invasive. Additionally, the materials used for sample collection, transport and storage are inexpensive compared to other biofluids. Individuals are swabbed with a piece of gauze or a cotton bud, both of which can be done at home or in clinic. The samples are then transported to the laboratory under ambient conditions where they are stored below 20°C prior to extraction (where required), before being analysed using hyphenated mass spectrometry methods (xC-MS). We have recently shown that this cold storage is not essential ***(22)***.

A Parkinson's disease (PD) associated odour was identified by Joy Milne, and we used GC-MS with an odour port to identify the volatile components of sebum which were characteristic of this PD odour ***(1)***. Subsequently, sebum has been analysed using LC-MS to identify changes in the metabolome and lipidome of PD sebum ***(2)***. From this dataset, functional metabolic network analyses implicated the steroid hormone, carnitine shuttle and sphingolipid metabolism pathways (amongst others) as significantly altered in PD pathogenesis. A rapid 3-minute analytical method for PD diagnosis has since been developed using Paper Spray Ion Mobility-Mass spectrometry ***(23)*** which revealed the up-regulation of large lipid moieties in the sebum of people with Parkinson's. More recent refinement of the TD GC-MS method has shown that this approach can identify Parkinson's disease with 97% accuracy ***(24)***. In REM-sleep behavioural disorders (RBD) (one of the prodromal signs of PD), sebum associated odour and volatile profile have also been identified ***(24)***. Sebum has also been used in a recent study comparing the diagnostic capabilities of serum and saliva for SARS-CoV-2. Whilst serum performed best (sensitivity and specificity of 0.97), the performance of sebum was only slightly worse than serum (sensitivity 0.92; specificity 0.84). Both biofluids reflected the lipid dysregulation present in patients infected with SARS-CoV-2 ***(25)***.

Here we detail accessible protocols for sebum analysis using mass spectrometry from samples obtained remotely from individuals in the form of swabs taken on cotton (gauze or cotton bud). We first outline a method to analyse volatile organic compounds (VOCs) direct from gauze using Thermal

https://doi.org/10.26434/chemrxiv-2023-5p4tg **ORCID:** https://orcid.org/0000-0002-7720-586X Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

Desorption Gas Chromatography-Mass Spectrometry (TD GC-MS). We then describe how to analyse the non-volatile components of sebum collected from cotton buds using a reverse phase Ultra High-Performance Liquid Chromatography-Mass Spectrometry (UPLC-MS) method. We typically expect our TD GC-MS method to produce between 200--600 features and our UPLC-MS method to provide in excess of 10,000 features, suggesting that the volatilome of sebum is formed from about 5% of the available components. Figure 1 shows a summary of the presented workflows.

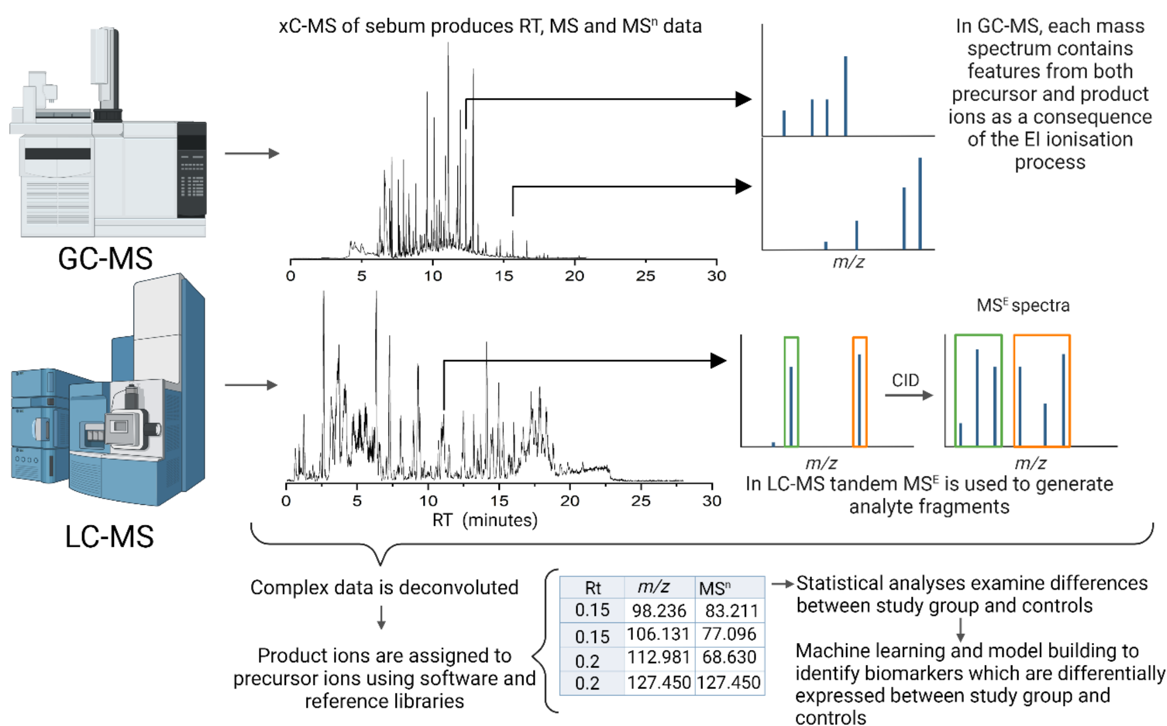

Figure 1: The workflow of our xC-MS methods. Sebum samples are separated chromatographically and metabolites undergo subsequent ionisation and detection in the mass spectrometer. Precursor ion fragmentation patterns are used alongside RT values and accurate m/z data to match against known spectral information in data libraries to assign putative identities to each analyte. Statistical analyses are performed using these data to differentiate between samples run in a study, and these findings are used alongside machine learning approaches to build models and identify differentially expressed metabolites.

https://doi.org/10.26434/chemrxiv-2023-5p4tg ORCID: https://orcid.org/0000-0002-7720-586X Content not peer-reviewed by ChemRxiv. License: CC BY-NC-ND 4.0

1.1 TD GC-MS

As detailed above, sebum contains many different molecular species with a range of chemical functionalities and molecular weights. A fraction of these are VOCs, some of which will contribute to disease and individual characteristic odours. An outline for the TD GC-MS method used to analyse the VOCs of sebum is described as follows. Prior to GC-MS analysis, a sample vial containing sebum collected on a piece of gauze is gently warmed and agitated. The headspace of this vial is then dynamically concentrated onto a thermal desorption tube (Tenax TA) with the aid of a steady flow of nitrogen. This tube is then transported to a thermal desorption unit (TDU) fitted in the inlet of the GC-MS instrument. The TDU is programmed to heat and liberate trapped VOCs into a cooled injection system (CIS) prior to the chromatography column. These VOCs are then injected by the CIS onto the GC column where a temperature gradient is applied over 22 minutes (from 40 °C to 300°C). As these analytes elute from the column they arrive at the electron impact (EI) ionisation source of the mass spectrometer where they are ionised (and subsequently a signal is detected at the detector). A consequence of EI ionisation is the fragmentation of precursor ions into patterns of smaller product ions. These fragmentation patterns are unique to each ion and can be used as a 'fingerprint' to identify an unknown parent ion/molecule. Assuming standards have been analysed on the same instrument, the retention time of a chromatographic peak (RT) can be combined with the *m/z* values of both precursor and product ions to annotate individual species in a mixture. For noting, in previous work we have used an odour port (attached to the front of the GC instrument) which allows a human nose to smell a portion of the GC eluent. This eluent is split post-column prior to routing to the mass spectrometer and is subsequently humidified at the odour port. We will not describe this olfactory method in this chapter although the interested reader is directed to an earlier publication ***(1)***.

