Evaluation of chemometric software for analysis of complex mixtures for biologically derived samples analysed using liquid-chromatography mass spectrometry

Megan Rose Scott

Thesis submitted for the degree of Doctor of Philosophy

June 2024

Keele University

Abstract

There is a risk of losing important information when choosing the best metabolomic workflow for untargeted chemometric analysis. Choosing the best software for conducting this analysis is crucial as wrong parameters with the wrong software could lead to false negatives, as well as over-saturating the data analysis with false positives. Over the course of this study, the intention was to show how a robust untargeted liquid chromatography-mass spectrometry method, followed by deconvolution then performing statistical analysis can determine consistent, concise and accurate markers that explain differences between datasets. Different software packages were used throughout to determine whether the chosen software affects the results.

Software packages for deconvolution and statistical analysis were compared over a range of different samples to evaluate the workflow over a range of sample types; plant samples, solid human products (hair samples) and liquid human products (blood samples) were used. The different deconvolution software packages used showed different results through the studies showing that the software used will affect the outcome. Though some markers were consistent in the statistical analysis performed with the same deconvolution, a lot of the results were different which shows that conducting the analysis in different types of software, results in different biomarker detection. This could lead to a potential oversight and loss of important information.

The data showed that deconvolution worked best in Mass Profinder then statistical analysis in MPP gave the most reliable results whilst being the easiest to navigate. However, where possible, it was concluded that more than one type of software should be used for reliable biomarker detection to reduce the risk of losing important information through the software choice.

Contents

List of Abbreviations

HPLC	High Performance Liquid Chromatography
UPLC	Ultra Performance Liquid Chromatography
ToF	Time of Flight
Q-ToF	Quadrupole Time of Flight (Mass Spectrometer)
GC	Gas Chromatography
LC	Liquid Chromatography
MS	Mass Spectrometry
LC-MS	Liquid Chromatography - Mass Spectrometry
TIC	Total Ion Count
EIC	Extracted Ion Chromatogram
MPP	Mass Profiler Professional
ECC	Extracted Compound Chromatogram
MP	Mobile Phase
PPM	Parts per Million
PPT	Parts per Trillion
EI	Electron Ionisation
CI	Chemical Ionisation
ESI	Electrospray Ionisation
APPI	Atmospheric Pressure Photoionization
APCI	Atmospheric Pressure Chemical Ionization
PCA	Principal Component Analysis
QqQ	Triple Quadrupole Mass Analyser
Da	Daltons (weight)
Ca.	Approximately
RCF	Relative Centrifugal Force
rpm	Revolutions per minute of rotor
HCT	haematocrit
MDD	Major Depression Disorder

Acknowledgements

I would like to take this opportunity to thank my supervisors, Dr Nicholas Powles and Dr David Thompson. Thank you both for your help and support over the past years and the chance to do this research project. Nick, words cannot express how grateful I am for your invaluable patience, support and the great opportunities throughout my time. Thank you for helping me grow as a person, and as a researcher.

I am thankful to everyone at Analytical Innovations for the advice and incredible knowledge that you have all passed onto me. Another thank you goes to Forensic Testing Services for the assistance with my research. Thank you to those at Agilent Technologies who have given me endless help and support over the years. A huge thank you to Professor Sabine Bahn and Dr Jakub Tomasik for your expertise and access to your samples and laboratories. Without you and your team, this research wouldn't have been possible.

My sincerest thank you goes to my Mum and Dad for always being there for me. This wouldn't have ever happened without your words of encouragement motivating me to carry on. Thank you to my wonderful fiancé Thomas Davison for unwavering optimism, love and support. Thank you to all my wonderful friends and family who have helped me through my PhD journey and assuring me that I could do this.

1 Introduction to Metabolomics

1.1 What is Metabolomics?

Metabolomics is a new approach used in different fields of analytical science, such as forensic, food authentication and clinical applications [1]. It is defined as the 'quantitative measurement of the dynamic multiparametric metabolic response of living systems to pathophysiological stimuli or genetic modification' [2-3]. A more simplistic definition would be 'the comprehensive analytical approach for the study of all low-molecular-weight species (metabolites) present in a given biological system of interest' [4]. Low molecular weight species are those typically less than 1000 Da, with the purpose of a metabolomics experiment being observing the effect of both intrinsic and extrinsic factors on the metabolome [1,4-8]. The metabolites that collectively make up the metabolome are generally organic species such as fatty acids, amino acids, carbohydrates, lipids and vitamins [9].

Together with the other "omics" approaches, metabolomics provides a complete picture of a living organism's chemical and functional signature [10]. The 'omics' technology includes genomics, transcriptomics, proteomics and metabolomics. Each 'omic' is a study of specific areas; the genome, gene expression, protein expression and the metabolism, respectively [11]. The term metabolome was first described by Oliver *et al*. as the 'complete set of low molecular weight compounds present in a cell that are required for its maintenance, growth and normal function and contributes to the metabolic reactions of a cell in a particular physiological or development stage'; it is a collection of small molecules in cells, tissues, urine, plasma etc., that shows what is currently happening in the system [12-13]. Therefore, metabolomics represents the small molecules that can be objectively and quantitatively measured in biofluids. As metabolites are immediate down-stream products of protein/gene transcription and translation, metabolomics provides a clearer picture of the phenotype in a biological system in comparison to genomic and proteomics. The genome tends to show what might happen and therefore it is not easy to change where-as the metabolites are easy to

change through diet, supplements, drugs etc. Overall, this makes metabolomics highly complex, yet gives a fully comprehensive overview of the system being investigated.

Genomics and transcriptomics do not tend to use much mass spectrometry as other techniques are preferable when determining the gene sequencing efficiently [14]. With proteomics and metabolomics however, mass spectrometry has a very central role. Mass Spectrometry (MS) and Nuclear Magnetic Resonance (NMR) spectroscopy have both been used in metabolomics however NMR poses some limitations when it comes to sensitivity, ease of access, cost and how it only looks at the sum of compounds in the mixture; it does not work well with complex mixtures. NMR is often used for structural confirmation on simple samples, such as the final stages of structural identification on a biomarker alongside MS-MS. Over the past decades, mass spectrometry has become the technique that many would choose for an untargeted approach to characterise complex biological systems. The improvements in technology in the past few years have shown that large sample sets can be analysed due to the increased sensitivity of detection and faster chromatographic separation. Due to the increased sample load in a short space of time, data processing has become a crucial step that limits productivity and potentially the quality of interpretation in raw mass spectrometer data. Determining a workflow with software that can reliably and rapidly process and produce accurate information is essential for the quality of analytical results and interpretation [15-17]. Commonly used commercially available software includes Mass Profiler / Mass Profiler Professional by Agilent, Progenesis QI by Waters, Compound Discoverer by Thermo Fisher Scientific and Bruker Metaboscape. However, since metabolomics is a fast-progressing area, the development of freely available software tools is also on the increase [18-19]. Popular open-source software programs such as MetaboAnalyst, XCMS online, MZmine2 and MS-Dial are providing the advanced tools to manage, explore, process, and annotate the increasingly complex data generated from mass spectrometry tools [20-24]. Some of these packages can perform statistical analysis, such as MetaboAnalyst and XCMS online, though some do not have the ability to deconvolute the data, i.e., MetaboAnalyst [25-27].

The goals of metabolomic studies when using LC-MS are:

- To collect highly reproducible chromatographic separations with a retention time error of +/- 0.02 min
- To use a high-resolution instrument to collect highly accurate, exact mass spectra defined by 4 decimal place *m/z* values, meaning that possible empirical formulas can be deduced
- To observe multiple levels of bioinformatics and multivariate statistics – 2000 to 20,000 features are the ideal number to compare
- Develop a therapy, develop a diagnostic test, or obtain new biochemical information (potentially through the use of biomarkers).

There are two different approaches that can be taken when looking at metabolomic studies: targeted or untargeted (**Figure 1.1 and 1.2**, respectively) [28-29].

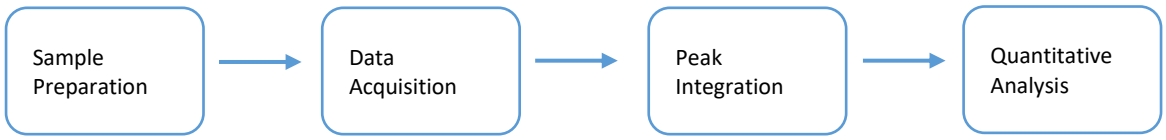

Figure 1-1. Targeted Metabolomics Workflow.

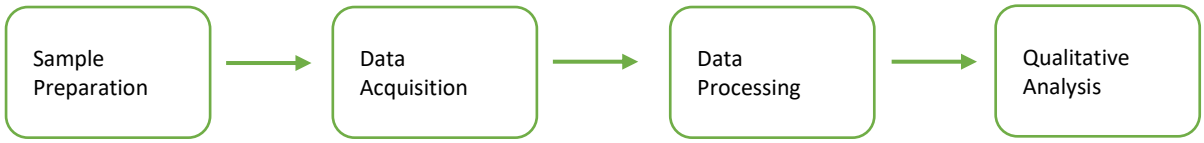

Figure 1-2. Untargeted Metabolomics Workflow.

1.1.2 **Targeted studies** are where a specific compound, or a small set of compounds can be quantitatively identified and examined. These are useful when the biological pathway is known to be affected by the investigated factor, and the targeted compound, or class of compounds, can be studied in detail with tailored preparation methods [28-29]. This typically uses triple quadrupole mass spectrometry (QQQ) which is not as data intensive as Q-ToF work and gives more accurate quantitation results, although it does make it more difficult to identify unknown metabolites given that it does not collect to 4 decimal places and so unknowns cannot have potential formulae deduced [30]. Targeted metabolomics is also known as metabolic profiling [31].

1.1.3 **Untargeted studies**, or metabolic profiling, use a compiled combination of high resolution, mass accuracy and chromatography to extract features. It aims to acquire and measure as many metabolites as possible, looking at any potential changes without focusing on the identification of every single feature [32-33]. It is used when it is not known if a biological pathway will be affected by factors (such as pH, storage or temperature) or not, and so as many small molecules/metabolites are profiled as possible to get a more holistic overview of the entire metabolome so the data can be investigated [31]. This type of experiment generates a large amount of data due to the vast amount of information gathered on the number of metabolites that may be present within a sample; many different sample classes can be captured in a single sample. Untargeted analysis is a useful starting point for investigations due to the ability to look at significant trends in the data using statistical workflows, however it is highly data intensive and is software dependent. It is possible to determine the identity of the compound(s) of interest, then further targeted studies can be conducted, and it may become a biomarker of interest for the factor that was investigated. Many different metabolites can be observed, and metabolic changes due to disease, environment or diet can be investigated over time [34]. Since untargeted metabolomics can be applied to a wide variety of different matrices and metabolites, the only limitation being the analytical instrument chosen, it is considered a true omics approach and a great starting point for metabolomics projects given the simple sample preparation [31].

1.2 Scientific Workflow

For untargeted metabolomics to be successful, a scientific workflow needs to be well structured and planned out. It should consist of several different stages, including sample collection/preparation, compound identification and targeted studies to ensure that high quality, reliable results can be obtained. **Figure 1.3** shows a generalised workflow for untargeted metabolomic studies.

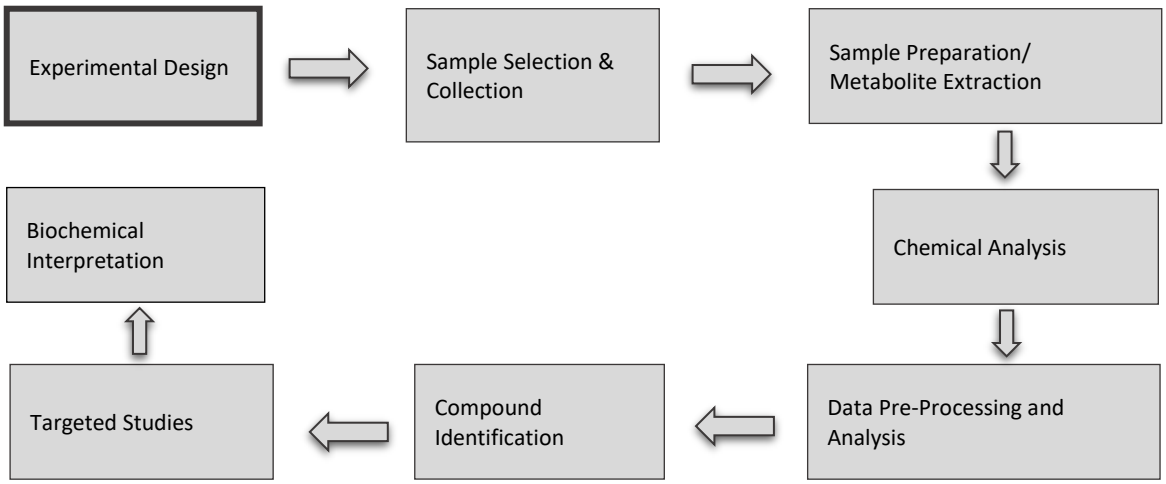

Figure 1-3. Generalised Untargeted metabolomics workflow

1.2.1 Experimental Design

The experimental design is the determining step of the study; if the initial setup of the experiment is flawed, then the results can be disregarded along with any interpretations made from this data. It is important that any potential variables in the study are controlled so that confidence can be had that the changes in the metabolic profiles are due to the factor that is being investigated, rather than a different uncontrolled variable [35]. To ensure the results are reliable, a minimum of three biological replicates are proposed, with five stated as preferable, where possible [36]. A biological replicate is where individual samples are prepared once; these are distinct samples that show biological variation [37]. A technical replicate is one sample prepared multiple times, however due to the high precision in LC-MS, this is not important unless you are developing a method and would like to determine instrument validation.

To monitor the drift throughout analysis, quality control (QC) samples should be added consistently throughout. These are usually pooled QC samples made up of aliquots of each individual sample adjusted with equal volumes that is used as a standard at the beginning, end and periodically throughout the sequence [38-39]. This gives the analyst an indication as to whether the results are reliable by identifying any potential changes in the QC throughout the analytical run. In most experiments, the LC column needs to be conditioned prior to analysis and so the QC sample is

injected till the results are consistent and reproducible. This is an ideal sample to condition the column as it consists of all the compounds in the analysis, and therefore you can be sure that the entire compound polarity extremes and mass ranges are covered.

The more replicates of samples in each sample set, the better. This is because the results are seen as more reliable if multiple replicates give the same answer and it reduces variation. There should also be a balanced number of samples from each group compared to one another and these should be fully randomised during the data acquisition, with a blinded sample list if possible.

1.2.2 Sample Selection and Collection

Sample selection is a critical step in the metabolomic workflow as all results depend on the suitability of the samples selected. The biological sample chosen depends on the aim of the project as different biological materials will be more suited to different analytical problems. There are plenty of biological samples available for experimental studies, though each one will be significantly different chemically [40]. For example, urine will have polar metabolites present within the short excretion window though the hair will store both the parent compound and its metabolites for months and years to come, respective to the length of the hair.

Sample collection is an important aspect of the metabolomic workflow as it can cause major changes to the results of a study, if not handled consistently. They should all be exposed to the same materials and temperatures etc., to prevent external factors influencing the results of metabolite profiles between samples. To minimalize metabolite degradation, samples should be stored at -80°C in the dark where possible, however this depends on numerous factors including sample matrix and length of time before sample preparation will occur [41].

1.2.3 Sample Preparation / Metabolite Extraction

Sample preparation and the extraction of metabolites is the next step of the scientific workflow, with the main purpose of releasing any metabolites present but removing interferences (i.e., proteins). The choice of the sample preparation method is important as it affects both the observed metabolite consistency and the biological interpretation [4]. This needs to be done in a consistent way across the samples so each sample can be compared directly to another using the proposed analytical technique [42]. Each different laboratory across the world will do their preparation differently, with different brands and/or grades of solvents, different consumables etc., all which may slightly affect the metabolite extraction [10]. For untargeted studies, the class of compounds is unknown and therefore a generalised extraction procedure is used to release as many small molecules for analysis, whilst removing as many large molecules as possible [43]. Once the experiment becomes targeted towards the compounds of interest, a more selective, tailored technique can be adopted for the compound groups being investigated. The sample preparation technique used is dependent on the matrix of the sample and the desired final sample type. The ideal method would be robust, reproducible and simple, but also be as non-selective as possible to ensure it covers the whole metabolomic consistency [10]. Ideally, it would also consist of a metabolism-quenching step to ensure the sample represents the true metabolome composition at the time of sampling [4]. Due to sample variability, using MS for data acquisition will not be the most robust technique; the analysis may retrieve different results on different days, even different results on different instruments. Therefore, in metabolomics studies the samples from all groups in a study should be extracted and analysed at the same time, keeping conditions as consistent as possible [10].

1.2.4 Chemical Analysis

Chemical analysis is the next step, where various analytical techniques have been used to gain insight into the metabolite profile. The different techniques vary from spectroscopic methods to mass spectrometry and other chromatographic techniques. An ideal scenario would involve simple, or even no, sample preparation, be rapid and have a high sensitivity equal for all compound classes that

may be available in the sample. It would ideally produce reproducible results with enough molecular information for metabolite identification but also be inexpensive and non-destructive to allow the sample to undergo further investigation when needed [44].

The two most common analytical techniques used for chemical analysis in metabolomic studies are NMR and MS due to the reproducibility and high precision, respectively.

1.2.4.1 Nuclear Magnetic Resonance

NMR is a non-destructive spectroscopic technique, producing highly reproducible results that are rich in molecular information, leading to identification of metabolites [5-6,45]. It is non-selective and therefore can detect multiple different compound classes, with little sample preparation. The analysis is relatively fast, meaning it is a high throughput technique desired for metabolomic studies. However, the instrumentation is expensive, and the low sensitivity means it can only detect a limited number of metabolites during the analysis. It is also a technique that shows everything in the sample all at once and therefore with more than one compound present, it will be very difficult to determine the structures of each separate metabolite. **Figure 1.4** shows a very simplified version of 3 separate compound NMR spectra. It then shows what the spectra would look like if these 3 simple sample spectra were all in one sample; it becomes very complicated and difficult to analyse.