1.2 UPLC-MS

For the analysis of less volatile, hydrophobic components of sebum it is possible to use LC-MS ***(2)***. Sebum contains many non-volatile species, and a typical LC-MS experiment with a 12 to 25 minute gradient can provide more than 10,000 features. To determine an optimum protocol for sebum-based LC-MS we refined LC methods used for serum and plasma-based metabolomics as well as lipidomics.

https://doi.org/10.26434/chemrxiv-2023-5p4tg **ORCID:** https://orcid.org/0000-0002-7720-586X Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

Our initial studies used medical gauze to collect sebum, however the subsequent method we used for the extraction of sebum for LC-MS was time consuming *(2)* (due to the length of time required for drying down large volumes of extraction solvents). For this reason, we instead collect sebum samples with cotton buds as these have a substantially smaller surface area (and consequently a much lower absorbent volume) and still yield sufficient material for analysis (data not shown). The method we present details both the extraction process and our UPLC-MS method for a single batch of samples (which includes gradients timed for both ionisation polarities). Following extraction, the non-volatile components of sebum experience chromatographic separation before being introduced to the electrospray ionisation (ESI) source of the mass spectrometer. Unlike EI ionisation (described in the TD GC-MS method above), ESI is a soft ionisation process and does not produce analyte fragments. In most LC-MS platforms it is both possible and common to use tandem MS to help identify features of interest *(8)*. In tandem MS workflows, fragment ions are instead generated downstream in the mass spectrometer through a process called collision induced dissociation (CID). This process involves the transmission of ions through a quadrupole containing a neutral buffer gas (typically argon). Ramping the energy of analyte ions inside of this quadrupole results in more energetic analyte-buffer gas collisions which fragments precursors into a series of product ions. In our method, we employ an MS^E workflow, where the CID energy is ramped from low to high to acquire both intact precursor and fragmented product ion spectra in a single acquisition. Fragmentation patterns are used alongside RT and *m/z* data to annotate species in a mixture.

2. Materials

2.1 Sampling and storage materials

Medical gauze swabs, 7.5 cm x 7.5 cm (Arco, UK) PN:4F9830

Cotton buds (Copan, VWR International, USA) PN:710-0097

Plastic zipper seal collection bags (GE Healthcare Whatman, UK) PN:11969134

https://doi.org/10.26434/chemrxiv-2023-5p4tg **ORCID:** https://orcid.org/0000-0002-7720-586X Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

Microcentrifuge tubes, 2 mL (Eppendorf, Germany) PN: 0030120094

For information on sampling methodology and sample storage see Notes 1 and 2.

2.2 Equipment

2.2.1 TD GC-MS

Glass vials, 14 mL (Scientific Glass Laboratories, UK) PN: T102/V2

Stainless steel tweezers (Scientific Laboratory Supplies, England) PN: INS4238

Glass headspace vials, 20 mL (Gerstel, Germany) PN: 093640-036-00

Magnetic caps for headspace vials (Gerstel, Germany) PN: 093640-040-00

VF-5MS column, 30 m x 0.25 mm x 0.25 um (Agilent Technologies, UK) PN: CP8944

Tenax TA adsorbent tubes (Gerstel, Germany) PN: 020810-005-00

Tenax liners (Gerstel, Germany) PN:012438-010-00

Ferrule for liners (Gerstel, Germany) PN: 007541-010-00

MPS dual head workstation robot (Gerstel, Germany)

The analysis we performed is based on the usage of a dynamic head space sampling (DHS) with a thermal desorption unit (TDU2) and a cooled injection system (CIS4) (Gerstel, Germany). These are coupled to a GC-MS instrument (GC/MSD 5975) which is interfaced by an electron impact source (Agilent Technologies, UK).

2.2.1 UPLC-MS

Mikro 120 centrifuge, 24 sample capacity (MedSupply Partners, United States) PN: HET-1204-01

MiVac vacuum centrifuge, 48 sample capacity (Genevac, UK)

Whirlimixer Vortex (Fisons, UK)

Sonicator (Sonicor, United States)

https://doi.org/10.26434/chemrxiv-2023-5p4tg **ORCID:** https://orcid.org/0000-0002-7720-586X Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

Acquity UPLC I class LC coupled to a Cyclic IMS Mass Spectrometer equipped with an electrospray ionisation source (Waters Corp, UK)

Acquity UPLC CSH C18 Column, 1.7 µm x 2.1 mm x 100 mm (Waters Corp, UK) PN: 186005297

Acquity CSH, C18 1.7 µm vanguard pre-column (Waters Corp, UK) PN:186005303

Amber Chromacol 300 µL insert chromatography LC vials (Thermo Fisher Scientific, UK) PN:03-FISV(A)

LC vial caps (Thermo Fisher Scientific, UK) PN: 9-SCK(B)-ST1

Microcentrifuge tubes, 2 mL (Eppendorf, Germany) PN: 0030120094

Glass vials, 14 mL (Scientific Glass Laboratories, UK) PN: T102/V2

Solvent bottles, 1 L (Waters Corp, UK) PN:186007089

Sartorius Secura Analytical Balance (Sartorius, Germany) PN: SA-SECURA225D-1S

https://doi.org/10.26434/chemrxiv-2023-5p4tg **ORCID:** https://orcid.org/0000-0002-7720-586X Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

2.3 Buffers and Standards

The use of Optima/LC-MS grade reagents is necessary to ensure quality and reproducibility in both analytical methods.

2.3.1 TD GC-MS

System Suitability Test : L-(-)-Carvone (Sigma Aldrich), δ-decalactone (Sigma Aldrich), Ethyl butyrate (Sigma Aldrich), Ethyl hexanoate (Sigma Aldrich), Hexadecane (Sigma Aldrich), Nonane (Sigma Aldrich),Vanillin (Sigma Aldrich), Ethanol

Tweezer cleaning solution : Methanol

2.3.2 UPLC-MS

Strong Wash: Isopropanol

Weak Wash: Acetonitrile, Water

Lock mass: Leucine Enkephalin (Waters Corp, UK) PN: 186006013

For a batch run in positive ionisation polarity:

Mobile Phase A: Ammonium Formate, Water, Acetonitrile, Formic acid

Mobile Phase B: Ammonium Formate, Water, Isopropanol, Acetonitrile, Formic acid

For a batch run in negative ionisation polarity:

Mobile Phase A: Ammonium Acetate, Water, Acetonitrile

Mobile Phase B: Ammonium Acetate, Water, Isopropanol, Acetonitrile

3. Methods

3.1 xC-MS batch preparation

For both of our xC-MS methods, a 'study' typically contains upwards of 100 samples and is divided into smaller groups of 20--30 samples called batches. A batch of samples should take no longer than a

https://doi.org/10.26434/chemrxiv-2023-5p4tg **ORCID:** https://orcid.org/0000-0002-7720-586X Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

day to be analysed and contains sebum samples from study participants, blank sampling materials, empty vials, and System Suitability Tests (SSTs) or Quality Controls (QCs) (for the TD GC-MS method and UPLC-MS method respectively). Every batch should be organised to have the same (or as close to as possible) number of samples from each class of participant in the study (See Note 3).

Prior to both the TD GC-MS and UPLC-MS method, the following steps are carried out to blind each sample and distribute participant classes within each batch:

1. Create a spreadsheet containing the IDs of each sample obtained in the study. Use an algorithm to randomly generate a range of numbers from 1 to N (where N is the total number of samples collected in a study). Assign these numbers to each sample to blind the samples.
2. Organise the blinded samples into groups of each participant class (i.e. positive/negative/control/unknown).
3. Divide the total number of samples in a study into batches of 25--30 samples. Assign equal numbers of each participant class to each batch.
4. Retrieve samples from storage and organise into the batches determined in the previous step.
5. Return each batch to storage in advance of the xC-MS method.

For the UPLC-MS method, a pooled QC ***(26)*** is produced which contains aliquots from each sebum sample in a study. Any features visible in each of the sebum samples should also be detectable in the pooled QC data; features in the data of the sebum samples which are not present in the majority of pooled QC injections are later disregarded during data analysis. A requirement of making a pooled QC is that every sample in a study is prepared up to step 12 in section 3.3.1 before a single batch may be analysed. The method in this chapter describes a batch run containing a total of 44 injections: 28 samples, 13 pooled QCs, two solvent blanks and one cotton bud blank. As we do not sample from the same vial more than four times (frequent piercings of an LC vial may introduce contaminants), four or five pooled QC vials are required per batch. The volume that is withdrawn from each sample in a study (V_{QC}, step 9 of section 3.3.1) to provide five QC vials per batch run can be calculated using the following relationship:

https://doi.org/10.26434/chemrxiv-2023-5p4tg **ORCID:** https://orcid.org/0000-0002-7720-586X Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

$$V_{QC} = \frac{5 * n_v * n_b}{n_s}$$

Where n_v is the target volume of each QC vial (in µL, step 14 section 3.3.1), n_b represents the number of batches in a study and n_s represents the total number of samples analysed in a study. The method in this chapter describes a small study containing a single batch of 28 samples (n_b = 1, n_s = 28), and takes 700 µL for n_v which results in V_{QC} = 125 µL (a 125 µL aliquot is required from each sample to make five 700 µL pooled QC vials for a batch).

When analysing a batch in both ionisation polarities, sample vials may be reused as these are pierced only once per batch analysis. As QC vials are pierced multiple times per batch, fresh QC vials are required when analysing a batch more than once. If the reader would like to analyse their batches in both ionisation polarities, double V_{QC} and use fresh QC vials when running a batch in the opposite polarity. For the TD GC-MS method, the gauze samples are analysed directly so a pooled QC cannot be easily produced. We use a SST ***(26)*** instead of a QC which is composed of a mixture of seven chemicals (Table 1). These chemicals have different RT values across the chromatography range and are used to adjust for any instrumental drift throughout the duration of a batch run. The method in this chapter describes one batch consisting of a total of 40 injections: 25 sample gauzes, 12 SST aliquots, two blank gauzes and one empty vial.

3.2 TD GC-MS (24 hours: sample preparation (~1 hour) and data acquisition (~23 hours))

Steps (11-15) describe the Gerstel GC-MS program which should be programmed in advance of a batch run.

1. Label 40 of the 20 mL glass headspace vials as the following: 25 as 'sample x' (where x is the blinded sample number), 12 as 'SST', two as 'blank' and one as 'empty'. Include the batch number on each vial. (10 minutes)
2. Prepare a mixture of the SST components in a 10 mL glass vial using the concentrations and volumes of the reagents provided in Table 1. (15 minutes)

https://doi.org/10.26434/chemrxiv-2023-5p4tg **ORCID:** https://orcid.org/0000-0002-7720-586X Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

3. Dilute this mixture to 10 mL with ethanol to create the SST stock solution (See Note 4). (5 minutes)
4. Vortex this mixture. (1 minute)
5. Aliquot 10 µL of this mixture into a 20 mL glass vial and dilute by a factor of 100 using ethanol. (2 minutes)
6. Aliquot 20 µL of this diluted SST stock into 12 of the 20 mL headspace vials labelled ‘SST’. (5 minutes)
7. Retrieve sample gauzes from storage in the -80°C freezer. (2 minutes)
8. Transfer the sample gauzes and two unsampled blank gauzes into the headspace vials labelled ‘sample x’ and ‘blank’ using tweezers. Dip tweezers in methanol after handling each sample to mitigate the transferral of volatiles between gauzes. (10 minutes)
9. Transfer each labelled headspace vial into their positions in the sample rack as described in Figure 2. (10 minutes)

10. Start the TD GC-MS run. Maintain the mass spectrometry operating parameters described in Table 2 throughout the batch.

https://doi.org/10.26434/chemrxiv-2023-5p4tg **ORCID:** https://orcid.org/0000-0002-7720-586X Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

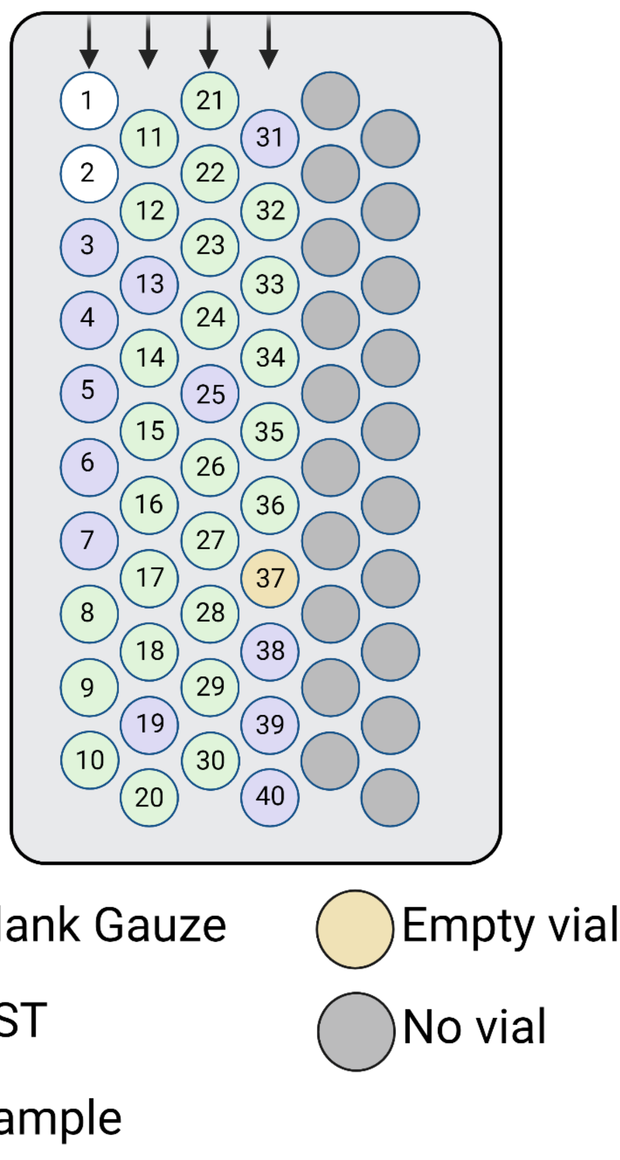

Figure 2: A plate map indicating the location of vials in the sample plate on the TD GC-MS instrument. Samples are injected sequentially from top to bottom, moving from left to right, starting with the sample located in position 1 and ending with the vial located in position 40. Blanks are injected at the start of a batch which are followed by five SST injections. From here, one SST injection occurs every five sample injections until the vial in position 37 is reached. The batch run ends with one empty injection and three SST injections.

The sorbent tube for the first sample is now transferred individually to the TD GC-MS inlet by the Gerstel workstation robot for the batch run. After each sample enters the GC column (step 14), the next sample on the plate is incubated (step 11) and purged (step 12) to save time. A batch run consists of the following steps for each vial:

https://doi.org/10.26434/chemrxiv-2023-5p4tg **ORCID:** https://orcid.org/0000-0002-7720-586X Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

11. Incubate the vial at 80 °C to preconcentrate the VOCs in the headspace of the vials. (10 minutes)
12. Purge the vial with dry nitrogen gas at a flow rate of 70 mL min^{-1}. (14.3 minutes)
13. Apply the TDU temperature ramp program detailed in Table 3. (6.5 minutes)
14. Apply the CIS and GC column temperature ramp programs detailed in Table 4 and Table 5 simultaneously. Release analytes from the CIS into the chromatography column, maintaining a column flow rate of 1 mL min^{-1} of helium for the first three minutes. (21 minutes)
15. Allow the GC column to cool down to the starting temperature (40 °C) before the next sample is introduced. (5 minutes)

3.3 UPLC-MS (26 hours: sample preparation (~5.5 hours) and data acquisition (~20.5 hours))

Steps (19--20) of section 3.3.2 describe settings for the UPLC-MS program which should be programmed in advance of a batch run.

3.3.1 Sample preparation (~5.5 hours, allow one working day)

Steps 1--12 must be completed for each sample in a study before step 13 and are described below for a single batch. Pooled QC (step 9) should be stored in the freezer during makeup across batches.

1. Label 34 of the 2 mL microcentrifuge tubes; 28 as ‘sample x’, 5 as ‘QC’, and one as ‘blank’. Where applicable, also include both the batch and sample numbers on each tube. (10 minutes)
2. Snap off the wooden stem of the 28 sample and single blank cotton bud and insert each of these cotton heads into the labelled 2 mL microcentrifuge tubes. (3 minutes)
3. Pipette 1 mL of methanol into each microcentrifuge tube. (5 minutes)
4. Vortex each tube for 10 seconds. (6 minutes)
5. Sonicate the tubes. (5 minutes)
6. Vortex each tube for 10 seconds. (6 minutes)

https://doi.org/10.26434/chemrxiv-2023-5p4tg **ORCID:** https://orcid.org/0000-0002-7720-586X Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

7. Remove, then discard the cotton bud heads from each microcentrifuge tube using tweezers. (See Note 5) (3 minutes)
8. Centrifuge the tubes at 13,800 g. (15 minutes)
9. Transfer 125 µL from the supernatant layer of each sebum sample into a 14 mL glass vial for the pooled QC (See Note 6). (8 minutes)
10. Transfer 700 µL from the supernatant layer of each sebum sample into new identically labelled microcentrifuge tubes. (15 minutes)
11. Vacuum centrifuge and concentrate these cotton bud extracts at ambient temperature. (4 hours)
12. Place all samples in the -80 °C freezer until day of analysis. (2 minutes)

Once all samples in a study have been aliquoted into the pooled QC:

13. Vortex the pooled QC. (10 seconds)
14. Aliquot 700 µL of the pooled QC into five different labelled microcentrifuge tubes for each batch. (2 minutes)
15. Vacuum centrifuge and concentrate the pooled QC extracts at ambient temperature. (4 hours)
16. Place these extracts into the -80 °C freezer with each batch until the day of analysis. (1 minute)

3.3.2 Batch run (positive ionisation mode) (20.5 hours, suggest sample run start AM)

The following method describes a batch analysis in positive ionisation mode. If the reader would like to run a batch analysis in negative ionisation mode, they are directed to section 3.3.3. To analyse a batch in both ionisation polarities, store samples in the freezer (in the LC vials) after finishing a batch run in either ionisation mode. These samples are then retrieved from storage and analysed again with fresh pooled QC vials in the opposite ionisation polarity (See Note 7).