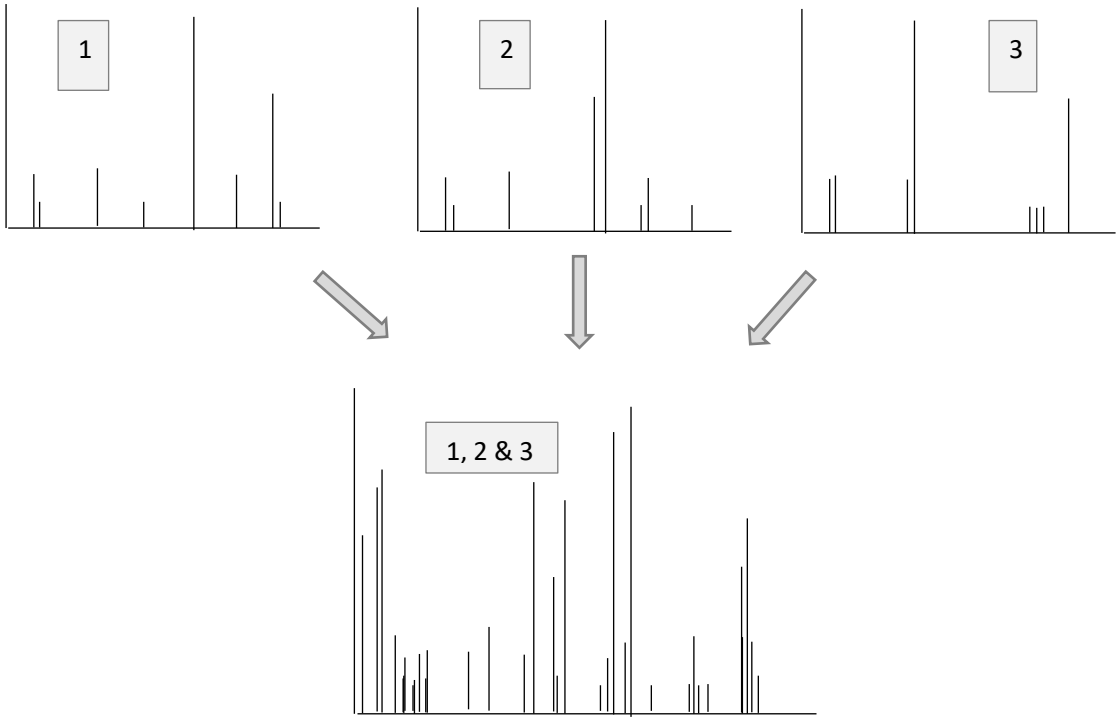

Figure 1-4. Simplified diagrams of NMR spectra showing what the spectra would look like if three compounds were analysed in one sample. It would be incredibly difficult to deduce any important information from the spectra.

1.2.4.2 Liquid Chromatography – Mass Spectrometry

Liquid Chromatography – Mass spectrometry (LC-MS) is used to analyse non-volatile analytes. The samples are usually diluted in an appropriate solvent before being introduced to the LC where analytes are separated based on their polarity. LC is beneficial due to the large variation of separations systems that can be used, such as normal phase (silica based), reversed phase (hydrophobic to non-polar molecules), ion exchange and hydrophilic interaction liquid ion chromatography (HILIC).

Metabolites are then ionised in the source, where Electrospray Ionisation (ESI) is typically used. ESI is a relatively soft, sensitive ionisation technique that has brought many new features to mass spectrometry, including the ability to use it with HPLC, and particularly for the identification of a range of analytes, from small molecules to proteins [46]. ESI is described as a soft ionisation technique since it provides little fragmentation, however it can produce ions from non-volatile, thermally liable compounds with multiple charged states [47]. The source, as seen in **Figure 1.5**, works by forcing the

analyte solution through a very fine capillary into an electric field creating charged droplets. The droplets will each possess a positive or negative charge, depending on the polarity of the capillary. The drying gas causes the droplets to decrease in size by evaporating the solvent, consequently increasing the charge density. As the droplet size decreases, repulsive forces between the charged ions increases and eventually, many smaller droplets are formed [48]. These repulsive forces are known as Coulomb force, where the surface tension attempts to keep the shape of the droplet but the charges on the surface of the droplet are repelling one another. The Coulomb explosion, or Coulomb fission, occurs when the Coulomb force becoming greater than the surface tension [14]. This causes the ions at the surface of the droplet to be released into the gaseous phase, allowing them to pass through the skimmer cone and onto the mass analyser, where it's mass to charge ratio is determined [14].

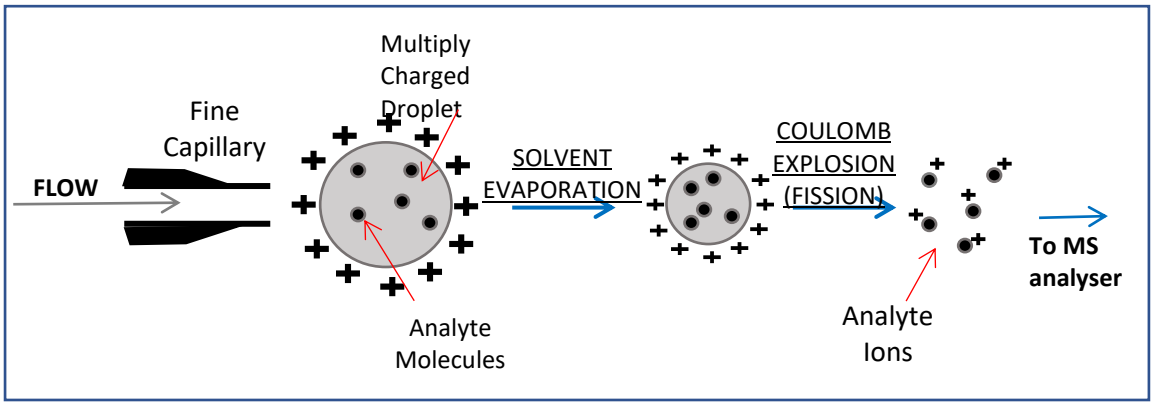

Figure 1-5. Simplified diagram of how ESI works.

Adapted from deHoffman and Stroobant [14].

ESI is often used to analyse biomolecules where the charge is generated from protonation of a basic site, $[M+H]^+$, or deprotonation of the acid group, $[M-H]^-$. Other adducts can be formed, such as a sodium, potassium or ammonium adduct when the compound is analysed in positive ionisation mode, i.e., $[M+Na]^+$. In negative mode, other adducts include the chlorination of an acidic site, or more complicated adducts such as formic acid, $[M+CH_3COOH]^-$. Since ESI produces little fragmentation, limited structural information can be obtained, although this means that the parent molecular ion is almost always observed [48]. This can be overcome by using other techniques such as tandem mass spectrometers (LC-MS/MS) which provide structural information.

The Quadrupole – Time of Flight (Q-ToF) mass spectrometer is a popular instrument. It is a tandem mass analyser is useful for determining structural information about unknown compounds. The ToF can screen for all possible compounds where-as the quadrupoles can be fragmenting and determining more structural information about any specific compounds at a given RT. Due to its sensitivity and high resolution, Q-Tof is considered a useful and powerful metabolomics tool [49-51].

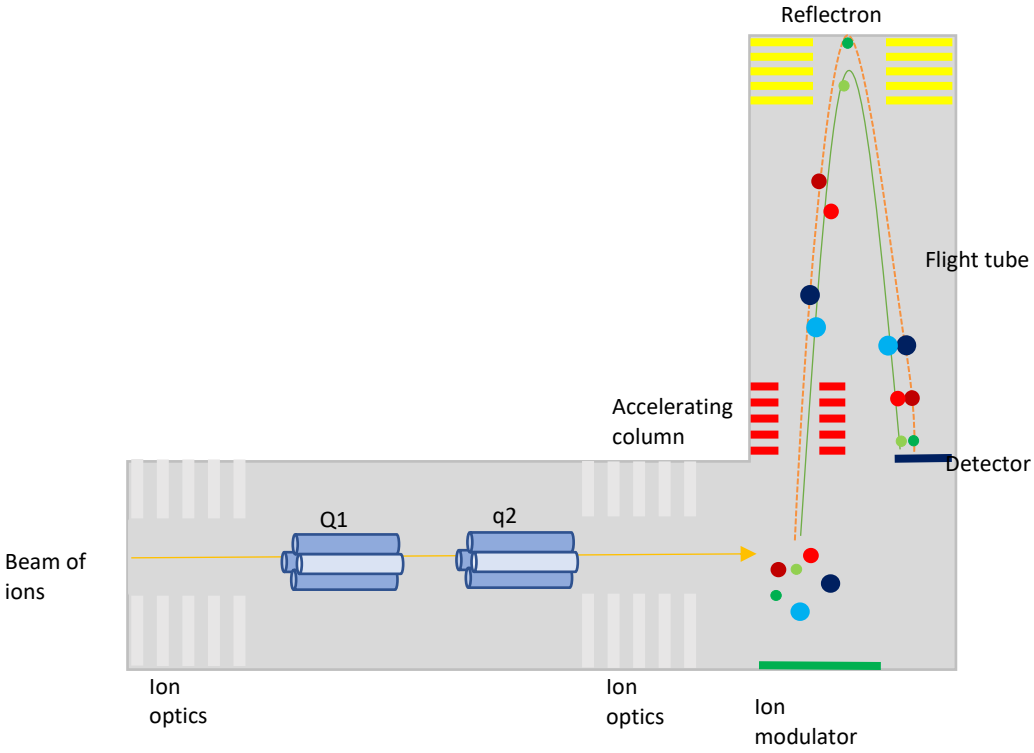

Figure 1-6. Diagram representative of a Q-ToF mass spectrometer. Q-ToF works by accelerating a beam of ions through the ion optics into Q1 (a mass filter that allows only ions of a certain m/z to be transmitted through to q2 (collision cell for fragmentation). Ions are pulsed and they accelerate into the flight tube. The ions are separated depending on m/z due to the difference in time taken to reach the detector. The reflectron is used to extend the flight path.

Diagram based on the diagram from Hoffman and Stroobant, and Chernushevich et al [14,52].

Overall, Mass Spectrometry (MS) is more sensitive than NMR, meaning it can detect a larger number of metabolites. MS can be used either by direct infusion where it is infused directly into the detector, or it can first go through chromatographic separation before being detected. Direct infusion can, however, lead to ion suppression or enhancement which affects the reliability of the analysis and therefore, chromatographic separation of the compounds is advised in most cases [7,28]. MS is a destructive technique; however, the results are reproducible and give enough information to identify each metabolite. It is also relatively non-selective and can detect multiple different compound

classes, but this is highly dependent on the volatility and polarity of each compound class. As the analytes are easily separated with chromatography columns, it means that the separate compounds can be identified easier than it would be using NMR. For all these reasons, this project uses MS and not NMR.

1.2.5 Pre-Processing and Analysis of Data

The main steps for data analysis in metabolomics are pre-processing, pre-treatment, processing, post-processing, validation, and interpretation [53]. Such information-rich, complex raw LC-MS data needs extensive processing to interpret the results properly [33]. With untargeted metabolomics there will be thousands of features detected, many of which will be not biologically interesting for many reasons. They may represent background ions from sample processing, or multiple different adducts, isotopes and/or fragmentation from the same ion, all of which do not add anything to the data analysis if they are not from biologically interesting parent ions [54]. In addition, the software used for feature detection can lead to false positives or negatives if incorrect filtering parameters are used, or imperfect integration may mean that noise signals have been integrated and falsely identified as features. If this step isn't carried out correctly, this can consequently affect the statistical analysis. If features of high quality are filtered out, they will not be considered as features for a potential biomarker when doing univariate statistical analysis. Conversely, insufficient filtering of noise may lead to false positives. Therefore, pre-processing of data must include filtering methods that remove noise before further investigation of the data. The process of how this is done depends on the personal preferences of the analyst. Programs such as MetaboAnalyst have valuable methods for performing this however many of them rely on the default cut-offs for filtering of such features, such as basing it upon the mean/median values, instead of determining appropriate thresholds based upon the specific dataset. This may lead to inappropriate filtering out of biologically interesting features or lead to incorrect 'missing values' in samples [54].

Chemometrics is about extracting the maximum information from the experiment to find discriminating features that can also show potential classification patterns in the dataset. There are a growing number of different software tools used for the data processing and analysis, however each tool has varied level of 'useful-ness' within the field. Each piece of software is created with different aims in mind and each one has different ranges of depth to the analysis. Some basic packages have been created by practitioners who have little software development experience but know a lot about the different forms of analysis needed in metabolomics. There are other packages that are more limited in the types of analysis but have more visual tools due to the more experienced developers involved. There are many instrument manufacturers who develop generic processing tools to meet a wide range of customer needs, such as Agilent Profinder. Waters Q-ToF's have their own software version called UNIFI which has some processing tools included in the package, however a lot of these vender software packages are fairly limited in the analysis they can conduct. Therefore, it is often necessary to create your own niche workflow by combining a sequence of different tools depending on your desired outcome and usage requirements.

Due to the large amounts of data that's produced, it needs to be pre-processed to convert it into a user-friendly format, ready for the data analysis. The end goal of the analysis needs to be determined before deciding on the software to use. The data pre-processing can sometimes be carried out using the same software as the statistical analysis, depending on the software being used. Other times, one piece of software may need to be used in conjunction with another depending on several factors, such as the data format, statistical analysis required and the resources available. Some overall types of software include:

- Data pre-processing software (peak picking, de-convolution, alignment of peaks/signals, smoothing, filtering, peak isotopes etc.)
- Molecular structure identification software (annotation software, matches processed peaks with known databases to present evidence for the presence of a specific compound, or predict a chemical structure)

- Statistical analysis software (univariate or multivariate analysis, identification, and comparison of features across the sample set, e.g., ANOVA, PCA, PLS-DA etc.)
- Functional analysis software (uses annotated peaks and their properties to infer biological interpretation, e.g., Analysing peak intensities across multiple samples to determine the changes in metabolic pathways)
- Chemical property prediction software (building libraries of chemical properties that can be measured to monitor in metabolomics analysis, i.e., *m/z*, retention time, chemical shift, relative intensity, fragments etc.)
- Metabolic modelling software (development of kinetic/flux models of metabolic networks to enable prediction of metabolic fluxes from metabolomics measurements) [55].

Different types of software each have a different type of programming language. The most common types are C-family (Java, C#, C, C++, PHP), Python and R. These all have specific coding language and so when picking the software to use, it is important to take into consideration the resources available for learning how to use each one sufficiently. This can be especially difficult when programs such as R, MatLab and Python are used due to their niche coding techniques rather than typical looking 'easy to navigate' software many analysts will be used to. The training is important; if the wrong code is recorded, then accurate results will not be obtained when used by other scientists in the future. Full training on programming needs to be conducted to avoid making some of the common mistakes. The import and export options also need to be considered as they are tailored towards their own needs, however there is usually a generic format such as '.cef' or '.csv' files. This is something to bear in mind if data will be transferred between different analysis tools. It is also important to keep a copy of the raw data in the instruments proprietary format as most of the generic exported files do not capture the entirety of the data.

1.2.5.1 Statistical Analysis

In metabolomics, both univariate and multivariate statistics are used to explore the data for different trends and variables of statistical significance. Untargeted metabolomics can be difficult given the thousands of features that need to be reviewed and therefore 'plug and play' software has been designed by various different companies to aid the analysis. Such platforms are freely available (e.g., XCMS online, MetaboAnalyst), while others are commercial products (e.g., Progenesis, Mass Profiler Professional [MPP]). At the higher level, this software has three basic functions: feature detection, feature alignment and statistical analysis. This allows a feature to be detected and compared across the sample run. Each performance, however, is parameter dependent. This means that even though the software does look simple to use, a substantial understanding of each parameter is required otherwise the processing will not be done properly, leading to false findings and unreliable results. Each different piece of software has its own limitations, and the desired outcome of the research means the optimal result may be needed to be performed on more than one piece of software. Due to the different parameters and the way each piece of software works, the same analysis may lead to different outcomes depending on which piece of software is used. However, if the parameters are set correctly, the same outcome should be produced each time [7,56]. One aim of this project is to determine whether the same results are obtained from each piece of software.

1.2.5.1.1 Univariate Statistics

Univariate analysis is where only one variable is observed at a time. It compares each individual compound between different sample sets to see whether there is a significant difference between the sample sets, or not [7,57]. Such tests include t-tests, ANOVA, Welch's test etc. They help to identify compounds that show significant differences across data to discriminate between different sample groups. Such compounds have potential to be a biomarker for the factor under investigation. When doing high throughput data analysis, multiple testing corrections can be applied to correct for potential 'flukes' in the raw *p*-value data as even when true differences are present, there are still chances that false positives will contaminate the data and significant findings may be missed. There

are different ways to correct, or adjust *p*-values for multiple testing such as Bonferroni corrections or Benjamini-Hochberg False Discovery Rate procedures.

1.2.5.1.2 Multivariate Statistics

Lots of different types of multivariate statistics are available; however, these tend to depend on the software you are using. For most experiments, advanced multivariate statistical analysis software is often required due to the large amount of data collected. Multivariate analysis (MVA) compares data from multiple different sample groups to visualise which features are significant to each sample, identify the potential patterns and how they vary throughout an experiment sample set, with more variable features than samples [55]. Each feature is a potential metabolite. There are two types of multivariate statistical analyses: supervised and unsupervised. Partial Least Squares Discriminant Analysis (PLS-DA) is the most popular type of supervised MVA, with the purpose of making predictions about the samples with unknown sample sets, based upon variables in the samples with known sample sets [29]. Principal Component Analysis (PCA) is the most popular type of unsupervised MVA, it explores and discovers any potential trends/patterns within the data, without prior labelling of the sample groups [58]. It is one of the most useful and powerful tools within chemometrics, hence why it is the most common multivariate statistic used [59]. It condenses multiple variables from large datasets into smaller Principal Components (PCs). There are no assumptions made about the distribution of the data and therefore is a general method used for data reduction [60].