On the day the batch is run:

https://doi.org/10.26434/chemrxiv-2023-5p4tg **ORCID:** https://orcid.org/0000-0002-7720-586X Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

1. Prepare a 25 mL stock solution of 1 M ammonium formate by dissolving 1.576 g of ammonium formate into 25 mL of water. (5 minutes)
2. Prepare 1 L of mobile phase A: Add 10 mL of the 1 M ammonium formate solution into 390 mL of water into a 1 L bottle. To this mixture, add 600 mL of acetonitrile and 1 mL of formic acid (See Note 8). (10 minutes)
3. Prepare 1 L of mobile phase B: Add 10 mL of the 1 M ammonium formate solution to 40 mL of water. To this mixture, add 850 mL of isopropanol, 100 mL of acetonitrile and 1 mL of formic acid. (10 minutes)
4. Prepare 1 L of strong wash: Add 1000 mL of isopropanol into a 1 L solvent bottle. (2 minutes)
5. Prepare 1 L of weak wash: Add 200 mL of water into a 1 L solvent bottle. Add 800 mL of acetonitrile to this mixture. (5 minutes)
6. Label 36 of the LC vials; 28 as 'sample x', 5 as 'QC', two as 'solvent blank' and one as 'cotton bud blank'. Where applicable, also include both the batch and sample numbers on each vial. (10 minutes)
7. Prepare a 1 mg/mL stock solution of leucine enkephalin (Leu-Enk) in water. (5 minutes)
8. Add 25 µL of this Leu-Enk stock solution to 250 mL acetonitrile. To this mixture, add 245 mL water and 5 mL of formic acid to make the lock mass solution which maintains mass calibration during batches (Waters Corp.). (10 minutes)
9. Sonicate both mobile phases. Ensure there is no salt visible in either solution before proceeding with the batch (see Note 9). (30 minutes)
10. Connect both mobile phases and the Leu-Enk solution to the LC system and purge as per manufacturers guidance. (5 minutes)
11. Retrieve the batch of microcentrifuge tubes from the freezer (samples and QCs). (2 minute)
12. Add 200 µL methanol to the dried extracts in each tube at ambient temperature. (8 minutes)
13. Vortex each sample for 10 seconds. (6 minutes)
14. Sonicate the samples. (5 minutes)
15. Centrifuge the samples at 13,800 g. (15 minutes)

https://doi.org/10.26434/chemrxiv-2023-5p4tg **ORCID:** https://orcid.org/0000-0002-7720-586X Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

16. Withdraw 160 μL from the supernatant layer of each tube and aliquot into the corresponding labelled LC-MS vials. (5 minutes)
17. Aliquot 160 μL of methanol into the two corresponding labelled LC-MS vials. (1 minute)
18. Transfer each LC-MS vial into their positions in the autosampler racks as described in Figure 3. (10 minutes)
19. Program the injection order on the UPLC-MS system using Figure 3 as guidance. (45 minutes)
20. Operate the LC program with the settings detailed in Table 6 and Table 7. Wash the inlet syringe with a 2 mL:2 mL (weak:strong) wash before each injection. (24 minutes per injection)

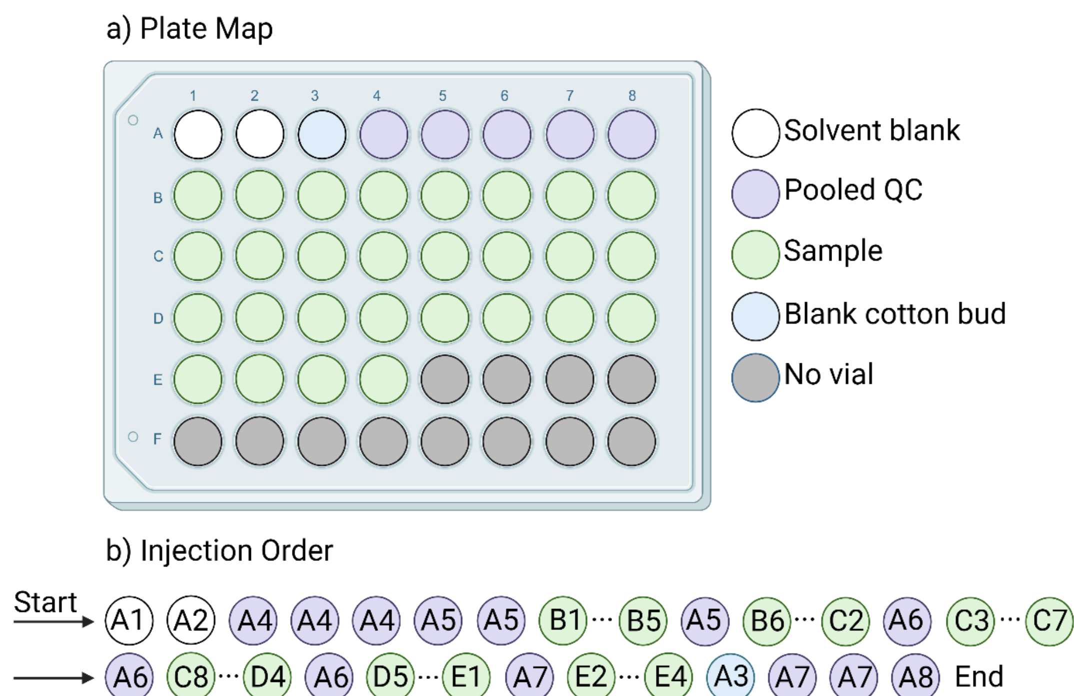

Figure 3: A plate map indicating a) the location of vials in the sample plate on the LC-MS instrument alongside b) the order of injection of these samples into the instrument. Blanks are injected at the start of a batch followed by five pooled QC injections. From here, one QC injection occurs every five sample injections until the end of the batch. The batch run ends with one blank injection and three QC injections.

3.3.3 Batch run (negative ionisation mode)

https://doi.org/10.26434/chemrxiv-2023-5p4tg ORCID: https://orcid.org/0000-0002-7720-586X Content not peer-reviewed by ChemRxiv. License: CC BY-NC-ND 4.0

When running a batch in negative ionisation polarity, replace steps 1--3 of section 3.3.2 with the following:

1. Prepare a 25 mL stock solution of 1 M ammonium acetate by dissolving 1.927 g of ammonium acetate into 25 mL of water. (5 minutes)
2. Prepare 1 L of mobile phase A: Add 10 mL of the 1 M ammonium acetate solution into 390 mL of water into a 1 L bottle. To this mixture, add 600 mL of acetonitrile (see Note 9) : Do not add isopropanol directly to the salt solution in step 2 or step 3 as the salt solution may precipitate. (10 minutes)
3. Prepare 1 L of mobile phase B: Add 10 mL of the 1 M ammonium acetate solution to 40 mL of water. To this mixture, add 850 mL of isopropanol and 100 mL of acetonitrile. (10 minutes)

4. Notes

Note 1: Further details on sampling methodology for both gauze and cotton buds can be found here ***(1,2,23,27)***.

Note 2: Following sampling, gauzes are placed in zipper seal collection bags and cotton buds are stored either in the plastic casing they arrive in or the wooden stems are snapped off and the cotton heads are stored in individual microcentrifuge tubes (if storage space is a concern). Both gauzes and cotton buds are stored at -80°C prior to analysis (although storage at ambient temperatures also yields similar results ***(22)***).

Note 3: Organising batches in this way allows sample variation to be distinguished from day to day instrument variations and mitigates the risk of losing a disproportionate amount of one sample class in a single failed batch.

Note 4: This stock solution remains stable in the fridge and may be reused for future batches.

Note 5: Squeeze each head against the side of the tube whilst removing to retain as much supernatant as possible.

https://doi.org/10.26434/chemrxiv-2023-5p4tg **ORCID:** https://orcid.org/0000-0002-7720-586X Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

Note 6: This volume will be less if you intend to run multiple batches in a study with more than 28 samples. (See section 3.1).