Principal Component Analysis works by finding the lowest possible number of dimensions needed to describe the largest data variation. It is used to explore interrelations across the data to see if there are any potential classifications, along with identifying outliers from the dataset [53,61]. If the entities were all plotted on a single graph, the longest line that could possibly be drawn would be the eigenvector with the largest eigenvalue and 1st principal component. The 2nd component is the eigenvector that has the next largest eigenvalue, orthogonal to the 1st principal component. The 3rd component is the one with the next largest eigenvalue that is orthogonal to the 2nd component [62]. An

example graph is shown in **Figure 1.7** where the principal component (first component) is represented as 'u' and the second principal component is represented as 'v'.

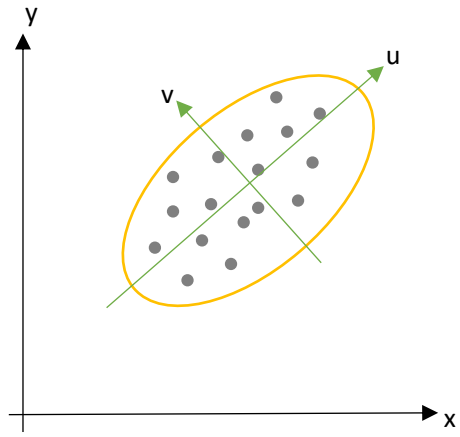

Figure 1-7. First two components on a 2D PCA plot, example graph.

The longest eigenvector is 'u', the principal component whereas 'v' is orthogonal to 'u' and is therefore the second component with the second longest eigenvector.

PCA generates scores which provide the coordinates of samples in space to allow visualisation of similarities and differences in samples and/or sample groups. It also generates loading values which show the amount each variable contributes to the PCs [63]. There are different ways that these plots can be visualised, and different parameters can be chosen when using different types of software. Sometimes the data can be standardised, which allows the scale to be converted to a measurement relative to each data group. The precision within the data can also be examined by looking at the QC samples and seeing whether they are tightly clustered or if they're more spread. QC samples assess the data quality and so if the dataset is reliable and has high precision, then the QC samples will be tightly clustered and sit in the middle of all the other groups. If the QC samples are more spread, this suggests that there has been an instrumental drift which has affected the results [53].

Figure 1.8 shows a simple, yet ideal PCA scores plot where all the QC samples are tightly grouped in the middle of all the sample groups. They should be injected from the same solution in the analytical run, hence why they are so tightly grouped; if the instrument is robust, each QC should be identical to another. Any deviations from this would indicate instrumental drift or an error, therefore showing that it should be investigated in more depth. There will always be small changes across an analytical

run, hence why tightly clustered rather than identical scores clustered points are accepted. Arrow 1 on the **Figure 1.8** represents the instrumental drift seen in the QC samples. Arrow 2 shows the difference between group 1 and group 2, and as arrow 2 is longer than arrow 1, it can be assumed that the differences between the sample groups is due to chemical differences seen by the instrument rather than instrumental instability. Group 2 and 3 on the PCA scores plot have less separation than to group 1, however due to the consistent QC injections showing tight clustering, and the grouping of all the groups it can be assumed the differences are genuine chemical differences in the composition. PCA plots offer an alternative means for visualising data to identify such patterns that may not be obvious from the raw data. However, to investigate these trends further, the raw data needs to be examined [64].

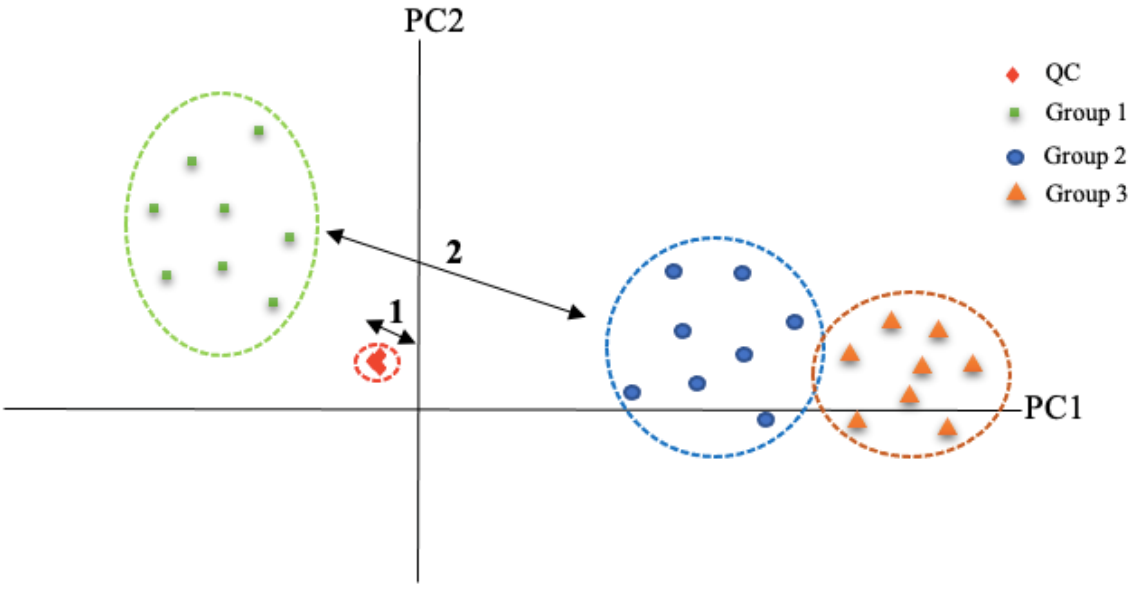

Figure 1-8. PCA Scores Plot.

Arrow 1 to indicate QC spread for instrumental drift, Arrow 2 indicates separation of two groups.

If the PCA plot shows good separation in different sample groups, they can be used going forward to give indication of potential biomarkers by looking at the features that are responsible for the separation. If the PCA plot does not show significant separation, then other statistical tests can be used to identify which features are most discriminating and could therefore be potential biomarkers.

Some other types of multivariate analysis that can be performed include cluster analysis and volcano plots:

Cluster analysis is an independent method that groups together variables based on their similarities or differences, forming 'clusters' [65]. The variables in one cluster should be more similar to each other than to variables in alternate clusters; the intra-cluster distance should be small whereas the inter-cluster distance should ideally be large. This type of cluster analysis can identify patterns in the data and helps to understand the data distribution.

Volcano Plots are a type of scatterplot that shows statistical significance (*p*-value) plotted against the magnitude of change (fold change). It is a visual way to determine features with a large fold change that are also statistically significant, identifying the most meaningful features that change between datasets [66].

1.2.6 Compound Identification, Targeted studies and Biochemical Interpretation

Once the statistical analysis is complete, any features of significant difference need to be identified, then targeted studies can be carried out to focus on just these compounds of interest [7,41]. Reliable identification is often said to be one of the most difficult steps in metabolomics. Target compounds are usually investigated using Q-ToF and once an annotation is given to the target, a standard can be purchased (where possible) and the analysis can be transferred over to a triple quadrupole mass analyser, QqQ, for confirmatory identification. This technique is targeted as it only focuses on the *m/z* values requested, at the retention time you specify. It is a tandem MS method where the first and third quadrupoles act just like mass filters, but the second quadrupole causes fragmentation of the targeted analyte by colliding with a neutral gas, typically argon or nitrogen. QqQ is a nominal (unit) mass analyser and so only masses to 1 d.p. are used but structural information can be obtained through some of the different modes available.

Once the identity of said compound has been confirmed, an analytical standard of known purity can be purchased and further targeted studies can be carried out, confirming the statistical significance between groups and obtain further quantitative information [7]. To complete the study, metabolic pathways involved in biosynthesis and degradation of these significant compounds need to be studied in further detail, through interpretation of the results [41,58]. The validation of the use of these novel biomarkers for the investigated factor can then be completed [67].

Figure 1.9 shows the pathway to creating the potential biomarkers. The validation step takes the longest time to complete due to the vigorous tests to be completed to ensure it is a validated biomarker. It is estimated that only 1% of 'hits' are followed through to be confirmed biomarkers. The objectives of such biomarkers are that they are consistent, reliable, and quantifiable; these objectives are what make it more difficult to validate each potential biomarker. They should be accurate at defining the 'disease' that they are a marker for and are defined from observational research.

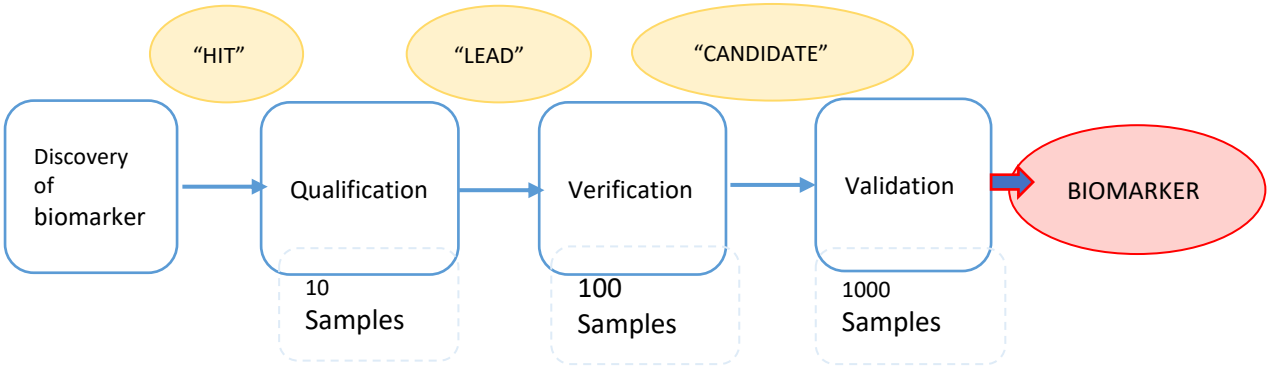

Figure 1-9. Pathway for creating biomarkers.

1.3 Software Packages for data analysis and comparison

To compare LC-MS data sets, the data needs to be extracted using retention time alignment and accurate mass; each retention time *and* mass that are aligned is named as a feature or entity [68].

1.3.1 Mass Profinder (Agilent)

Data extraction software includes MassHunter Profinder; a standalone software program that has been developed for feature extraction (including batch feature extraction) and alignment of data. It is compatible with data collected using Gas Chromatography-Mass Spectrometry (GC-MS), Liquid Chromatography-Mass Spectrometry (LC-MS), Capillary Electrophoresis-Mass Spectrometry (CE-MS), when coupled to a ToF or Q-ToF instrument. It also works for GC-MS nominal mass data. Good quality data needs to be collected to produce good results in Profinder, ideally this includes Gaussian peak shapes that have good chromatographic separation, and each sample set should be highly reproducible. Bad quality raw data may lead to increased false peak detection, incorrect identifications and/or completely missed peaks, hence leading to incorrect potential biomarker analysis. Feature extraction within data is an important step is ensuring that the number of false positives and negatives are kept to a minimum. This step also significantly reduces data file size by removing information that is not useful to the user [69].

Profinder can be used in both a targeted and non-targeted way for batch feature extraction. For non-targeted analysis, either 'molecular feature extraction' (MFE) or 'recursive feature extraction' (RFE) can be used. MFE extracts the feature and combines the different charged states, isotopes and adducts in one group with an assigned neutral mass. A sum of all the ions associated to the neutral mass is then put onto a compound chromatogram [70]. With RFE, the user can set different thresholds for batch molecular feature extraction, which repeats each feature extraction across the whole batch of samples. The files are then re-extracted using the additional information given which reduces the number of compounds that are missed. Once the features have been re-extracted into Profinder, a list of compound groups found in the samples is displayed. One compound can be selected at a time

so that its details can be displayed as a chromatogram/spectrum or in a table and the ability to do this allows each compound to be checked, ensuring the correct filtering parameters have been used and that each peak is “real”. It is also possible to manually integrate peaks that have been missed, or to delete compounds that aren’t suitable. Should the filtering parameters that have been used not give suitable results, the extraction can be repeated using revised parameters.

Profinder can also be used as more of a visual way to check for potential biomarkers if there are not too many samples. An example of two features in Profinder can be seen in **Figure 1.10**. The sample names can be seen in the top line of each chromatogram, along with the *m/z* of the feature shown. The scale can be linked if this is something of use, though it can then make it difficult to see the peak shape in each sample. The chromatograms show the Extracted Ion Chromatogram, EIC, of the selected feature across all samples, in which you can see whether it is present in every sample or missing in some. The features shown in **Figure 1.10** are some that increase and decrease overtime. This example shows samples that have undergone microwave digestion across multiple different temperatures; the temperature is highlighted on each chromatogram. These particular features shown by the *m/z* are seen to decrease as the temperature increases, which highlighted these features as useful information for optimal microwave temperatures in this project. Profinder is a useful tool when looking for potential biomarkers in a small selection of samples; however it can be very time consuming when looking at more than 50 features, as well as checking the integration for every feature and sample.

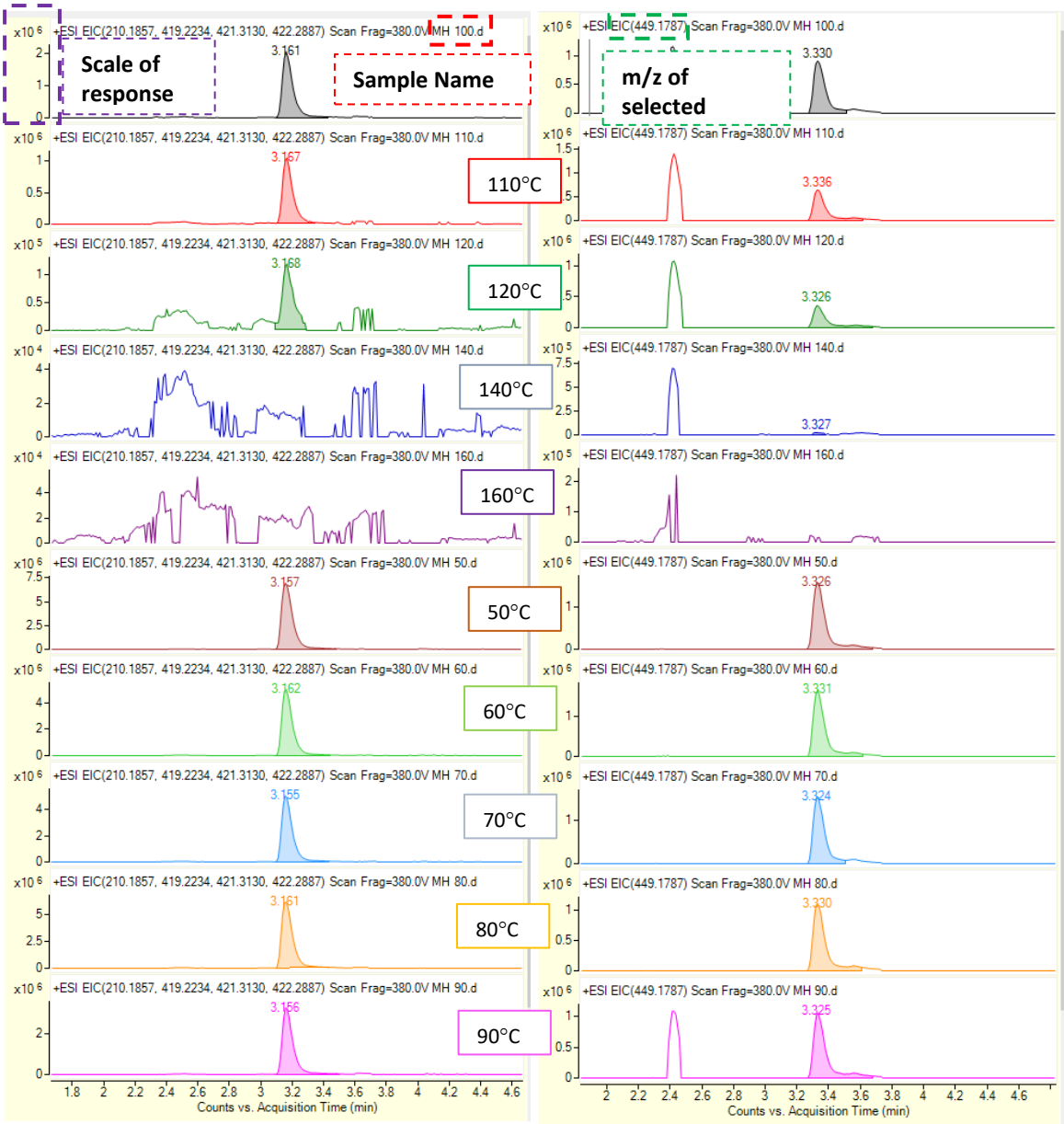

Figure 1-10. Screenshots from MassHunter Profinder 10.0 showing the EIC diagrams.

This shows how the software can be useful for seeing trends across datafiles. It is not possible to change the order of chromatograms, hence the miss-matched order. This example shows that these compounds decrease in abundance as temperature increases which is a useful trend to know as they are not present in samples above 120 degC.

1.3.2 MassHunter Profiler (Agilent)

Data extraction software also includes MassHunter Profiler (Agilent). This works in a very similar was to Profinder, however it can only be used for group one vs group two, or disease vs control. This software gives a table showing a feature summary, with the option for graphs showing entities that are unique to only one group. It shows a comparison of each feature against the control group, with a log score for the fold change. This is useful for seeing potential biomarkers, or those features that

are causing the largest differences between the groups. It is also possible to export the features for identification using databases (e.g. NIST) or personal preference of libraries.

1.3.3 Mass Profiler Professional

Once the data extraction files have been checked and deconvoluted in Mass Profinder, Mass Profiler or MassHunter Qualitative Analysis, the data can be exported as '.cef files' to compress the file to easily transfer it into Mass Profiler Professional (MPP). This is a powerful chemometrics platform which is specially designed to handle highly complex mass spectrometer data, giving the options for statistical analysis and visualisation tools. It can be used for any mass spectrometry based differential analysis that has two or more sample groups (and/or variables) and is compatible with GC-MS, LC-MS, CE-MS and Inductively Coupled Plasma-Mass Spectrometry (ICP-MS) which makes it a great tool for use in metabolomic studies. It also has an automated 'sample class predictor' for qualitative analysis of unknown samples across many different applications [71]. Like in Profinder, the user can set thresholds throughout for different filters and this can be used in conjunction with each other; lenient filters can be used in Profinder to create the data files and then stricter filtering can be applied in MPP. It is possible to filter the results by frequency or by flags which is useful as it allows 'one hit wonders' to be filtered out across the repeat sample sets. Because the data extraction is completed first in another type of software, the results given by MPP depend on the quality of the data extraction data. Good quality data and data extraction should give reliable statistical results in MPP.

As an example, the changes occurring in an herbal remedy sample over the period of 3 months on stability trials. There are 14 samples; 1-4 represents 1 month, 5-6 is 28 days in use, 7-10 is 2 months and 11-14 represents 3 months. 'In use' means that it is an opened bottle of the herbal remedy, the other samples are all unopened bottles. Each horizontal line represents a different entity, and the different colours show the abundance of that particular entity with the lowest abundant entities represented as a blue line and the highest abundance shown as a red line. The large number of

entities found in these samples are making the graphics look very complex. The type of graphic image shown in **Figure 1.11** can look very complicated due to the large number of compounds present in each sample.

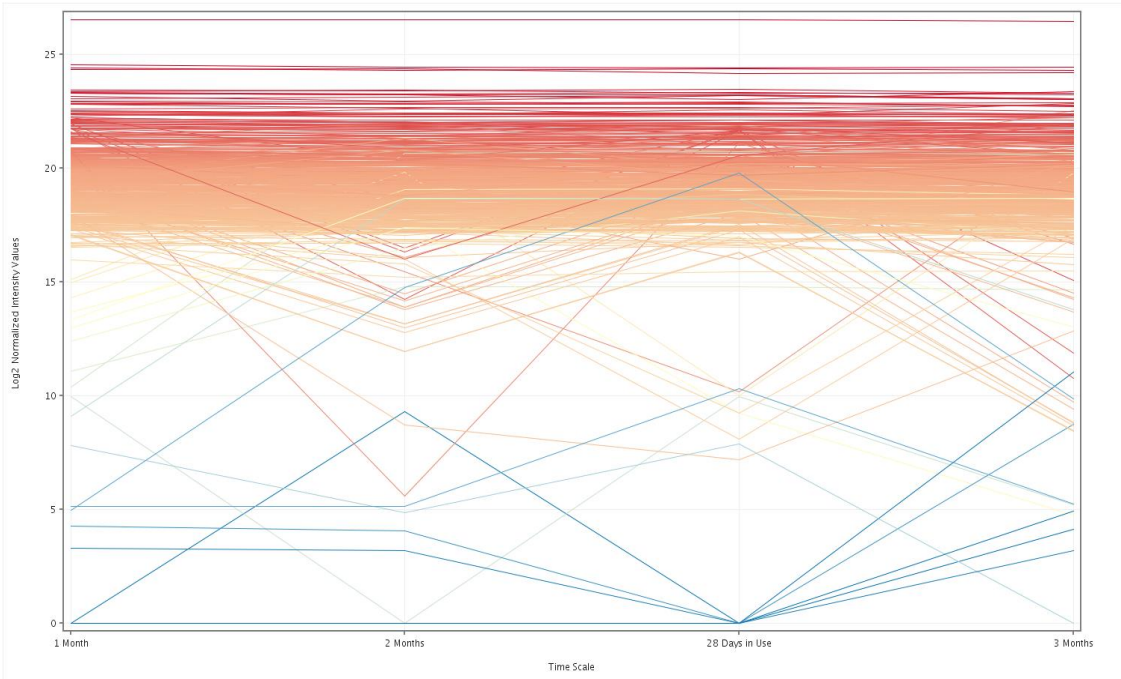

Figure 1-11. The abundance of each entity across all compounds, grouped by sample group type. Each line represents a single entity, with blue showing the lower abundant entities and red showing those of highest abundance. The lines reaching 0 show that the entity is not present in that sample group, or has a very low abundance and those across the top show those that are present in high abundance in the sample groups. It is a great way to see trends in the data.

There is a filter available in which you display only a few selected compounds. In **Figure 1.12**, only compounds with a fold change of 8.0 or above, in reference to '28 days in Use' initial time scale group have been chosen. This is a useful visualisation tool that can show trends in the data quickly by eye. Any obvious issues with stability will be shown by these lines as the compounds in the sample will change dramatically, showing a huge increase or decrease in some entities. It is important to note that when the line touches the x-axis, this means that a value is absent. However, the 'missing' value may be due to the abundance of this entity falling below the peak height cut-off parameter used for the initial data extraction or the parameters selected in MPP. Of course, it is also possible that the entity is completely missing from those samples also.

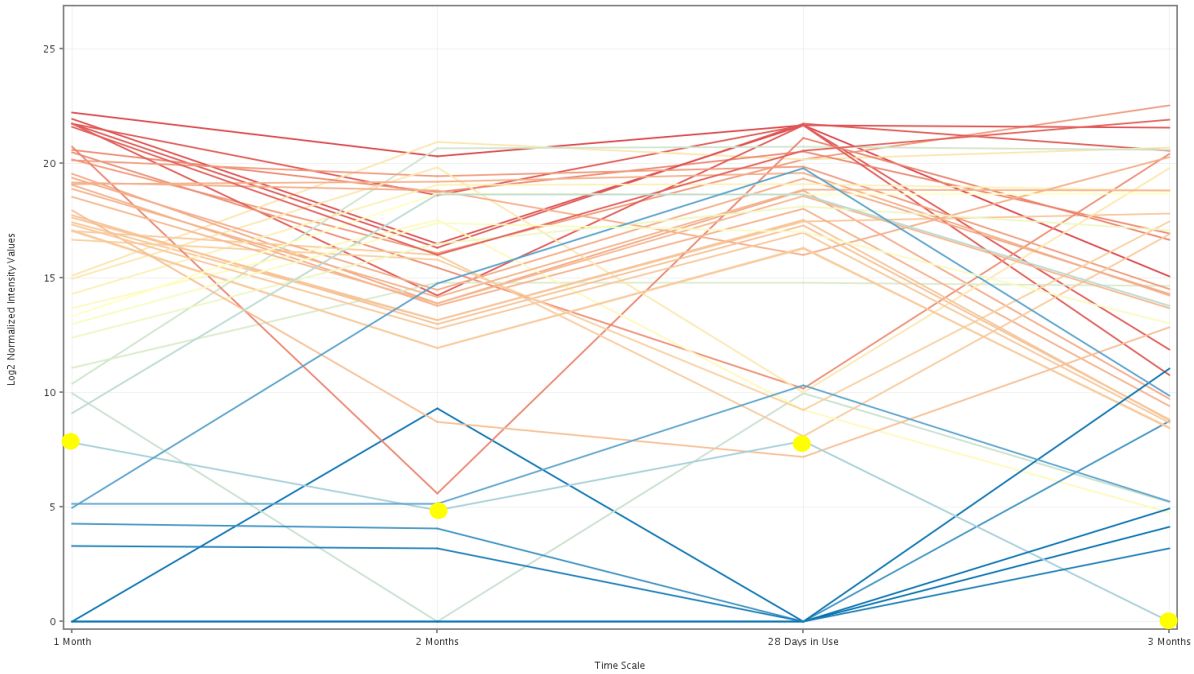

Figure 1-12. Filtered entity list based on fold change of 8.0, in reference to '28 Days in Use' time scale group.

Each continuous line represents a single entity across all samples, averaged by group type. The yellow dots are highlighting one entity and showing the difference in normalised abundance across the different sample groups. The sample groups are not in order, though it is possible to visualise that the abundance of this entity creeps down over time till it reaches the lower cut off in '3 months' sample group. This plot highlights the trends in entities across the groups.

In **Figure 1.12**, there is one feature line highlighted by the yellow dots to show why these plots can be useful. These yellow dots show the normalised abundance at each of these time scales though unfortunately it is not possible to change the order of the groups and therefore, they do not go in time order. It is still easy to see that the highlighted feature is highest at '28 days in use', then of similar abundance in the '1 month' group. It then decreases in 2 months and further decreases to near 0 at 3 months. This identifies that the selected feature possibly degrades overtime, showing that this compound in the sample is unstable.

1.3.3.1 Principal Component Analysis Plots

The three-dimensional principal component analysis (3D-PCA) plot is one of the most controversial features that MPP produces. It allows quick and easy data visualisation though can force differences on the 3rd component (z-axis) that may make samples look more different than they are. It can be useful for seeing if there is further grouping between samples other than the initial 2 components. A

two-dimensional principal component analysis plot (2D-PCA) is the most widely used plot in metabolomics which only shows the first two components on the x- and y-axes.

When the selected files are loaded into MPP, each data file contains information on each entity present in each sample. Each entity is characterised by its retention time, mass and the ion types present therefore, it can detect whether the same entity is present in more than one sample. Ion abundance is recorded for each entity, therefore up or down regulated entities can also be examined. Good reproducibility is shown by tight clustering within each sample group with the pooled QC samples in the middle of all the other samples. This would also show that there is no instrumental bias and that the instrument is working well.

1.3.3.2 Venn Diagram

Another useful feature of MPP is to display the as a Venn diagram, allowing entities to be identified as similar to, or different, to a particular group. This can be a useful tool to find entities that are in only one group and therefore could be used as an identifying biomarker for that particular group. To do this kind of interpretation, the sample grouping is inputted into the software and then the entities are filtered either by frequency, or by flags. This filter allows the user to select how many samples in the group an entity must be in, or if it can be accepted though it is potentially missing in a few samples. Below, **Figure 1.13** shows an example of a Venn diagram with 3 sample groups, labelled A, B and C. Overlapping segments show entities that are common to the segments, those not overlapping are entities that are unique to that group. The entity lists can be exported and identified. This makes it a great tool for identifying potential biomarkers in metabolomic analysis.

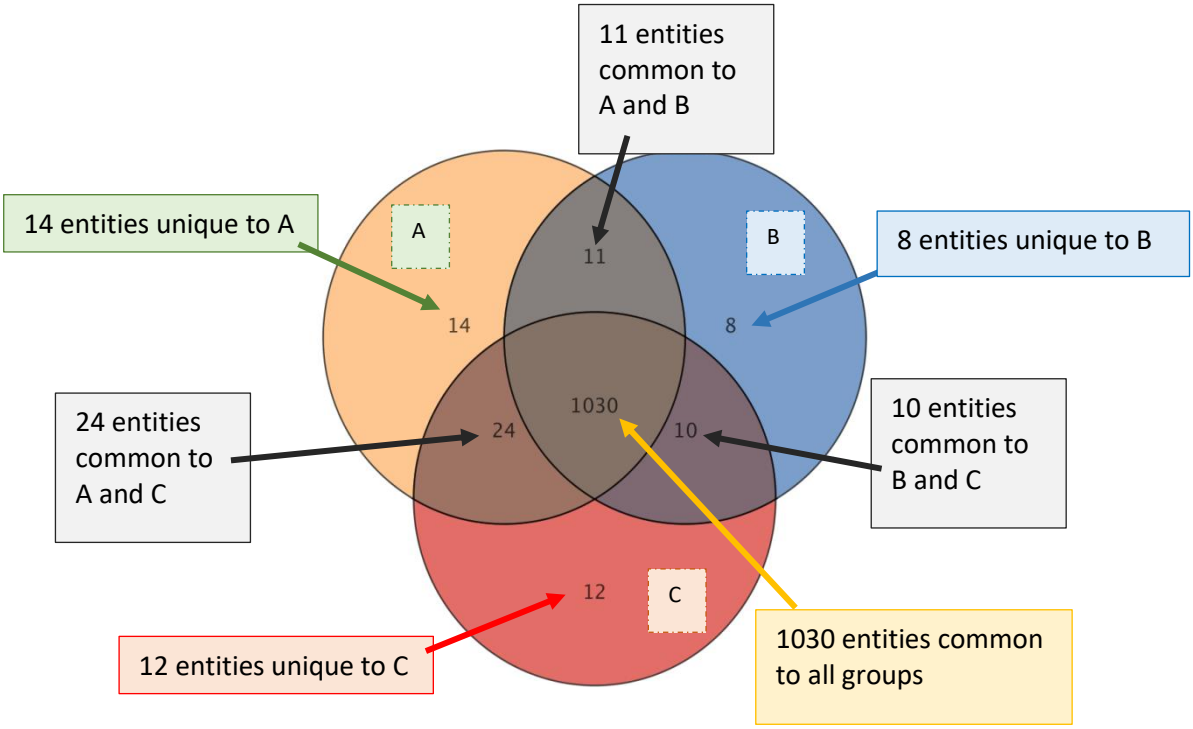

Figure 1-13. Example of a Venn diagram showing the entities common and/or unique to sample groups. The overlapped circles show features similar to only the groups overlapping. Therefore, it is possible to visualise features unique to sample group, or features similar in two sample groups or three sample groups.

1.3.3.3 Clustering

Another useful tool in MPP is clustering, there are various different ways to cluster your samples however hierarchical clustering is a great way to visually see any overall similarities or differences between samples. **Figure 1.14** shows a diagram for hierarchical clustering on 3 different batches of an herbal remedy, and a QC sample. It shows which samples are most similar to each other, then confirms that the intra-batch samples are most similar to each other than to those of a different batch.

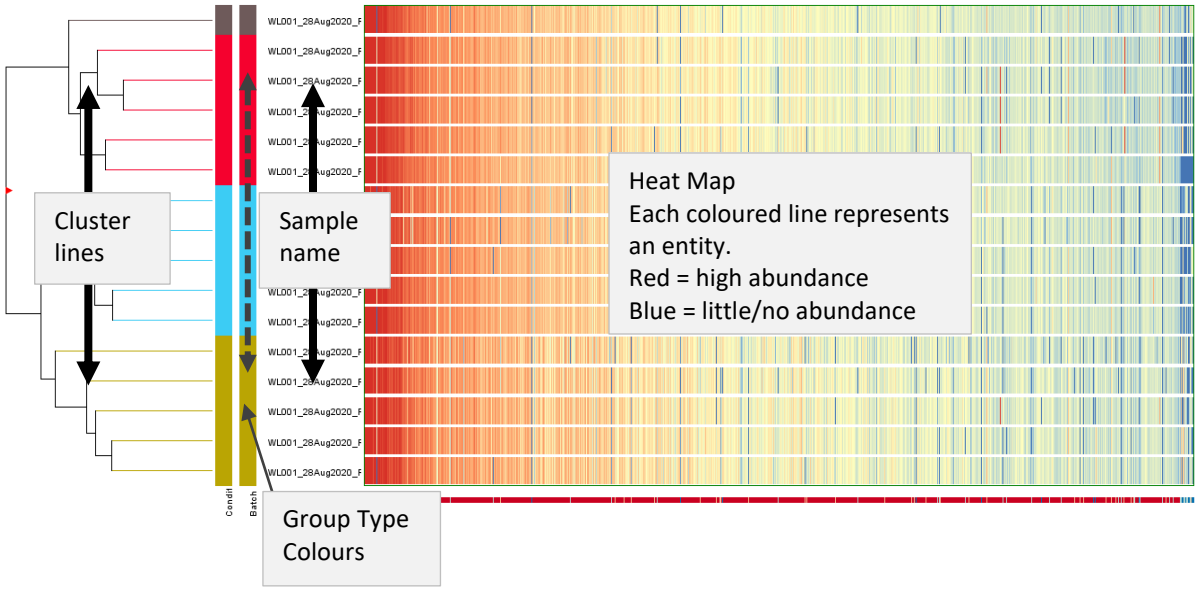

Figure 1-14. Hierarchical cluster map example.

This map consists of different batches of an herbal remedy with one QC sample (grey). The arches on the left-hand side show which sample is most similar to another.

The coloured lines on the right-hand side are similar to the complex plot in the **Figure 1.11**. The red coloured lines show entities of the highest abundance, through to the blue lines which show no abundance, 0. Therefore the blue lines indicate that the entity is missing from that sample. The colours in the middle (orange/yellow/green) show the abundances in between highest and lowest.

Overall, MPP is a great tool as it lets you easily classify, compare, and analyse sample groups. Something to watch out for is the authenticity of the entities that it highlights. These are highly dependent on many of the customised parameters and therefore if just one value isn't optimised correctly, such as lower abundance cut off value, then it is possible to get false positives, or even false negatives. Once the parameters are optimised, it is not only able to find the 'needle in the haystack', but it can also characterise each potential feature.

1.3.4 R

'R' is a popular statistical programming language originating from 'S'; R is a dialect of S. The term 'R' is used for both the programming language to write scripts, and the software that interprets said scripts. R is a lot more in depth than other software, with a lot more training required to do basic commands. A long string of written commands is needed to do analysis, but also means that the same scripts can be used if the data was altered [72]. For example, if you collect more data then you can just run the script again on the data without having to remember what exactly had been done for each step of obtaining the results. The script enables clear steps in the analysis so that another person can review and possibly give feedback on how to improve the analysis. Each step is shown in code so that it can be reviewed in depth should a mistake occur, giving a deeper understanding of what you are doing, furthering the user's knowledge. This is known as open source, meaning there's less chance for a mistake given that it is completely transparent. If mistakes are found, there are places to report and fix these bugs [72].