Note 7: Leaving samples in the running solution for longer than 10 days is non-optimal (data not shown) and if there is a time gap due to instrument availability, each sample should be dried down as described in step 11 of section 3.3.1 where they may be reconstituted prior to analysis as required (step 12 of section 3.3.2).

Note 8: Do not add isopropanol directly to the salt solution in steps 2--3 as the salt solution may precipitate.

Note 9: If there is still salt visible in either solution after 30 minutes, sonicate the mobile phase(s) for an extra 10 minutes (or as long as required).

https://doi.org/10.26434/chemrxiv-2023-5p4tg **ORCID:** https://orcid.org/0000-0002-7720-586X Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

5. Supporting Information

5.1 Exemplar Data

5.1.1 TD GC-MS

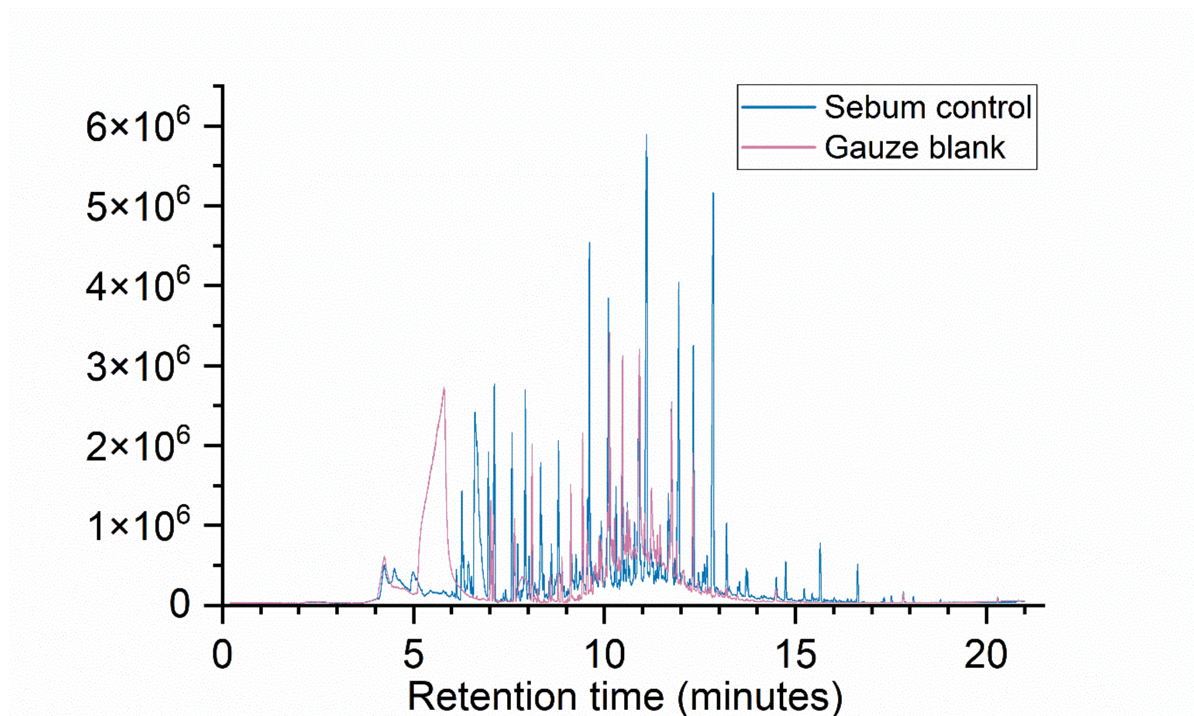

Figure 4: Exemplar data comparing the base peak intensity total ion chromatograms of two samples prepared and analysed using the TD GC-MS method. A sebum control gauze (blue) is overlaid on top of a blank gauze (pink) to aid in the comparison of features between the two samples.

Figure 4 above shows typical GC-MS TICs for sebum swabbed and blank gauze. We typically expect between 200--600 metabolite features per sample from our TD GC-MS method. After finishing a batch run, raw data files are converted to .mzxml files using Proteowizard ***(28)*** with processing parameter choice and tuning that needs to be refined by the user. Once the data is processed putative annotations can be found using different available databases. Examples of post-acquisition processing, feature annotation and machine learning approaches are provided in the data processing section.

https://doi.org/10.26434/chemrxiv-2023-5p4tg **ORCID:** https://orcid.org/0000-0002-7720-586X Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

5.1.2 UPLC-MS

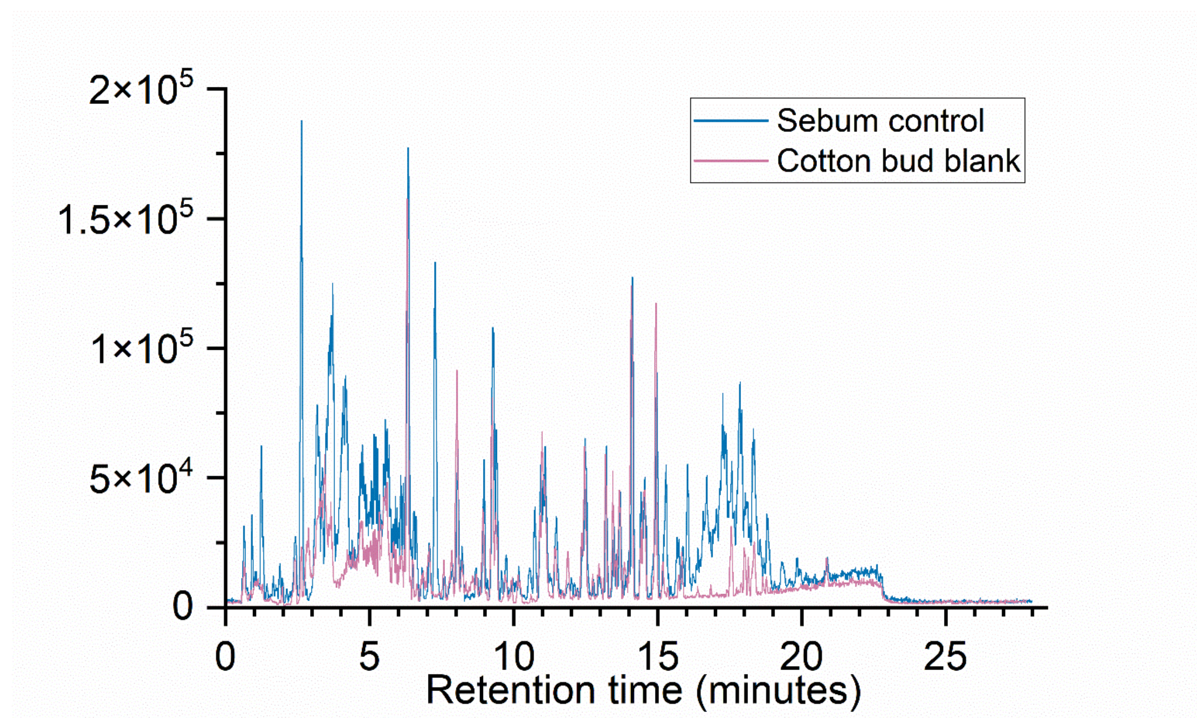

Figure 5: Exemplar data comparing the base peak intensity total ion chromatograms of two samples prepared and analysed using the UPLC-MS method. The sebum control (blue) is overlaid on top of a blank cotton bud (pink) to aid in the comparison of features between the two samples.

Figure 5 above shows typical TICs for extracted cotton buds both sebum swabbed and a blank. We typically expect over 10,000 metabolite features per sample in our UPLC-MS method. After a batch is analysed the raw LC-MS data is aligned to a QC injection which is selected by, deconvoluted and peak picked in Progenesis (Waters). Similar vendor software, alongside packages such as XCMS ***(29)*** or MetaboAnalystR ***(30)*** in R ***(31)*** may also be used. Putative identifications (MSI level 2 ***(32)***) are then assigned using Lipid Maps ***(33,34)***, Lipid Blast ***(35)***, HMDB ***(36)*** and Metlin **(37)**. The features from Progenesis can be filtered using variations between (and missing values in) QC injections in addition to the fold change compared to cotton bud blanks. The resultant output can be analysed using the approach discussed in the data processing section as before with the TD GC-MS method due to the shared inherent multidimensionality in the data.

https://doi.org/10.26434/chemrxiv-2023-5p4tg **ORCID:** https://orcid.org/0000-0002-7720-586X Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

5.2 xC-MS data processing

For TD GC-MS data, raw data is converted to .mzxml files using Proteowizard. Proteowizard parameters which may require tuning include binary precision (64 bit), zlib compression, TPP compatibility and the write index. The .mzxml files are then deconvolved in R using eRah (a package in R)***(38)*** and inhouse scripts which include peak picking parameters and alignment, alongside the annotation of deconvolved features using imported libraries. These scripts are available on request. Putative annotations can be assigned (MSI level 2 ***(32)***) using the GOLM database ***(39)*** or NIST 14.0 library. We recommend a spectral match threshold set at 70%.