Reproducibility is a main advantage of R: the same results can be obtained when any person conducts analysis on the same data set. This is due to the same code that can be used and providing a good data set is obtained, the analysis cannot be miss-interpreted. An increasing number of journals understand how reproducible R is and knowingly give you an edge should you include R analysis to back up any findings. R works on a large range of data sets, whether there is tens or thousands of lines, there will not be a noticeable difference on time or efficiency with the software. R can connect many data formats, such as spreadsheets or databases, on either a web or from personal computer files.

As well as all the advantages to R, there are obviously some limitations. The main limitation is that it is based upon 50-year-old technology [72]. There have been significant updates to the software as previously it was unable to handle the graphics that are now available. It also advanced in areas of memory: each statistical analysis package is now available to download so that R does not take up all

the computer memory with large data sets. Another limitation could be that the scripts are based on customer demand made by voluntary users and therefore is no one has made the script that you need for your analysis, then you will have to create it yourself. However, the R community is forever growing and so there are always plenty of other users who would be more than happy to help in any of the online forums should any difficulties arise.

As well as statistical analysis, high quality graphics are also available in such a way that conveys the best message about the data. This can be adapted in many ways to get the most efficient plots, such as colours, shapes, labels etc. An example of a PCA plot made by coding in R is shown in **Figure 1.15** where the different objects represent a sample, and the group with the ellipse represents a group type. The details can be seen in the key on the right-hand side of the Figure.

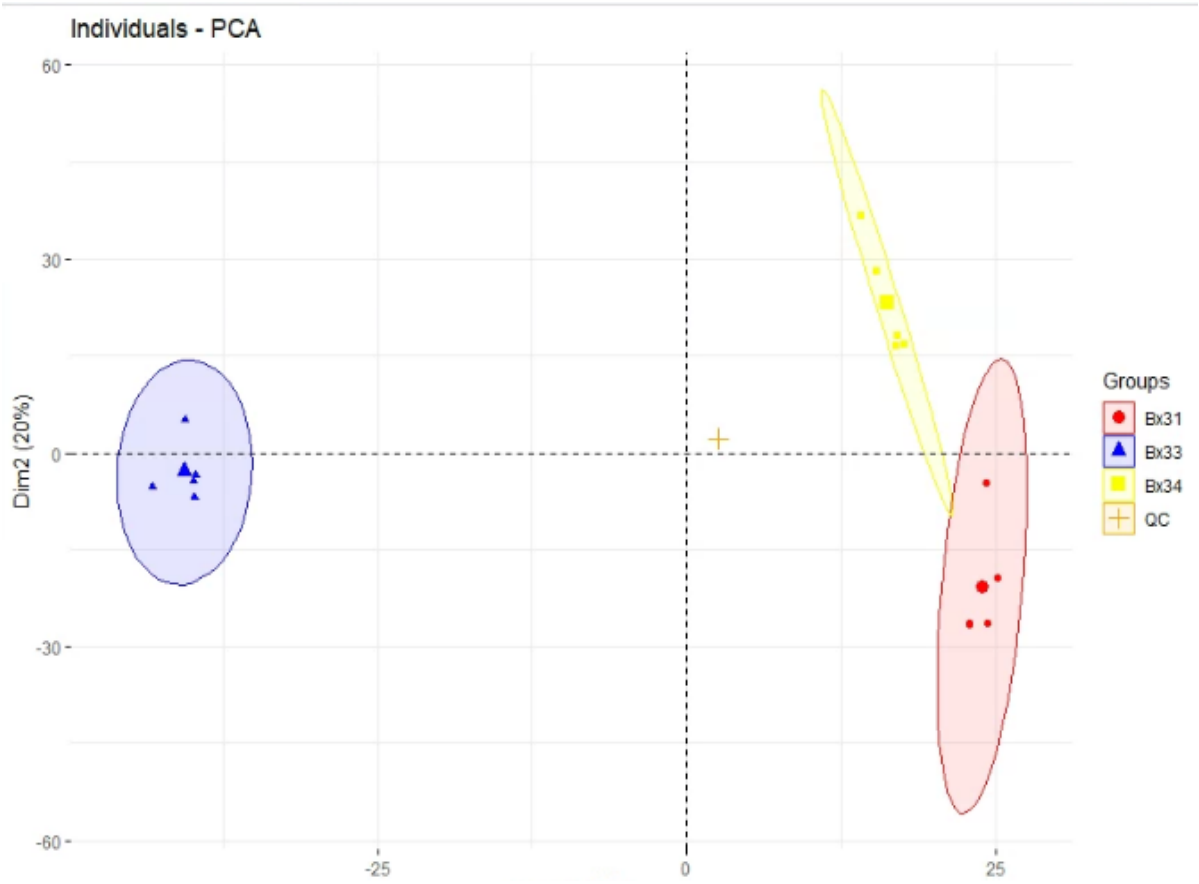

Figure 1-15. An example PCA plot made in R.

Ellipses of 95% confidence are around each sample group. The key on the right-hand side shows the sample groups, plus the QC sample, with a distinct colour used for each. It is therefore possible to quickly determine that the three sample groups are different to each other, and each sample within the group are clustered closely to those with its own sample type. It shows that Bx33 is very different to the other batches with the 44.6% spread along the x-axis from Bx31 and Bx34, which are clustered more closely together. These are separated by the y-axis, with a 20% spread. The pooled QC shown by the orange cross sitting in the middle of the plot shows that the instrument and method are both working sufficiently as it is sitting in between them all.

Not only is R a piece of software available for free, but it is also available across all platforms (Windows, Mac and Linex). R has a large community of people, most of which will help if there's any problems encountered, and typical websites such as Stack Overflow show solutions to these problems.

1.3.5 MetaboAnalyst

There are other software packages available that give also statistical information, though are freely available online. There is a website called 'MetaboAnalyst' which also gives statistical features like volcano plots and PCA plots as well as univariate statistical analysis. It also allows the building of

sample group predictions, as well as metabolite identification and metabolic pathway analysis [73]. To use MetaboAnalyst, the raw data must first be analysed in peak picking software such as XCMS online, giving results in a format that is readable by MetaboAnalyst (i.e. .txt, mzXML, .csv, mzData). The peak intensity table obtained from raw data is not interpretable for most researchers, however it can be exported and uploaded straight into MetaboAnalyst where through a series of steps, it can convert the data into visual features so that it can be understood by many [67]. It is also possible to combine MetaboAnalyst with R to give a more tailored data analysis, of which can be downloaded in R as an R package [74].

1.3.6 XCMS online

Over the past years, software has evolved and XCMS has become one of the most popular open-source tools for processing raw data, with it currently being the most cited pre-processing software used to date in the metabolomics literature [33]. XCMS online picks the peaks, filters them, matches them across samples and corrects the retention times to align each data file accordingly. Any missing values are filled with a small number that does not affect the analysis in any way as statistical analysis cannot be performed if there are any missing values. The aligned data files are grouped and the reported peaks are displayed as a peak table, ready for statistical analysis. Visual tools such as EICs, PCA plots and metabolomic cloud plots are available. A cloud plot is a visual graphic where the features are displayed on a plot, separated by *m/z* on the y-axis and RT on the x-axis. Each feature is represented by a circle, where the larger the circle, the larger the average abundance within a sample group. It shows up and down regulated features, with the different sample groups represented by a different colour. **Figure 1.16** shows an example plot of this metabolomic cloud plot, the green circles representing up regulated compared to the control group and red circles showing the down regulated features compared to the control group.

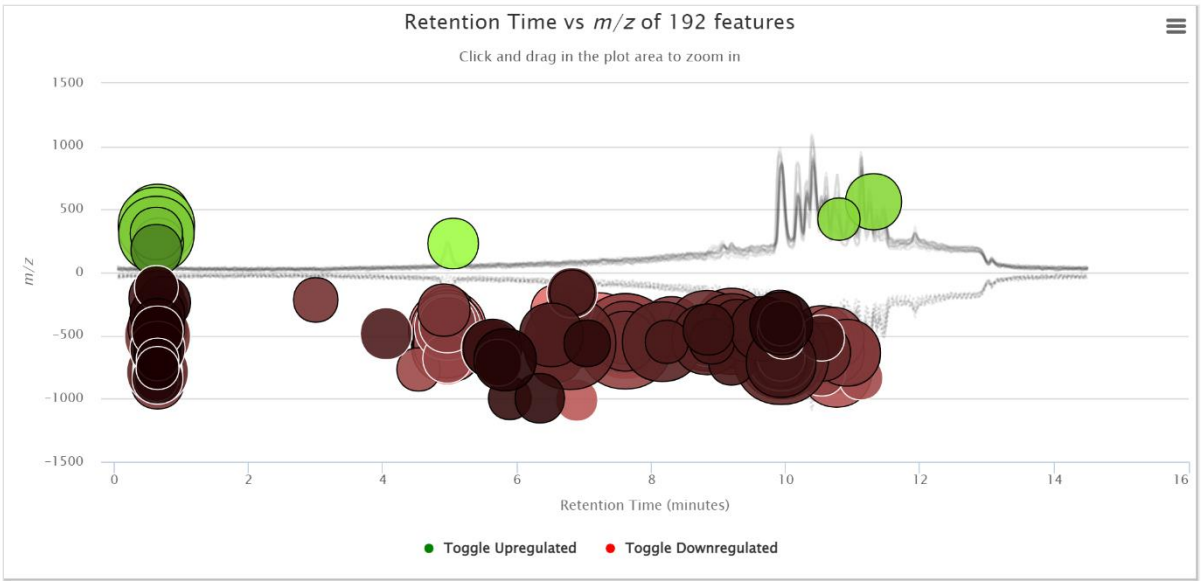

Figure 1-16. Example of a metabolomic cloud plot in XCMS online

1.4 Comparison of Software

During untargeted metabolomics projects, the software used to analyse data is another added variable to consider for the analysis. Several studies have compared the differences of feature and biomarker detection across the different untargeted data processing software but each one has discovered that there are several differences in the results. A comparison of MetAlign, XCMS and MZ Mine by Coble et al. showed that there are significant differences in feature detection and quantification [75]. Rafiei et al. discovered that less than 10% of compound peaks were common to all four types of software that was compared: Peakview, Markerview, Metabolite Pilot and XCMS Online [76]. Myers et al. recently reported several problems in the peak detection algorithm, *centWave*, used in both XCMS and MZ Mine 2 that are partly responsible for many false positives and negatives throughout datasets [77]. Li et al. more recently observed significant differences in the true features detected, and quantified, by using MS-Dial, MZ Mine 2, XCMS, MarkerView and Compound Discoverer [33]. All these mentioned studies only analyse a small number of samples within type of dataset, which is significantly below the complexity of real metabolomic samples.

It is key when doing non-targeted metabolomics to use a reproducible and reliable method and therefore a fully validated method must be developed to ensure the variability is from only the data itself, not from the method being used. QC's should be used to ensure that the instrument is working sufficiently and signal is repeatable across the batch and allow signal correction. Samples should be injected in a random order to ensure that there is no instrumental bias. Pre-validation can save data analysis problems down the road as you can trust the instrument, hence trust the data set. In chemometric analysis with a non-targeted approach, the large data sets can be presented by class prediction in multivariate statistical analysis. There are multiple different ways that this can be produced, though it is important that this is done reliably. Validation of such statistical models can be done to avoid overfitting the model, though this is out of the scope for this work. This is something that could be investigated in the future.

1.5 Roadmap of this Thesis

This thesis consists of five chapters, including an introduction and an overall conclusion with any final remarks and future work. All chapters highlight the fact that good datasets are the most fundamental parts of achieving reliable results. Deconvolution is the next step that is shown to be very important, as without the right peaks being picked consistently throughout the data available, reliable statistical analysis cannot be conducted. Chapters 2 and 3 show that XCMS online does not provide accurate results, with little to no accurate statistical analysis being performed. The same datasets showed that potential biomarkers for different types of tea (Chapter 2) and biomarkers for bleached hair (Chapter 3) are available, using Mass Profinder or Mass Profiler for deconvolution. Chapter 4 includes data that showed with a standard C18 column on RP-LC, no differences are found between the dried blood spots of major depressive disorder (MDD) patients and 'healthy' patients. However, further work showed that it is likely that the data was collected in a way that does not capture everything as it was later found that there are multiple lipid-type molecules found to be of different concentrations in the blood of these patients, possibly highlighting the discovery of a potential biomarker for MDD in

blood. Chapter 5 summarises all the findings discovered throughout the thesis, followed by opinions on the software and recommendations on future work that could be considered.

1.6 Aims and Objectives

There is a major risk of losing important data when choosing the best deconvolution and statistical analysis software packages to create a metabolomic workflow for untargeted LC-MS analysis. Theoretically, if the same parameters are chosen, it shouldn't matter which software is used as they should all give the same results. However, recreating analysis in different software packages has been seen to give contradicting results and therefore a full investigation into how different these results are, and which give the most reliable results is of great interest to those in the metabolomics community to minimise the risk of losing important information. To ensure that the software packages can be used reliably throughout the metabolomics community, it would be ideal to use datasets across a range of disciplines so it doesn't continue to hinder the application of a chemometric approach.

The objectives of this thesis include:

- Investigate the current approaches for statistical analysis used in metabolomics.
- Investigate the influence on using different deconvolution packages using different datasets from multiple disciplines.
- Investigate the use of different statistical analysis software packages using different datasets from multiple disciplines.
- Use the deconvolution and statistical analysis software packages to create a robust, reliable workflow for discovery of biomarkers.
- Compare the different software packages to determine which combined workflow gives the most reliable results and creates the biggest impact on the user's experience through ease of use, cost and reliability.

1.7 References

1 Brigante F, Podio N, Wunderlin D, Baroni M. Comparative metabolite fingerprinting of chia, flax and sesame seeds using LC-MS untargeted metabolomics. Food Chemistry. 2021;371:131355.

2 J. K. Nicholson, J. C. Lindon, and E. Holmes. Metabolomics: understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. Xenobiotica. 1999;29(11):1181-1189.

3 J. C. Lindon, E. Holmes, and J. K. Nicholson. So what's the deal with metabolomics. Analytical Chemistry. 2003;75(17):384-391.

4 D. Vuckovic. Current trends and challenges in sample preparation for global metabolomics using liquid chromatography–mass spectrometry. Analytical and Bioanalytical Chemistry. 2012;403(6):1523-1548.

5 H. Issaq, Q. Van, T. Waybright, G. Muschik, and T. Veenstra. Analytical and statistical approaches to metabolomics research. Journal of Separation Science. 2009;32(13):2183-2199.

6 D. Yoon, M. Lee, S. Kim, and S. Kim. Applications of NMR spectroscopy based metabolomics: A review. Journal of the Korean Magnetic Resonance Society. 2013;17(1):1-10.

7 S. Putri, S. Yamamoto, H. Tsugawa, and E. Fukusaki. Current metabolomics: Technological advances. Journal of Bioscience and Bioengineering. 2013;116(1):9-16.

8 Jacobs D, van den Berg M, Hall R. Towards superior plant-based foods using metabolomics. Current Opinion in Biotechnology. 2021;70:23-28.

9 Zhang A, Sun H, Wang P, Han Y, Wang X. Modern analytical techniques in metabolomics analysis. The Analyst. 2012;137(2):293-300.

10 Gong Z, Hu J, Wu X, Xu Y. The Recent Developments in Sample Preparation for Mass Spectrometry-Based Metabolomics. Critical Reviews in Analytical Chemistry. 2017;47(4):325-331.

11 Subramanian I, Verma S, Kumar S, Jere A, Anamika K. Multi-omics Data Integration, Interpretation, and Its Application. Bioinform Biol Insights. 2020;14(1):1-24

12 Oliver S. Systematic functional analysis of the yeast genome. Trends Biotechnol. 1998;16(9):373-378.

13 Fiehn O. Metabolomics – the link between genotypes and phenotypes. Plant Mol Biol. 2002;48(1/2):155-171.

14 deHoffman E, Stroobant V. Mass Spectrometry: Principles and Applications. Toronto: Wiley. 2003;1:133.

15 Johnson C, Ivanisevic J, Benton H, Siuzdak G. Bioinformatics: The Next Frontier of Metabolomics. Analytical Chemistry. 2015;87(1):147-156.

16 Katajamaa M, Orešič M. Data processing for mass spectrometry-based metabolomics. Journal of Chromatography A. 2007;1158(1-2):318-328.

17 Wan, E, Masson, P. Processing and analysis of GC/LC-MS-based metabolomics data. Methods Mol Biol . 2011;708:277–298.

18 Cambiaghi, A.; Ferrario, M.; Masseroli, M. Analysis of metabolomic data: Tools, current strategies and future challenges for omics data integration. Brief. Bioinform. 2016;18:498–510

19 Stanstrup J, Broeckling C, Helmus R, Hoffmann N, Mathé E, Naake T et al. The metaRbolomics Toolbox in Bioconductor and beyond. Metabolites. 2019;9(10):200.

20 Pang Z, Chong J, Li S, Xia J. MetaboAnalystR 3.0: Toward an Optimized Workflow for Global Metabolomics. Metabolites. 2020;10(5):186.

21 Huang S, Toh W, Benke P, Tan C, Ong C. MetaboNexus: an interactive platform for integrated metabolomics analysis. Metabolomics. 2014;10(6):1084-1093.

22 Smith C, Want E, 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(3):779-787.

23 Pluskal T, Castillo S, Villar-Briones A, Orešič M. MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. BMC Bioinformatics. 2010;11(1):1.

24 Clasquin, M.F, Melamud, E, Rabinowitz, J.D. LC-MS Data Processing with MAVEN: A Metabolomic Analysis and Visualization Engine. Curr Protoc Bioinform. 2012;37(1): 1-23.

25 Tautenhahn R, Böttcher C, Neumann S. Highly sensitive feature detection for high resolution LC/MS. BMC Bioinformatics. 2008;9(1):1.

26 Tautenhahn R, Patti G, Rinehart D, Siuzdak G. XCMS Online: A Web-Based Platform to Process Untargeted Metabolomic Data. Analytical Chemistry. 2012;84(11):5035-5039.

27 Gowda H, Ivanisevic J, Johnson C, Kurczy M, Benton H, Rinehart D et al. Interactive XCMS Online: Simplifying Advanced Metabolomic Data Processing and Subsequent Statistical Analyses. Analytical Chemistry. 2014;86(14):6931-6939.

28 E. M. Lenz and I. D. Wilson. Analytical strategies in metabolomics. Journal of Proteome Research. 2007;6(2):443-458.

29 S. Ren, A. Hinzman, E. Kang, R. Szczesniak, and L. Lu. Computational and statistical analysis of metabolomics data. Metabolomics. 2015;11(6):4192-1513.

30 Mozzi F, Ortiz M, Bleckwedel J, De Vuyst L, Pescuma M. Metabolomics as a tool for the comprehensive understanding of fermented and functional foods with lactic acid bacteria. Food Research International. 2013;54(1):1152-1161.

31 Griffin J, Shockcor J. Metabolic profiles of cancer cells. Nature Reviews Cancer. 2004;4(7):551-561.

32 Dettmer K, Aronov P, Hammock B. Mass spectrometry-based metabolomics. Mass Spectrometry Reviews. 2006;26(1):51-78.

33 Li Z, Lu Y, Guo Y, Cao H, Wang Q, Shui W. Comprehensive evaluation of untargeted metabolomics data processing software in feature detection, quantification and discriminating marker selection. Analytica Chimica Acta. 2018;1029:50-57.

34 P. Masson, A. C. Alves, T. M. D. Ebbels, J. K. Nicholson, and E. J. Want. Optimization and evaluation of metabolite extraction protocols for untargeted metabolic profiling of liver samples by UPLC-MS. Analytical Chemistry. 2010;82(18):7779-7786.

35 S. Barnes, H. Benton, K. Casazza, S. Cooper, X. Cui, X. Du, J. Engler, J. Kabarowski, S. Li, W. Pathmasiri, J. Prasain, M. Renfrow, and H. Tiwari. Training in metabolomics research. i. Designing the Experiment, collecting and extracting samples and generating metabolomics data. Journal of Mass Spectrometry. 2016;51(7):461-475.

36 L. Sumner, A. Amberg, D. Barrett, M. Beale, R. Beger, C. Daykin, T. Fan, O.Fiehn, R. Goodacre, J. Grin, T. Hankemeier, N. Hardy, and J. Harnly. Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). Metabolomics. 2007;3(3):211-221.

37 Blainey P, Krzywinski M, Altman N. Points of significance: replication. Nature methods. 2014 Sep 1;11(9):879.

38 G. Theodoridis, H. Gika, and I. Wilson. LC-MS-based methodology for global metabolite profiling in metabolomics/metabolomics. Trends in Analytical Chemistry. 2008;27(3):251-260.

39 T. Sangster, H. Major, R. Plumb, J. A. Wilson, and I. D. Wilson. A pragmatic and readily implemented quality control strategy for HPLC-MS and GC-MS-based metabolomic analysis. The Analyst. 2006;131(10):1075-1078.

40 Álvarez-Sánchez B, Priego-Capote F, Luque de Castro M. Metabolomics analysis I. Selection of biological samples and practical aspects preceding sample preparation. TrAC Trends in Analytical Chemistry. 2010;29(2):111-119.

41 B. Zhou, J. Xiao, L. Tuli, and H. Ressom. LC-MS-based metabolomics. Molecular BioSystems. 2012;8(2):470-481.

42 B. Alvarez-S´anchez, F. Priego-Capote, and M. D. L. de Castro. Metabolomics analysis ii. Preparation of biological samples prior to detection. Trends in Analytical Chemistry. 2010;29(2):120-127.

43 Gika H, Theodoridis G. Sample preparation prior to the LC–MS-based metabolomics/metabolomics of blood-derived samples. Bioanalysis. 2011;3(14):1647-1661.

44 H. R. Tang and Y. L. Wang. Metabolomics: A revolution in progress. Progress in Biochemistry and Biophysics. 2006;33(5);401-417.

45 W. B. Dunn, N. J. C. Bailey, and H. E. Johnson. Measuring the metabolome: Current analytical technologies. Analyst. 2005;130(5);606-625.

46 Loo J. Studying noncovalent protein complexes by electrospray ionization mass spectrometry. Mass Spectrometry Reviews. 1997;16(1):1-23.

47 Smith R, Loo J, Edmonds C, Barinaga C, Udseth H. New developments in biochemical mass spectrometry: electrospray ionization. Analytical Chemistry. 1990;62(9):882-899.

48 C. Ho, M. Chan, R. Cheung, L. Law, L. Lit, K. Ng, M. Suen, and H. Tai. Electrospray Ionisation Mass Spectrometry: Principles and Clinical Applications,. The Clinical Biochemist Reviews. 2003;24(1):3–12.

49 Yin P, Zhao X, Li Q, Wang J, Li J, Xu G. Metabolomics Study of Intestinal Fistulas Based on Ultraperformance Liquid Chromatography Coupled with Q-TOF Mass Spectrometry (UPLC/Q-TOF MS). Journal of Proteome Research. 2006;5(9):2135-2143.

50 Wilson I, Nicholson J, Castro-Perez J, Granger J, Johnson K, Smith B et al. High Resolution "Ultra Performance" Liquid Chromatography Coupled to oa-TOF Mass Spectrometry as a Tool for Differential Metabolic Pathway Profiling in Functional Genomic Studies. Journal of Proteome Research. 2005;4(2):591-598.

51 O'Connor D, Mortishire-Smith R. High-throughput bioanalysis with simultaneous acquisition of metabolic route data using ultra performance liquid chromatography coupled with time-of-flight mass spectrometry. Analytical and Bioanalytical Chemistry. 2006;385(1):114-121.

52 Rubinson J, Rubinson K. Contemporary instrumental analysis. Prentice Hall; 1999.

53 Man K, Chan C, Tang H, Dong N, Capozzi F, Wong K et al. Mass spectrometry-based untargeted metabolomics approach for differentiation of beef of different geographic origins. Food Chemistry. 2021;338:127847.

54 Schiffman C, Petrick L, Perttula K, Yano Y, Carlsson H, Whitehead T et al. Filtering procedures for untargeted LC-MS metabolomics data. BMC Bioinformatics. 2019;20(1):334.

55 Chang, H., Colby, S., Du, X., Gomez, J., Helf, M., & Kechris, K. et al. A Practical Guide to Metabolomics Software Development. Analytical Chemistry. 2021;93(4):1912-1923.

56 D. Barupal, S. Fan, and O. Fiehn. Integrating bioinformatics approaches for a comprehensive interpretation of metabolomics datasets. Current Opinion in Biotechnology. 2018;54:1-9.

57 E. Saccenti, H. Hoefsloot, A. Smilde, J. Westerhuis, and M. Hendriks. Reflections on univariate and multivariate analysis of metabolomics data. Metabolomics. 2014;10(3):361-374.

58 Barnes, H. Benton, K. Casazza, S. Cooper, X. Cui & X. Du, et al. Training in metabolomics research. ii. processing and statistical analysis of metabolomics data, metabolite identification, pathway analysis, applications of metabolomics and its future. Journal of Mass Spectrometry. 2016;51(8):535-548.

59 R. G. Brereton. Applied Chemometrics for Scientists. Chichester, England: John Wiley & Sons Ltd., 2007.

60 Analytical Methods Committee AMCTB. Multivariate statistics in the analytical laboratory (1): an introduction. Analytical Methods. 2021;13(2):274–277.

61 Boccard J, Rudaz S. Harnessing the complexity of metabolomic data with chemometrics. Journal of Chemometrics. 2013;28(1):1-9.

62 Everitt B, Dunn G. Applied multivariate data analysis. 2nd ed. Chichester: John Wiley & Sons; 2001.

63 F. Marini, Chemometrics in Food Chemistry. Oxford, UK: Elsevier, first ed.,2013

64 Sauzier G, van Bronswijk W, Lewis S. Chemometrics in forensic science: approaches and applications. Analyst. 2021;146(8):2415-2448.

65 Kumar S. Monitoring Novel Corona Virus (COVID-19) Infections in India by Cluster Analysis. Annals of Data Science. 2020;7(3):417-425.

66 Cui, X.; Churchill, G. A. Statistical tests for differential expression in cDNA microarray Experiments. Genome Biology. 2003;4(4):210.

67 S. Naz, M. Vallejo, A. Garc´ıa, and C. Barbas. Method validation strategies involved in non-targeted metabolomics. Journal of Chromatography A. 2014;1(1353):99-105.

68 Palmblad M, Mills D, Bindschedler L, Cramer R. Chromatographic alignment of LC-MS and LC-MS/MS datasets by genetic algorithm feature extraction. Journal of the American Society for Mass Spectrometry. 2007;18(10):1835-1843.

69 Fernández-Ochoa Á, Quirantes-Piné R, Borrás-Linares I, Cádiz-Gurrea M, Alarcón Riquelme M, Brunius C et al. A Case Report of Switching from Specific Vendor-Based to R-Based Pipelines for Untargeted LC-MS Metabolomics. Metabolites. 2020;10(1):28.

70 Sana, T. Profinder webinar. The Most Advanced Batch Feature Extraction Software for Metabolomics. Retrieved from Agilent library: Accessed 14 April 2020.

71 Mass Profiler Professional Software Resource Package. Retrieved from Agilent Technologies: http://www.agilent.com/en-us/products/software-informatics/masshunter- suite/masshunter-for-life-science-research/mass-profiler-professional-software. Accessed 14 April 2020.

72 Nicolotti L, Hack J, Herderich M, Lloyd N. MStractor: R Workflow Package for Enhancing Metabolomics Data Pre-Processing and Visualization. Metabolites. 2021;11(8):492.

73 Xia J, Wishart D. Using MetaboAnalyst 3.0 for Comprehensive Metabolomics Data Analysis. Current Protocols in Bioinformatics. 2016;55(1).

74 Pang Z, Chong J, Zhou G, de Lima Morais D, Chang L, Barrette M et al. MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights. Nucleic Acids Research. 2021;49(W1):388-396.

75 Coble J, Fraga C. Comparative evaluation of preprocessing freeware on chromatography/mass spectrometry data for signature discovery. Journal of Chromatography A. 2014;1358:155-164.

76 Rafiei A, Sleno L. Comparison of peak-picking workflows for untargeted liquid chromatography/high-resolution mass spectrometry metabolomics data analysis. Rapid Communications in Mass Spectrometry. 2015;29(1):119-127.

77 Myers O, Sumner S, Li S, Barnes S, Du X. Detailed Investigation and Comparison of the XCMS and MZmine 2 Chromatogram Construction and Chromatographic Peak Detection Methods for Preprocessing Mass Spectrometry Metabolomics Data. Analytical Chemistry. 2017;89(17):8689-8695.

2 - The Development and Evaluation of an Untargeted Metabolomic Workflow on a Range of Tea Samples

2.1 Aims

The aims of this experiment were to develop and evaluate an untargeted metabolomic workflow for tea samples. An investigation using liquid chromatography-mass spectrometry and comparison software techniques was conducted, with the objective to determine whether potential significant features that are unique, or up regulated to a sample group could be obtained. Using tea samples with simple sample preparation, instrumental methods could be developed and the workflow could be evaluated. Furthermore, the untargeted chemometric software techniques were then evaluated to compare whether the same biomarkers are determined throughout.

2.2 Introduction

2.2.1 Tea Plant

Tea, brewed from leaves of the plant *Camellia sinensis*, is one of the most popular beverages in the world with about three billion kilograms produced and consumed yearly [1]. This plant is native to Southeast Asia however it is currently cultivated in more than 30 countries around the world. Once the leaves are picked from *camellia sinensis*, they can be made into black tea by allowing them to fully oxidise under controlled temperature and humidity until the leaves turn fully black or brown, then they are dried. During the oxidation process, the oxygen interacts with the tea plant's cell walls, turning them into the brown to black colour whilst also altering the flavour profile [1]. This process is also called fermentation, however no actual fermentation takes place and the active enzyme is polyphenol oxidase which is found naturally in plants [2]. The same tea leaves can be made into green tea by heating them to halt the oxidation process immediately after harvesting them. This is usually done by steaming or pan-firing the leaves to ensure the leaves do not undergo fermentation, and

therefore they will stay bright green. Tea is produced and consumed in different forms but 78% is produced as black tea usually in Western countries, where-as Asian countries tend to drink green tea, correlating to roughly 20% of all tea consumed. In Asia, besides water, tea is the most consumed beverage [2]. Southern China tend to partially ferment their tea producing oolong tea, which is the remaining 2% of tea consumed across the world. Within the last few centuries, white tea has also come onto the market after been harvested primarily in China. There is little agreement on a definition on what white (albino) tea is, though it is currently referred to tea that has no additional processing after it has been dried and is usually only the tips of the leaves that are picked [3]. Yellow tea is not very well known, though it is gradually getting more recognition in the Western countries due to its unique flavour and apparent health-promoting properties [4]. It is made by a similar procedure to green tea; however, it undergoes light fermentation before drying, a step that is often referred to as "sealed yellowing". This step gives the tea the bright yellow appearance with a sweet, mellow taste in comparison to the grassy taste that green tea gives [4]. Purple tea has also recently come into the market, providing potentially greater health benefits than those documented about green tea. This type of tea is slightly different in that although it originates from the same *Camellia sinensis* plant, the purple appearance is caused by a genetic mutation which produces anthocyanins. These anthocyanins are the same powerful compound class that are found in blueberries and cause the colourfulness of many fruits and vegetables. It was originally found in small quantities in China, however a partnership with Kenya to isolate the mutation has seen the tea being mass produced in Kenya due to the high altitude causing very high levels of antioxidants being produced, protecting the leaves from damage. **Figure 2.1** shows a simplified pathway of how each of the different types of tea are made. After this end point of the graph, the leaves can be oxidised with potassium hydroxide (KOH) to produce tea extracts.

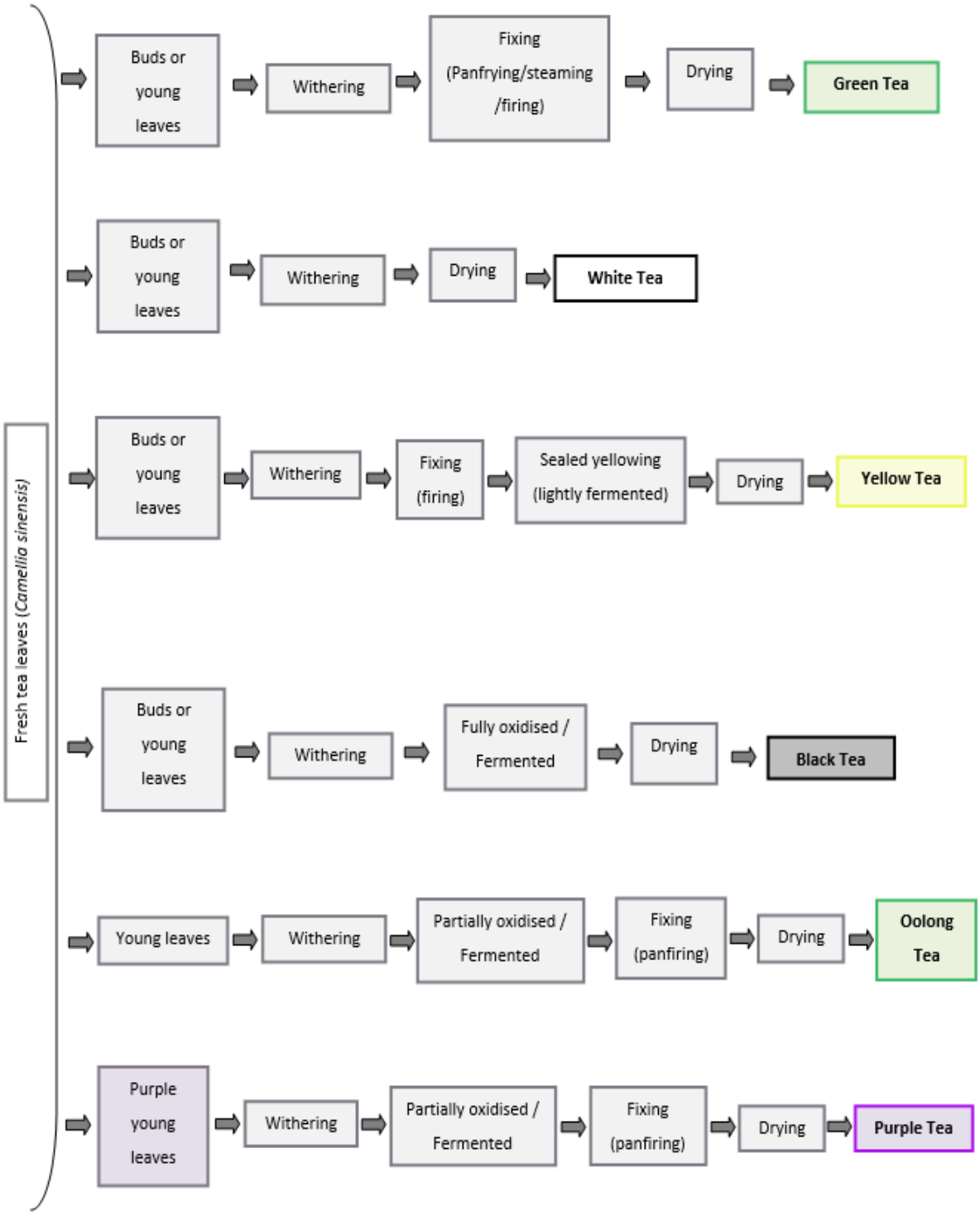

Figure 2-1. Simplified pathway of how different types of teas are made.