The deconvolved and putatively annotated output can be analysed using various statistical platforms, which may include open source resources such as R ***(31)*** or Python ***(40)***, commercial software such as Origin(Pro), MATLAB ***(41)*** or SIMCA-P ***(42)***. Online servers such as MetaboAnalyst ***(43)*** may be used for univariate analyses, multivariate analyses, machine learning approaches, metabolic pathway analysis and data visualisation. Commonly used univariate approaches include Student's *t*-test, Analysis of Variance (ANOVA), Receiver Operating Characteristic analysis (ROC) and correlation analysis. For Multivariate approaches, dimension reduction techniques (such as Principal Component Analysis (PCA) or Multi-Dimensional Scaling (MDS)) and clustering methods (such as k-Nearest Neighbours (kNN) or Partial Least Squares Discriminant Analysis (PLS-DA) and its variants) may be employed. Machine learning approaches include simple Linear Discriminant Analysis (LDA), Support Vector Machines (SVM), and decision tree based methods and ensemble methods. Due to the nature of LC-MS and GC-MS data, dimension reduction and feature selection is often mandatory prior to machine learning. This can be achieved using a variety of ranking methods including Variable Importance in Projections (VIP) scores, the ReliefF algorithm and recursive feature elimination (RFE) (or any similar approaches).

6. References

(1) Trivedi, D. K.; Sinclair, E.; Xu, Y.; Sarkar, D.; Walton-Doyle, C.; Liscio, C.; Banks, P.; Milne, J.; Silverdale, M.; Kunath, T.; Goodacre, R.; Barran, P. Discovery of Volatile Biomarkers of Parkinson's Disease from Sebum. *ACS Cent. Sci.* **2019**, *5* (4), 599–606. https://doi.org/10.1021/acscentsci.8b00879.

https://doi.org/10.26434/chemrxiv-2023-5p4tg **ORCID:** https://orcid.org/0000-0002-7720-586X Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

(2) Sinclair, E.; Trivedi, D. K.; Sarkar, D.; Walton-Doyle, C.; Milne, J.; Kunath, T.; Rijs, A. M.; de Bie, R. M. A.; Goodacre, R.; Silverdale, M.; Barran, P. Metabolomics of Sebum Reveals Lipid Dysregulation in Parkinson's Disease. *Nat. Commun.* **2021**, *12* (1), 1592. https://doi.org/10.1038/s41467-021-21669-4.

(3) Millington, D. S.; Kodo, N.; Terada, N.; Roe, D.; Chace, D. H. The Analysis of Diagnostic Markers of Genetic Disorders in Human Blood and Urine Using Tandem Mass Spectrometry with Liquid Secondary Ion Mass Spectrometry. *Int. J. Mass Spectrom. Ion Process.* **1991**, *111*, 211–228. https://doi.org/10.1016/0168-1176(91)85056-R.

(4) Sugimoto, M.; Wong, D. T.; Hirayama, A.; Soga, T.; Tomita, M. Capillary Electrophoresis Mass Spectrometry-Based Saliva Metabolomics Identified Oral, Breast and Pancreatic Cancer-Specific Profiles. *Metabolomics* **2010**, *6* (1), 78–95. https://doi.org/10.1007/s11306-009-0178-y.

(5) Hu, J.-R.; Coresh, J.; Inker, L. A.; Levey, A. S.; Zheng, Z.; Rebholz, C. M.; Tin, A.; Appel, L. J.; Chen, J.; Sarnak, M. J.; Grams, M. E. Serum Metabolites Are Associated with All-Cause Mortality in Chronic Kidney Disease. *Kidney Int.* **2018**, *94* (2), 381–389. https://doi.org/10.1016/j.kint.2018.03.008.

(6) Chang, K.-H.; Cheng, M.-L.; Tang, H.-Y.; Huang, C.-Y.; Wu, Y.-R.; Chen, C.-M. Alternations of Metabolic Profile and Kynurenine Metabolism in the Plasma of Parkinson's Disease. *Mol. Neurobiol.* **2018**, *55* (8), 6319–6328. https://doi.org/10.1007/s12035-017-0845-3.

(7) Marksteiner, J.; Blasko, I.; Kemmler, G.; Koal, T.; Humpel, C. Bile Acid Quantification of 20 Plasma Metabolites Identifies Lithocholic Acid as a Putative Biomarker in Alzheimer's Disease. *Metabolomics* **2017**, *14* (1), 1. https://doi.org/10.1007/s11306-017-1297-5.

(8) Thomas, S. N.; French, D.; Jannetto, P. J.; Rappold, B. A.; Clarke, W. A. Liquid Chromatography–Tandem Mass Spectrometry for Clinical Diagnostics. *Nat. Rev. Methods Primer* **2022**, *2* (1), 96. https://doi.org/10.1038/s43586-022-00175-x.

(9) Chace, D. H. Mass Spectrometry in Newborn and Metabolic Screening: Historical Perspective and Future Directions. *J. Mass Spectrom. JMS* **2009**, *44* (2), 163–170. https://doi.org/10.1002/jms.1528.

(10) Bernardo-Bermejo, S.; Xue, J.; Hoang, L.; Billings, E.; Webb, B.; Honders, M. W.; Venneker, S.; Heijs, B.; Castro-Puyana, M.; Marina, M. L.; van den Akker, E. B.; Griffioen, M.; Siuzdak, G.; Giera, M.; Sánchez-López, E. Quantitative Multiple Fragment Monitoring with Enhanced In-Source Fragmentation/Annotation Mass Spectrometry. *Nat. Protoc.* **2023**, *18* (4), 1296–1315. https://doi.org/10.1038/s41596-023-00803-0.

(11) Maurer, H. H. Multi-Analyte Procedures for Screening for and Quantification of Drugs in Blood, Plasma, or Serum by Liquid Chromatography-Single Stage or Tandem Mass Spectrometry (LC-MS or LC-MS/MS) Relevant to Clinical and Forensic Toxicology. *Clin. Biochem.* **2005**, *38* (4), 310–318. https://doi.org/10.1016/j.clinbiochem.2005.01.014.

(12) Kentsis, A.; Monigatti, F.; Dorff, K.; Campagne, F.; Bachur, R.; Steen, H. Urine Proteomics for Profiling of Human Disease Using High Accuracy Mass Spectrometry. *Proteomics – Clin. Appl.* **2009**, *3* (9), 1052–1061. https://doi.org/10.1002/prca.200900008.

(13) Zhan, X.; Li, J.; Guo, Y.; Golubnitschaja, O. Mass Spectrometry Analysis of Human Tear Fluid Biomarkers Specific for Ocular and Systemic Diseases in the Context of 3P Medicine. *EPMA J.* **2021**, *12* (4), 449–475. https://doi.org/10.1007/s13167-021-00265-y.

https://doi.org/10.26434/chemrxiv-2023-5p4tg **ORCID:** https://orcid.org/0000-0002-7720-586X Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

(14) Stewart, M. E.; Downing, D. T. Chemistry and Function of Mammalian Sebaceous Lipids. In *Skin Lipids*; ELIAS, P. M., Ed.; Advances in Lipid Research; Elsevier, 1991; Vol. 24, pp 263–301. https://doi.org/10.1016/B978-0-12-024924-4.50013-4.

(15) Géhin, C.; Tokarska, J.; Fowler, S. J.; Barran, P. E.; Trivedi, D. K. No Skin off Your Back: The Sampling and Extraction of Sebum for Metabolomics. *Metabolomics Off. J. Metabolomic Soc.* **2023**, *19* (4), 21. https://doi.org/10.1007/s11306-023-01982-3.

(16) Schneider, M. R.; Paus, R. Sebocytes, Multifaceted Epithelial Cells: Lipid Production and Holocrine Secretion. *Int. J. Biochem. Cell Biol.* **2010**, *42* (2), 181–185. https://doi.org/10.1016/j.biocel.2009.11.017.

(17) Nicolaides, N. Skin Lipids: Their Biochemical Uniqueness. *Science* **1974**, *186* (4158), 19–26. https://doi.org/10.1126/science.186.4158.19.

(18) Nikkari, T. Comparative Chemistry of Sebum. *J. Invest. Dermatol.* **1974**, *62* (3), 257–267. https://doi.org/10.1111/1523-1747.ep12676800.

(19) Ishikawa, A.; Ito, J.; Shimizu, N.; Kato, S.; Kobayashi, E.; Ohnari, H.; Sakata, O.; Naru, E.; Nakagawa, K. Linoleic Acid and Squalene Are Oxidized by Discrete Oxidation Mechanisms in Human Sebum. *Ann. N. Y. Acad. Sci.* **2021**, *1500* (1), 112–121. https://doi.org/10.1111/nyas.14615.

(20) Ro, B. I.; Dawson, T. L. The Role of Sebaceous Gland Activity and Scalp Microfloral Metabolism in the Etiology of Seborrheic Dermatitis and Dandruff. *J. Investig. Dermatol. Symp. Proc.* **2005**, *10* (3), 194–197. https://doi.org/10.1111/j.1087-0024.2005.10104.x.

(21) Grice, E. A.; Segre, J. A. The Skin Microbiome. *Nat. Rev. Microbiol.* **2011**, *9* (4), 244–253. https://doi.org/10.1038/nrmicro2537.

(22) Walton-Doyle, C.; Sinclair, E.; Begum, H.; Hollywood, K.; Trivedi, D.; Barran, P. Does Storage Post Sampling Influence the Stability of Sebum When Used for Mass Spectrometry Metabolomics Analysis? ChemRxiv October 4, 2023. https://doi.org/10.26434/chemrxiv-2023-wxv2b.

(23) Sarkar, D.; Sinclair, E.; Lim, S. H.; Walton-Doyle, C.; Jafri, K.; Milne, J.; Vissers, J. P. C.; Richardson, K.; Trivedi, D. K.; Silverdale, M.; Barran, P. Paper Spray Ionization Ion Mobility Mass Spectrometry of Sebum Classifies Biomarker Classes for the Diagnosis of Parkinson's Disease. *JACS Au* **2022**, *2* (9), 2013–2022. https://doi.org/10.1021/jacsau.2c00300.

(24) Caitlin Walton-Doyle; Beatrice Heim; Eleanor Sinclair; Katherine A Hollywood; Joy Milne; Evi Holzknecht; Ambra Stefani; Birgit Högl; Klaus Seppi; Monty Silverdale; Werner Poewe; Perdita Barran; Drupad K Trivedi. Investigation of Volatile Metabolites in Sebum as Prodromal Indicators of Parkinson's Disease. *bioRxiv* **2023**, 2023.03.01.530578. https://doi.org/10.1101/2023.03.01.530578.

(25) Spick, M.; Lewis, H.-M.; Frampas, C. F.; Longman, K.; Costa, C.; Stewart, A.; Dunn-Walters, D.; Greener, D.; Evetts, G.; Wilde, M. J.; Sinclair, E.; Barran, P. E.; Skene, D. J.; Bailey, M. J. An Integrated Analysis and Comparison of Serum, Saliva and Sebum for COVID-19 Metabolomics. *Sci. Rep.* **2022**, *12* (1), 11867. https://doi.org/10.1038/s41598-022-16123-4.

(26) Broadhurst, D.; Goodacre, R.; Reinke, S. N.; Kuligowski, J.; Wilson, I. D.; Lewis, M. R.; Dunn, W. B. Guidelines and Considerations for the Use of System Suitability and Quality Control Samples in Mass Spectrometry Assays Applied in Untargeted Clinical Metabolomic Studies. *Metabolomics* **2018**, *14* (6), 72. https://doi.org/10.1007/s11306-018-1367-3.

(27) Sinclair, E.; Walton-Doyle, C.; Sarkar, D.; Hollywood, K. A.; Milne, J.; Lim, S. H.; Kunath, T.; Rijs, A. M.; de Bie, R. M. A.; Silverdale, M.; Trivedi, D. K.; Barran, P. Validating Differential

https://doi.org/10.26434/chemrxiv-2023-5p4tg **ORCID:** https://orcid.org/0000-0002-7720-586X Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

Volatilome Profiles in Parkinson's Disease. *ACS Cent. Sci.* **2021**, *7* (2), 300–306. https://doi.org/10.1021/acscentsci.0c01028.

(28) Chambers, M. C.; Maclean, B.; Burke, R.; Amodei, D.; Ruderman, D. L.; Neumann, S.; Gatto, L.; Fischer, B.; Pratt, B.; Egertson, J.; Hoff, K.; Kessner, D.; Tasman, N.; Shulman, N.; Frewen, B.; Baker, T. A.; Brusniak, M.-Y.; Paulse, C.; Creasy, D.; Flashner, L.; Kani, K.; Moulding, C.; Seymour, S. L.; Nuwaysir, L. M.; Lefebvre, B.; Kuhlmann, F.; Roark, J.; Rainer, P.; Detlev, S.; Hemenway, T.; Huhmer, A.; Langridge, J.; Connolly, B.; Chadick, T.; Holly, K.; Eckels, J.; Deutsch, E. W.; Moritz, R. L.; Katz, J. E.; Agus, D. B.; MacCoss, M.; Tabb, D. L.; Mallick, P. A Cross-Platform Toolkit for Mass Spectrometry and Proteomics. *Nat. Biotechnol.* **2012**, *30* (10), 918–920. https://doi.org/10.1038/nbt.2377.

(29) Smith; C.A.; Want; E.J.; O'Maille; G.; Abagyan,R.; Siuzdak; G. XCMS: Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching and Identification. *Analytical Chemistry*, 2006, *78*, 779–787.

(30) Chong, J.; Xia, J. MetaboAnalystR: An R Package for Flexible and Reproducible Analysis of Metabolomics Data. *Bioinformatics* **2018**, *34* (24), 4313–4314. https://doi.org/10.1093/bioinformatics/bty528.

(31) R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2021.

(32) Sumner, L. W.; Amberg, A.; Barrett, D.; Beale, M. H.; Beger, R.; Daykin, C. A.; Fan, T. W.-M.; Fiehn, O.; Goodacre, R.; Griffin, J. L.; Hankemeier, T.; Hardy, N.; Harnly, J.; Higashi, R.; Kopka, J.; Lane, A. N.; Lindon, J. C.; Marriott, P.; Nicholls, A. W.; Reily, M. D.; Thaden, J. J.; Viant, M. R. Proposed Minimum Reporting Standards for Chemical Analysis. *Metabolomics* **2007**, *3* (3), 211–221. https://doi.org/10.1007/s11306-007-0082-2.

(33) Fahy, E.; Sud, M.; Cotter, D.; Subramaniam, S. LIPID MAPS Online Tools for Lipid Research. *Nucleic Acids Res.* **2007**, *35* (suppl_2), W606–W612. https://doi.org/10.1093/nar/gkm324.

(34) Sud, M.; Fahy, E.; Cotter, D.; Brown, A.; Dennis, E. A.; Glass, C. K.; Merrill, J., Alfred H.; Murphy, R. C.; Raetz, C. R. H.; Russell, D. W.; Subramaniam, S. LMSD: LIPID MAPS Structure Database. *Nucleic Acids Res.* **2006**, *35* (suppl_1), D527–D532. https://doi.org/10.1093/nar/gkl838.

(35) Kind, T.; Liu, K.-H.; Lee, D. Y.; DeFelice, B.; Meissen, J. K.; Fiehn, O. LipidBlast in Silico Tandem Mass Spectrometry Database for Lipid Identification. *Nat. Methods* **2013**, *10* (8), 755–758. https://doi.org/10.1038/nmeth.2551.

(36) Wishart, D. S.; Guo, A.; Oler, E.; Wang, F.; Anjum, A.; Peters, H.; Dizon, R.; Sayeeda, Z.; Tian, S.; Lee, B. L.; Berjanskii, M.; Mah, R.; Yamamoto, M.; Jovel, J.; Torres-Calzada, C.; Hiebert-Giesbrecht, M.; Lui, V. W.; Varshavi, D.; Varshavi, D.; Allen, D.; Arndt, D.; Khetarpal, N.; Sivakumaran, A.; Harford, K.; Sanford, S.; Yee, K.; Cao, X.; Budinski, Z.; Liigand, J.; Zhang, L.; Zheng, J.; Mandal, R.; Karu, N.; Dambrova, M.; Schiöth, H. B.; Greiner, R.; Gautam, V. HMDB 5.0: The Human Metabolome Database for 2022. *Nucleic Acids Res.* **2022**, *50* (D1), D622–D631. https://doi.org/10.1093/nar/gkab1062.