Adapted from reference 22

2.2.2 Brewed Tea

Brewed tea is known to contain many different chemical compounds but is constituted primarily of polyphenols which account for the aroma but also several studies have shown that they reduce the risk of a variety of diseases [5-7]. Green tea is often credited with the most health benefits and antioxidant properties as the polyphenols are available in much higher concentrations than they are in black or oolong tea [8]. It can be noted that studies on purple tea comparison to the green teas are more difficult to come across due to the more recent discovery of purple tea. The different types of polyphenols in tea include catechins (flavan-3-ols), tannins and flavonoids. In green tea, the content of the catechin epigallocatechin-3-gallate (EGCG) is higher than it is in other teas and is the polyphenol that is often the main subject of the health studies with regards to tea. Some other catechins that are present in tea include epigallocatechin, epicatechin gallate and epicatechin, a general structure of catechins can be seen in **Figure 2.2**. Alternate flavanols found in tea include kaempferol, quercetin and myricetin. Flavonoids that are found in black tea include theaflavin (TF-1), theaflavin-3-gallate (TF-2) and theaflavin-3,3-digallate (TF-3), of which a general simplified structure can be seen in **Figure 2.2**. Tannins are bitter polyphenolic compounds that hold the ability to bind and precipitate organic compounds, making them typically much larger than other polyphenols [9]. An example of a tannin found in tea is ellagitannin, an example of a hydrolysable tannin that can be seen in **Figure 2.2**. Condensed tannins, also known as proanthocyanidins, are often formed from two or more molecules of flavan-3-ols (catechins), hence their present in teas.

Due to the various reported health benefits, polyphenol rich food and drink is of interest to many researchers and therefore further investigation into different tea products was conducted to gather more information about the levels of polyphenols in tea beverages, as well as any other compounds with potentially great health benefits.

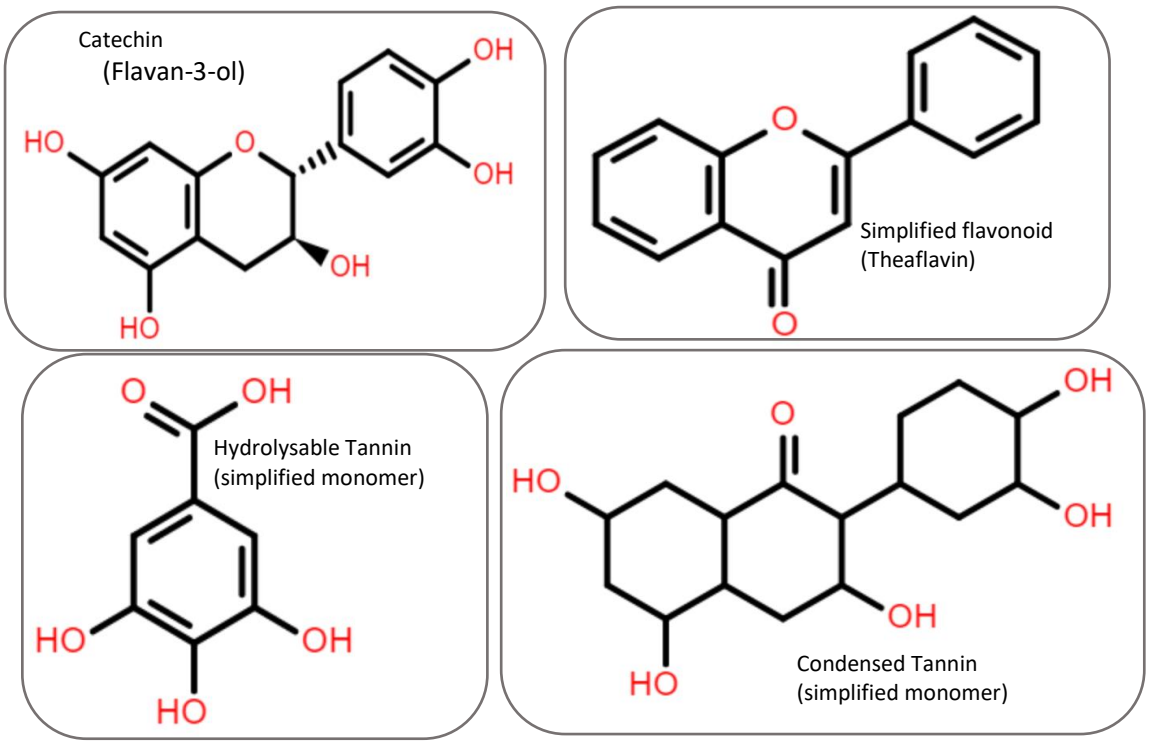

Figure 2-2. Chemical Structures of some simplified structures of common tea compounds

2.2.1.1 Anthocyanins

Anthocyanins, a type of polyphenol, are of interest as these also have potentially great health benefits for humans. In tea, anthocyanidin glycosides (anthocyanidins) are often present, though these originate from their parent anthocyanin, of which the basic structure can be seen in **Figure 2.3**. These compounds are analysed in positive mode.

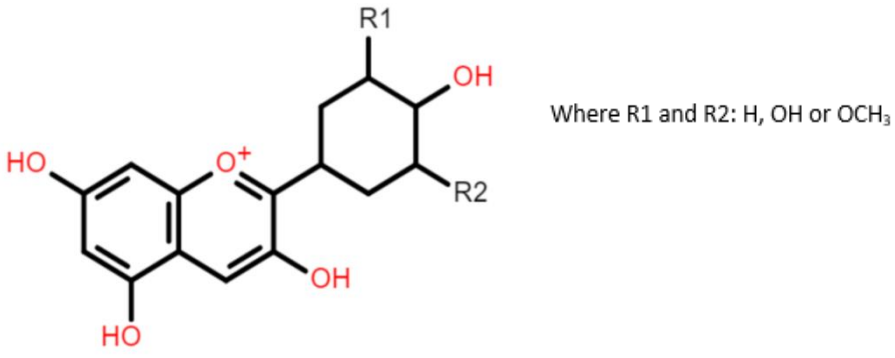

Figure 2-3. Basic chemical structure of an anthocyanin.

Anthocyanins give the coloured pigment to fruits and vegetables when in their glycosylated forms, such as in berries. Pelargonidin is an example of an anthocyanin that potentially gives the red-purple colour to a purple tea, as well as blackberries, blueberries and raspberries. Polyphenols have many

health benefits and so the higher the concentration, the 'better' the tea would be as it would be more beneficial for the consumer. The health benefits are something that many people around the world are searching for, especially given the increase in heart disease, obesity and other diseases.

2.3 Materials

Water, methanol (MeOH), acetonitrile (ACN) and formic acid of optima LCMS grade were obtained from Fisher Scientific (Loughborough, U.K.). ESI-L low concentration tuning mix and reference peak markers were purchased from Agilent Technologies (California, USA). The samples were provided by Finlay Beverages Ltd. The analysis was performed using Agilent Technologies 1290 Infinity II UHPLC system, coupled to an Agilent Technologies 6550 Accurate-Mass Quadrupole-Time-of-Flight mass spectrometer with iFunnel.

2.4 Experimental

The difference between different coloured tea extracts and leaves is something that was also under investigation. It was also interesting to determine whether there is any benefit to the newly discovered purple tea, or whether it is too similar to green and yellow tea. Tea extracts and tea leaf samples were provided for sample comparison analysis. There were 11 green tea extracts, 1 black tea extract, 1 purple tea extract, 4 yellow tea extracts, 6 purple tea leaf, 19 green tea leaf and 1 albino tea leaf. Each tea extract was prepared in the same way, then the method was adapted for the tea leaf samples due to the change in concentration of compounds present.

Approximately 100 mg (± 5 mg) of each tea extract sample was weighed into separate 50 mL volumetric flasks in duplicate. 25 mL of hot water (< 60 °C) was added and mixed to dissolve the sample. This was then left to cool to room temperature. To this, 5 mL of LCMS-grade acetonitrile was added and LCMS water was used to fill to the mark and the solution was mixed well. Approximately 1

mL of the sample solution was taken and centrifuged at 9800 RCF for 10 minutes, and then the supernatant was transferred to an LC vial for LCMS analysis.

Approximately 20 mg of tea leaf sample was weighed into a 1.5 mL Eppendorf tubes in duplicate then 1.0 mL of a 70% v/v LCMS methanol in LCMS water solution was pipetted into each tube. They were shaken well, placed in the oven at 70 °C for 10 min with shaking. The tubes were removed from the oven, allowed to cool and centrifuged at 9800 RCF for 10 minutes. The supernatants were decanted into 5 mL volumetric flasks. A further 1.0 mL aliquot of 70 %v/v methanol in water solution was pipetted into each tube, shaken and placed in the oven again at 70 °C for 10 min with shaking. The tubes were removed from the oven, cooled to room temperature and centrifuged at 9800 RCF for 10 minutes. The supernatants were combined with the first extractions by decanting them into the same 5 mL volumetric flask. The volumetric was made to the line with LCMS water and 1 mL of each solution was filtered through a 0.45 µm PTFE syringe filter into LC vials for analysis.

Two different LC-MS methods were used for this analysis. The first method shown in **Table 2.1** was used for all samples, all tea extract and tea leaf samples. This was developed and optimised using some polyphenol standards, along with a randomly selected extract and leaf sample, and no compounds of interest were extracted after a mobile phase B composition of 38%.

Table 2.1. LC-MS Method for All Tea Extract and Tea Leaf samples.

This was conducted in positive and negative ionisation modes.

HPLC-MS					
Instrument	Agilent 1290 Infinity II UHPLC + 6550 Q-ToF with iFunnel				
Column	Agilent Eclipse Plus C18 2.1mm x 100mm, 1.8µm				
Oven (ºC)	40°C				
Pump	Mobile Phase A	0.1% Formic Acid in Water			
	Mobile Phase B	0.1% Formic Acid in Acetonitrile			
	Flow Rate	0.4 mL/min			
	Gradient	Time / (min)	%A	%B	
		0.0	99	1	
		7.0	62	38	
		9.0	1	99	
		10.0	1	99	
		10.1	99	1	
	Runtime (min)	14.0			
Injector	Volume (µL)	2			
MS	QTOF/QQQ/TOF	Mass Spec Type:	QTOF	Mode	+ve & -ve
	Source –Dual Jet Stream ESI	Gas temp	280°C	Gas flow	14 L/min
		Sheath Gas temp	350°C	Sheath gas flow	11 L/min
		Nebuliser pressure	310 kPa	Nozzle Voltage	100 V
		VCap	4000 V	Fragmentor	380 V

A further LC-MS method was developed to focus on the anthocyanin type compounds that can cause the differences between purple tea leaf samples. This LC-MS method can be found in **Table 2.2** where it was conducted only in positive mode.

Table 2.2. LC-MS Method for Purple Tea Leaf samples.

This was conducted in positive ionisation mode only.

HPLC-MS					
Instrument	Agilent 1290 Infinity II UHPLC + 6550 Q-ToF with iFunnel				
Column	Agilent Eclipse Plus C18 2.1mm x 100mm, 1.8µm				
Oven (ºC)	40°C				
Pump	Mobile Phase A	0.1% Formic Acid in Water			
	Mobile Phase B	0.1% Formic Acid in Acetonitrile			
	Flow Rate	0.4 mL/min			
	Gradient	Time / (min)	%A	%B	
		0.0	95	5	
		20.0	80	20	
		30.0	10	90	
		32.0	10	90	
		33.0	95	5	
	Runtime (min)	35.0			
Injector	Volume (µL)	5			
MS	QTOF/QQQ/TOF	Mass Spec Type:	QTOF	Mode	+ve
	Source –Dual Jet Stream ESI	Gas temp	200°C	Gas flow	14 L/min
		Sheath Gas temp	350°C	Sheath gas flow	11 L/min
		Nebuliser pressure	241 kPa	Nozzle Voltage	0 V
		VCap	3500 V	Fragmentor	380 V

2.5 Results

There are more than two sample groups and therefore Mass Profiler cannot be used as it only investigates '1 vs 1' sample groups.

2.5.1 All Tea Extract and Tea Leaf Samples

The analysis was checked and negative ionisation mode showed the most compounds of interest therefore became the focus of this part of the experiment. The samples shown in **Table 2.3** were all run on the method shown in **Table 2.1**.

Sample Table:

Table 2.3. Different tea samples used throughout.

Sample Name	Reference Number
Green Tea Extract	Std 604 (3 Repeats)
Green Tea Extract	Std 101
Green Tea Extract	Std 607K (4 Repeats)
Green Tea Extract	DF302 (3 Repeats)
Black Tea Extract	Std 652K
Purple Tea Extract	FEI 8000
Green Tea Leaf	LA1 : CA 635 pearl
Green Tea Leaf	LA2 : CA 635 60%
Green Tea Leaf	LA3 : CA 635 30%
Green Tea Leaf	LA4 : CA 635 open
Green Tea Leaf	LB1 : CA 609 pearl
Green Tea Leaf	LB2 : CA 609 60%
Green Tea Leaf	LB3 : CA 609 30%
Green Tea Leaf	LB4 : CA 609 open
Green Tea Leaf	LC1 : CG 29W8 pearl
Green Tea Leaf	LC2 : CG 29W8 60%
Green Tea Leaf	LC3 : CG 29W8 30%
Green Tea Leaf	LC4 : CG 29W8 open
Green Tea Leaf	SC1 : SC 12/28
Green Tea Leaf	SC2 : SF 32/186
Green Tea Leaf	SC3 : S 15/10
Green Tea Leaf	SC4 : CHM61/60
Green Tea Leaf	SC5 : U864
Green Tea Leaf	SC6 : SMK 30/52
Green Tea Leaf	SC7 : KPT 7/124
Purple Tea Leaf	PURP : TRI 306
Albino Tea Leaf	ALB : Albino Tea
Purple Leaf	TRI 306 (5 Repeats)
Purple Leaf	JFK CG1 (5 Repeats)
Purple Leaf	JFK 27/1 (5 Repeats)
Purple Leaf	JFK 27/4 (5 Repeats)
Purple Leaf	JFK 27/6 (5 Repeats)
Yellow Tea Extract	Lot CN18-37
Yellow Tea Extract	Lot CN18-50
Yellow Tea Extract	Lot CN18-103
Yellow Tea Extract	Lot CN19-06

2.5.1.1 Mass Profinder

Mass Profinder was conducted to extract the data as .cef files for use in MPP, however given the many sample groups, it is not ideal to conduct this comparison analysis using the data tables.

Recursive Feature Extraction (RFE) was performed on the batch of all data files in Mass Profinder. It first performs Molecular Feature Extraction, which involves chromatographic deconvolution and aligning the features across the selected data files using mass and retention time. It then uses the mass and retention time of each feature to perform targeted feature extraction across the whole batch of samples to compare each file against each other.

The generalised parameters include picking peaks with a height above 500 counts and a charge state of 1 – 2. It aligns each feature with an across the batch allowance of +/- 0.3 min retention time difference and 20 ppm mass difference, but only those features with a height of $\geq$1% of the relative height from the largest peak. Each feature must be present in at least 2 files in at least 1 sample group.

2.5.1.2 MPP

Using the sample data from Mass Profinder, the deconvoluted samples were exported as .cef files and entered into MPP. Some further filtering of samples was conducted but this was minimal. The minimum abundance of peaks used was raised to minimum of 1500 counts but still used only those features present in 100% of samples in at least one sample group. The feature had to be present in at least one sample due to some samples only having one data file in the group.

PCA Plot

The PCA plot can be seen in **Figure 2.4** with a key for the diagnosis of the three sample groups for each different colour.

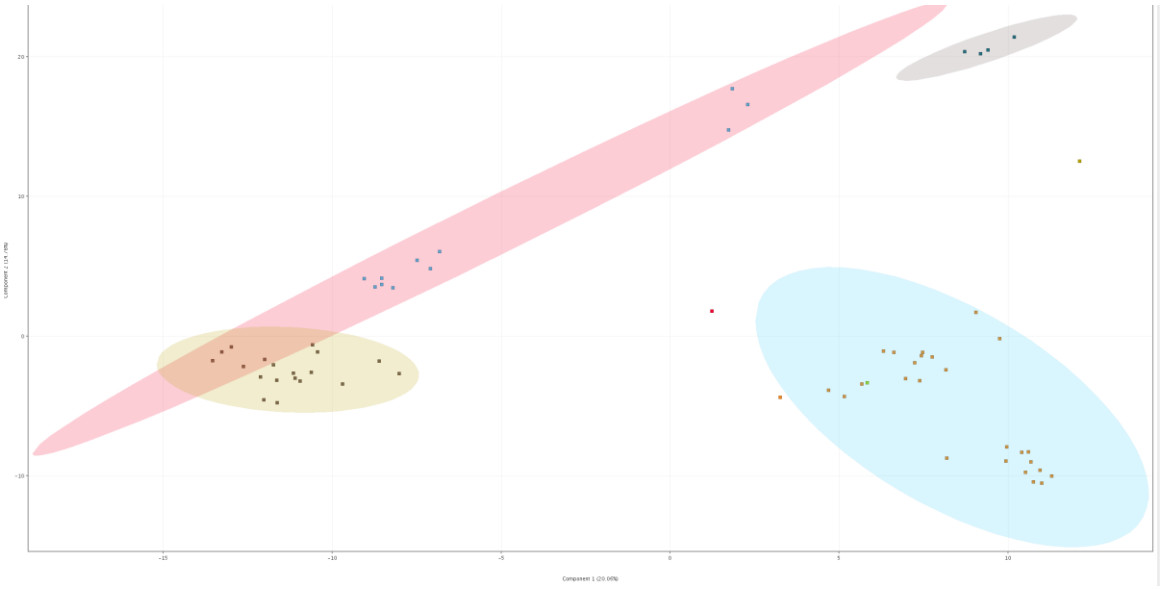

Figure 2-4. 2D PCA plot showing all tea samples.
The x-axis shows the largest variance of 20.06% and the y-axis shows the next largest variance at 14.78%. The samples have 95% confidence ellipses round them and can be seen to be clustered into distinct groups corresponding to their sample types.