(37) Smith, C. A.; O'Maille, G.; Want, E. J.; Qin, C.; Trauger, S. A.; Brandon, T. R.; Custodio, D. E.; Abagyan, R.; Siuzdak, G. METLIN: A Metabolite Mass Spectral Database. *Ther. Drug Monit.* **2005**, *27* (6), 747–751. https://doi.org/10.1097/01.ftd.0000179845.53213.39.

(38) Domingo-Almenara, X.; Brezmes, J.; Vinaixa, M.; Samino, S.; Ramirez, N.; Ramon-Krauel, M.; Lerin, C.; Díaz, M.; Ibáñez, L.; Correig, X.; Perera-Lluna, A.; Yanes, O. eRah: A Computational Tool Integrating Spectral Deconvolution and Alignment with Quantification and Identification of

https://doi.org/10.26434/chemrxiv-2023-5p4tg **ORCID:** https://orcid.org/0000-0002-7720-586X Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

Metabolites in GC/MS-Based Metabolomics. *Anal. Chem.* **2016**, *88* (19), 9821–9829. https://doi.org/10.1021/acs.analchem.6b02927.

(39) Kopka, J.; Schauer, N.; Krueger, S.; Birkemeyer, C.; Usadel, B.; Bergmüller, E.; Dörmann, P.; Weckwerth, W.; Gibon, Y.; Stitt, M.; Willmitzer, L.; Fernie, A. R.; Steinhauser, D. GMD@CSB.DB: The Golm Metabolome Database. *Bioinformatics* **2004**, *21* (8), 1635–1638. https://doi.org/10.1093/bioinformatics/bti236.

(40) Van Rossum, G.; Drake Jr, F. L. *Python Reference Manual*; Centrum voor Wiskunde en Informatica Amsterdam, 1995.

(41) Inc, T. M. MATLAB Version: 9.13.0 (R2022b), 2022. https://www.mathworks.com.

(42) Wu, Z.; Li, D.; Meng, J.; Wang, H. Introduction to SIMCA-P and Its Application. In *Handbook of Partial Least Squares: Concepts, Methods and Applications*; Esposito Vinzi, V., Chin, W. W., Henseler, J., Wang, H., Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2010; pp 757–774. https://doi.org/10.1007/978-3-540-32827-8_33.

(43) Pang, Z.; Zhou, G.; Ewald, J.; Chang, L.; Hacariz, O.; Basu, N.; Xia, J. Using MetaboAnalyst 5.0 for LC–HRMS Spectra Processing, Multi-Omics Integration and Covariate Adjustment of Global Metabolomics Data. *Nat. Protoc.* **2022**, *17* (8), 1735–1761. https://doi.org/10.1038/s41596-022-00710-w.

7. Figure Captions

Figure 1: The workflow of our xC-MS methods. Sebum samples are separated chromatographically and metabolites undergo subsequent ionisation and detection in the mass spectrometer. Precursor ion fragmentation patterns are used alongside RT values and accurate *m/z* data to match against known spectral information in data libraries to assign putative identities to each analyte. Statistical analyses are performed using these data to differentiate between samples run in a study, and these findings are used alongside machine learning approaches to build models and identify differentially expressed metabolites.

Figure 2: A plate map indicating the location of vials in the sample plate on the TD GC-MS instrument. Samples are injected sequentially from top to bottom, moving from left to right, starting with the sample located in position 1 and ending with the vial located in position 40. Blanks are injected at the start of a batch which are followed by five SST injections. From here, one SST injection occurs every five sample injections until the vial in position 37 is reached. The batch run ends with one empty injection and three SST injections.

https://doi.org/10.26434/chemrxiv-2023-5p4tg **ORCID:** https://orcid.org/0000-0002-7720-586X Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

Figure 3: A plate map indicating a) the location of vials in the sample plate on the LC-MS instrument alongside b) the order of injection of these samples into the instrument. Blanks are injected at the start of a batch followed by five pooled QC injections. From here, one QC injection occurs every five sample injections until the end of the batch. The batch run ends with one blank injection and three QC injections.

Figure 4: Exemplar data comparing the base peak intensity total ion chromatograms of two samples prepared and analysed using the TD GC-MS method. A sebum control gauze (blue) is overlaid on top of a blank gauze (pink) to aid in the comparison of features between the two samples.

Figure 5: Exemplar data comparing the base peak intensity total ion chromatograms of two samples prepared and analysed using the UPLC-MS method. The sebum control (blue) is overlaid on top of a blank cotton bud (pink) to aid in the comparison of features between the two samples.

https://doi.org/10.26434/chemrxiv-2023-5p4tg **ORCID:** https://orcid.org/0000-0002-7720-586X Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

8. Tables

Table 1. Composition of the SST for the TD GC-MS method.

Reagent	Concentration (µg/mL)	Volume added (µL)
L(-)-Carvone	2,500	106
δ-decalactone	10,000	107
Ethyl butyrate	10,000	114
Ethyl hexanoate	5,000	57.3
Hexadecane	5,000	129
Nonane	5,000	139
Vanillin	10,000	100 (mg)

Table 2. Mass spectrometry parameters for the Agilent GC/MSD 5975 mass spectrometer used in the TD GC-MS method.

Parameter	Value
EI source (°C)	230
Quadrupole temperature (°C)	150
GC-MS transfer line (°C)	300
m/z acquisition range (Th)	30-800

Table 3. Temperature ramp settings for the TDU program for the TD GC-MS method.

TDU Program steps	Duration (s)
Set and maintain the temperature at 30 °C	60
Apply a temperature ramp of 12 °C s^{-1} from 30 °C to 280 °C	21
Maintain the temperature at 280 °C	300

https://doi.org/10.26434/chemrxiv-2023-5p4tg **ORCID:** https://orcid.org/0000-0002-7720-586X Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

Table 4. Temperature ramp settings for the CIS program for the TD GC-MS method.

CIS Program steps	Duration (s)
Set and maintain the temperature at 10 °C	120
Apply a temperature ramp of 12 °C s^{-1} from 10 °C to 280 °C	23
Maintain the temperature at 280 °C	300

Table 5. Temperature ramp settings for the GC column for the TD GC-MS method.

GC program steps	Duration (s)
Set and maintain the temperature at 40 °C	60
Apply a temperature ramp of 25 °C min^{-1} from 40 °C to 180 °C	336
Apply a temperature ramp of 8 °C min^{-1} from 180 °C to 240 °C	450
Maintain the temperature at 240 °C	60
Apply a temperature ramp of 20 °C min^{-1} from 240 °C to 300 °C	180
Maintain the temperature at 300 °C	174

Table 6. The composition of the UPLC-MS solvent gradient throughout a batch. Where solvent system A is acetonitrile:water:buffer (v:v:v, 60:39:1), system B is isopropanol:acetonitrile:water:buffer (v:v:v:v, 85:10:4:1). The buffer solution constitutes either 1 M ammonium formate and 1 mL of formic acid (for a batch run in positive ionisation polarity) or 1 M ammonium acetate (for a batch run in negative ionisation polarity).

Time (minutes)	Solvent A (%)	Solvent B (%)	Column Flow Rate (mL min^{-1})
0	90	10	0.4
4	50	50	0.4
11.6	30	70	0.4
11.7	26	74	0.4

https://doi.org/10.26434/chemrxiv-2023-5p4tg ORCID: https://orcid.org/0000-0002-7720-586X Content not peer-reviewed by ChemRxiv. License: CC BY-NC-ND 4.0

16	13	87	0.4
18	5	95	0.4
18.1	90	10	0.4
24	90	10	0.4

Table 7. Mass spectrometry parameters for the Waters Cyclic IMS mass spectrometer used in the UPLC-MS method.

Parameter	Positive Mode	Negative Mode
Capillary voltage (V)	2000	1500
Source temperature (°C)	120	120
Desolvation gas flow (mL min^{-1})	1200	1200
Desolvation gas temperature (°C)	550	550
Nebuliser gas flow (mL min^{-1})	6	6
Auto trap collision energy (eV)	6	-6
Auto transfer MS collision energy (eV) (low)	15	-15
Auto transfer MS collision energy (eV) (high)	45	-45
m/z acquisition range (Th)	50-2000	50-2000

https://doi.org/10.26434/chemrxiv-2023-5p4tg **ORCID:** https://orcid.org/0000-0002-7720-586X Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0