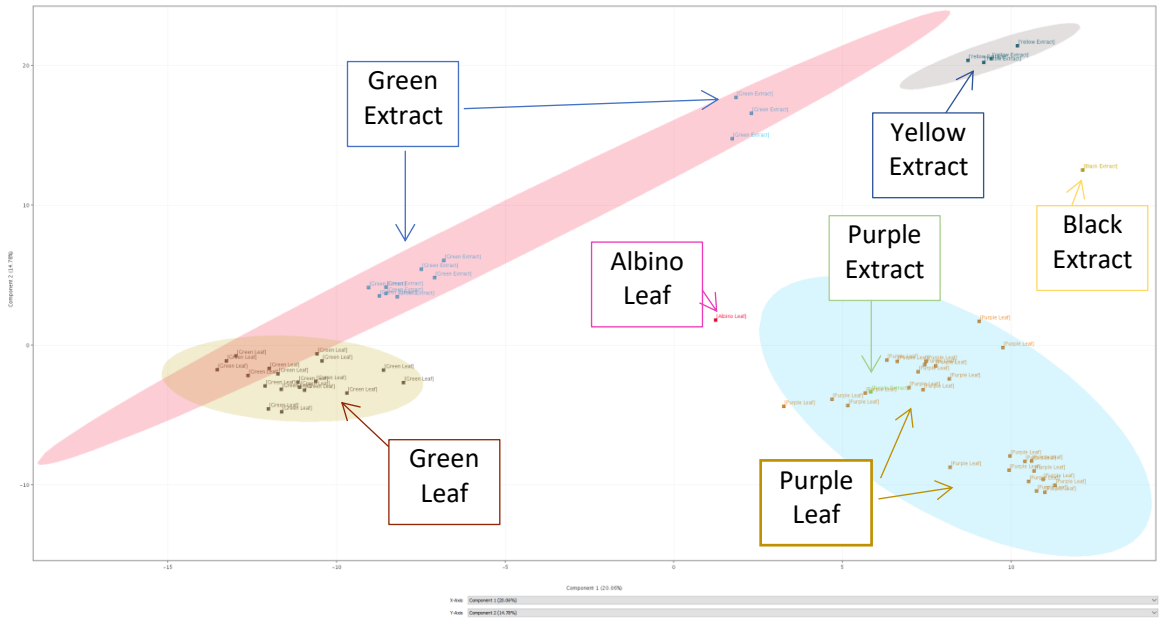

Figure 2-5. Labelled 2D PCA plot showing all tea samples.

As visible in the PCA plots in **Figures 2.4 – 2.5**, the repeats of the same 'type' of tea are all tightly grouped and therefore show stability in the instrument and method. There are clear clusters within some of these groups showing differences between different groups and growing conditions. Even

the green extracts and leaf samples overlap showing that they're similar, despite the changes in chemical processes between them and concentrations of the compounds between the extract and leaf. This is the same for the purple tea leaf and extracts.

There is a difference within the green extract groups and 4 groups within the purple leaves, even though the purple leaves are 5 repeats of 5 samples. The green tea extracts have two different clone names in the sample names also, suggesting the reason for this split within the blue dots. The PCA shows that each of these repeats are tightly grouped showing that the difference in genuine within the samples, not from the instrument or method. The differences between clones are due to factors such as concentrations of contents (i.e., polyphenol content, catechins) which can be caused by differences like soil contents and shading.

The main difference seen along the x-axis is between the purple leaf and the green leaf, with the greatest variance of 20.06%. The y-axis, representing the next largest variance of 14.7%, shows the spread between the yellow extract and the purple leaf. The PCA plot shows distinct grouping of each sample type however, the leaf and extracts are prepared in slightly different ways. Hence, the comparison between them cannot be truly determined. Therefore, MPP was repeated but only looking at the tea leaf samples.

2.5.2 Only Tea Leaf Samples

The samples used are those from **Section 2.5.1**, which were run on the LC method from **Table 2.1**.

2.5.2.1 MPP

PCA Plot

The PCA plot can be seen in **Figure 2.6** with a label of the three sample groups for each different colour. The samples are only the tea leaves, looking at the differences between purple tea leaves, green tea leaves and the one albino tea leaf.

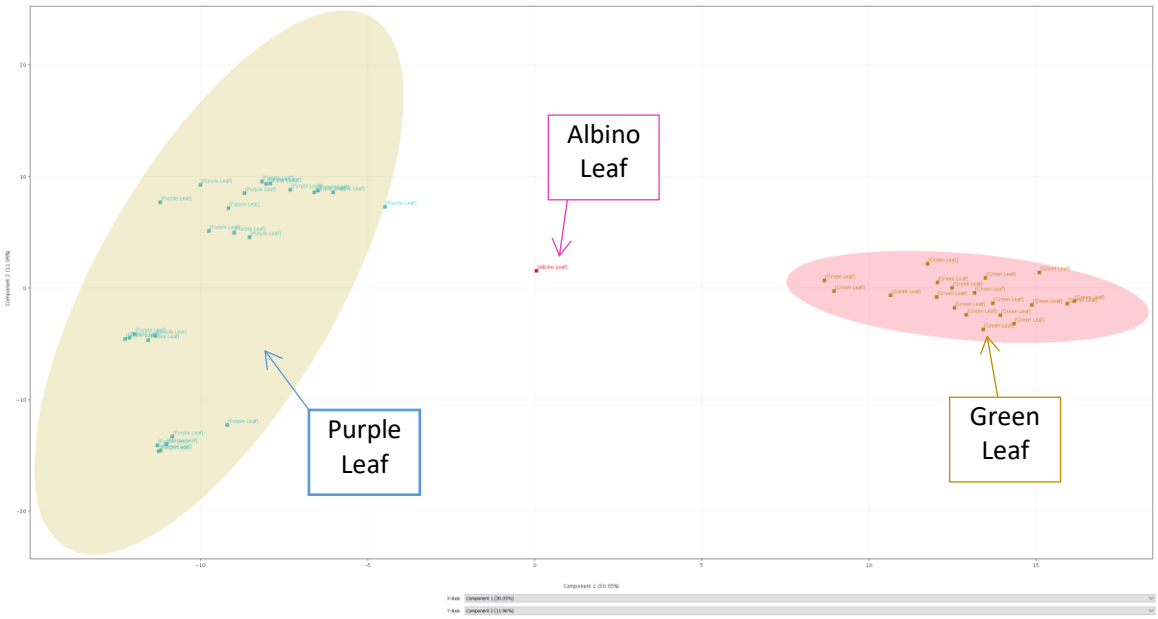

Figure 2-6. MPP PCA Plot of all tea leaf samples.

It shows that the samples are all distinctly grouped, with some further separation within the purple tea leaf sample group also. The key is shown in this figure that relates each coloured dot to the sample group. The x-axis represents 30.05% variance and y-axis 11.96%.

The PCA plot shows that the largest variance, represented along the x-axis (principal component 1), is the difference between the purple and green leaf sample groups. The albino tea leaf is also shown on these plots directly in between the two sample groups, however, this cannot be confirmed without repeat samples. The green and purple tea leaf groups form its own cluster.

Figure 2.7 shows a PCA plot of only the purple and green tea leaf samples.

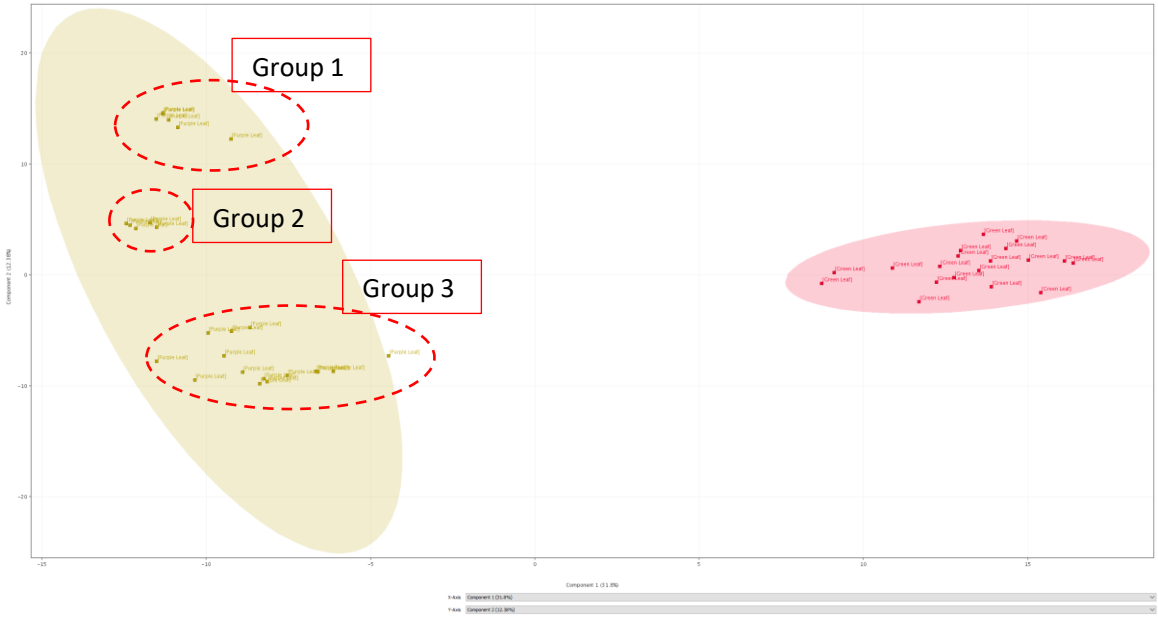

Figure 2-7. PCA Plot, labelled, of only purple and green tea leaf samples.

It shows the separation within the purple tea leaf samples, all of which are clustered into individual groups, with some sample groups showing closer relation than others.

The green leaf samples do not cluster into distinct groups as the purple tea leaf samples do. The purple tea leaf samples seem to form three groups, of which group 3 seems to be multiple clones similar to each other (JFK 27/1, JFK 27/4 and JFK 27/6). Group 1 is one sample group of five repeats, and group 2 is also one sample group of five repeats. The compounds causing this difference is of interest due to the nature of the sample's potential worth, it would be a disadvantage to have less polyphenols and anthocyanins in some clones in comparison to others.

2.5.3 Further Investigation into the Difference Between Purple Tea Leaf Samples

It was of interest to determine how different the clones of purple tea are to each other to see which one holds the most value for the tea market, with respect to the health benefits. Using the comparison software platforms, the objective is to create a list of compounds that show the differences between the purple tea leaf clones. This will hopefully give more insight into which sample contains the most compounds that give the most interesting health benefits. Two repeats of

each sample were run on the LC method shown in **Table 2.2** in positive ionization mode to investigate whether the anthocyanin content is as different between the samples.

For this analysis, a Yorkshire tea sample and a purple tea extract were also added to see how different they are from each other in the preliminary analysis.

2.5.3.1 MPP

PCA Plot

The **Figure 2.8** shows the 2D PCA plot.

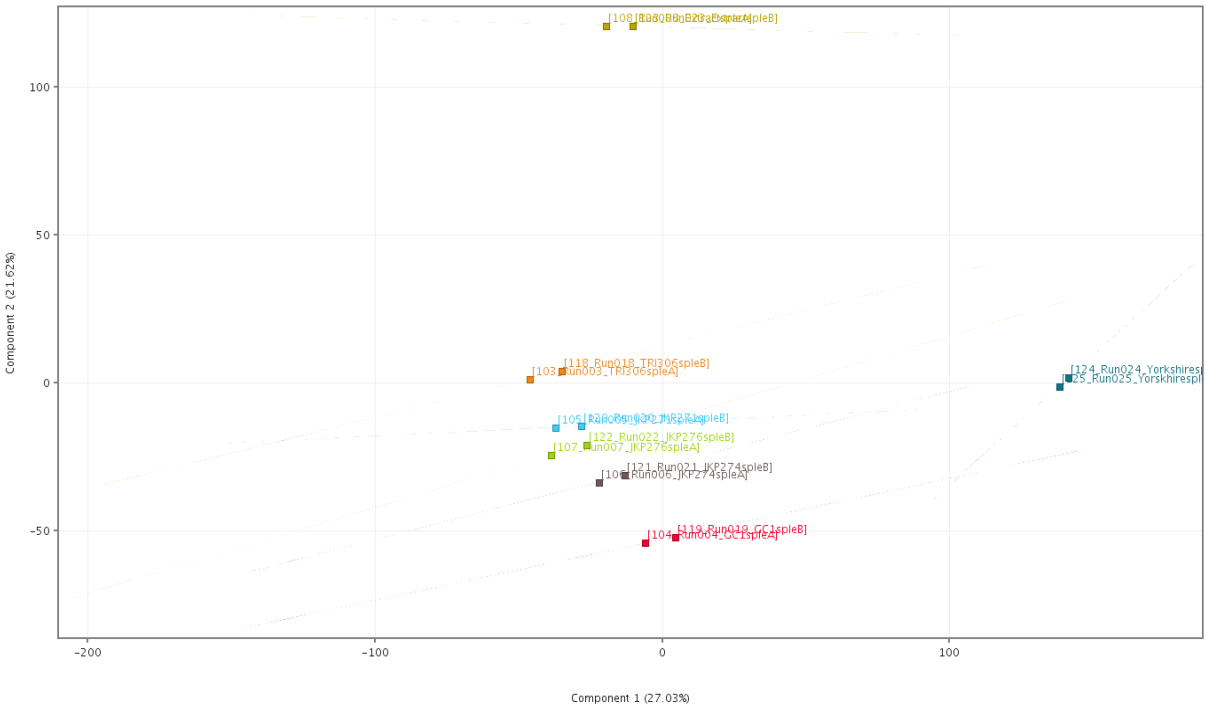

Figure 2-8. 2D PCA plot of all samples run in positive mode. PC1 represents 27.03% variance and PC2 represents 21.62% variance.

The purple tea leaves are clustered closer together, with the Yorkshire tea sample spread to the right on the x axis. The purple tea extract is spread up the y-axis. There are some very small 95% ellipses barely visible on the plot due to only two repeat samples in each group. The samples were run in a random order and show close clustering within sample group.

The PCA plot from **Figure 2.8** is repeated below in **Figure 2.9** however is now labelled. The software does not allow for zoomed labelling; therefore, the labels are extremely difficult to read, hence the need for hand labelling. This then allows for the differences between the types of tea to be seen more, and view how they are grouped.

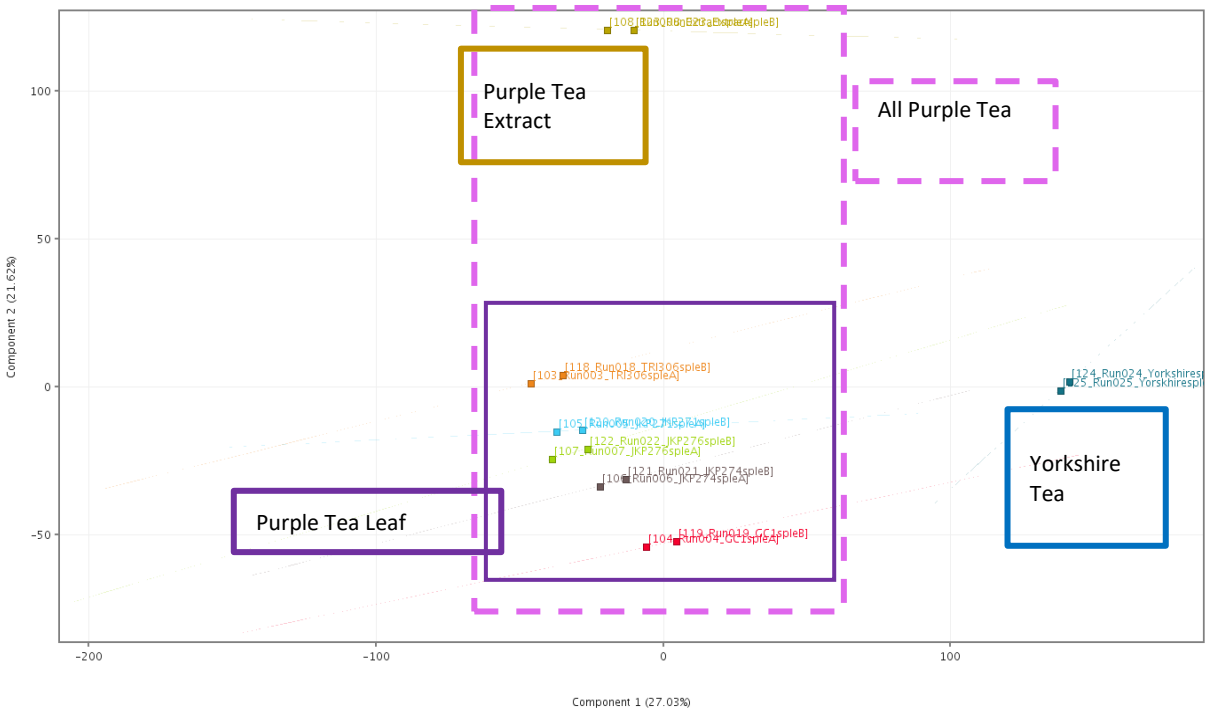

Figure 2-9. MPP 2D PCA Plot on all purple tea leaf, extract and Yorkshire tea samples.

The purple tea can be seen to be clustered within one large group but are spread from the Yorkshire tea sample along the x-axis. The purple tea is separated up the y-axis, separated into the tea leaf and the tea extract.

Figure 2.9 shows the most different samples are the Yorkshire tea compared to all the purple tea samples. This is due to Yorkshire tea being black tea, where-as across the x-axis, the purple tea samples are all fairly close together. They are spread up PC2 which is representing nearly 22% difference between the purple tea leaves and the purple tea extract.

Figure 2.10 shows a zoom of the purple tea leaf samples on the 2D PCA scores plot.

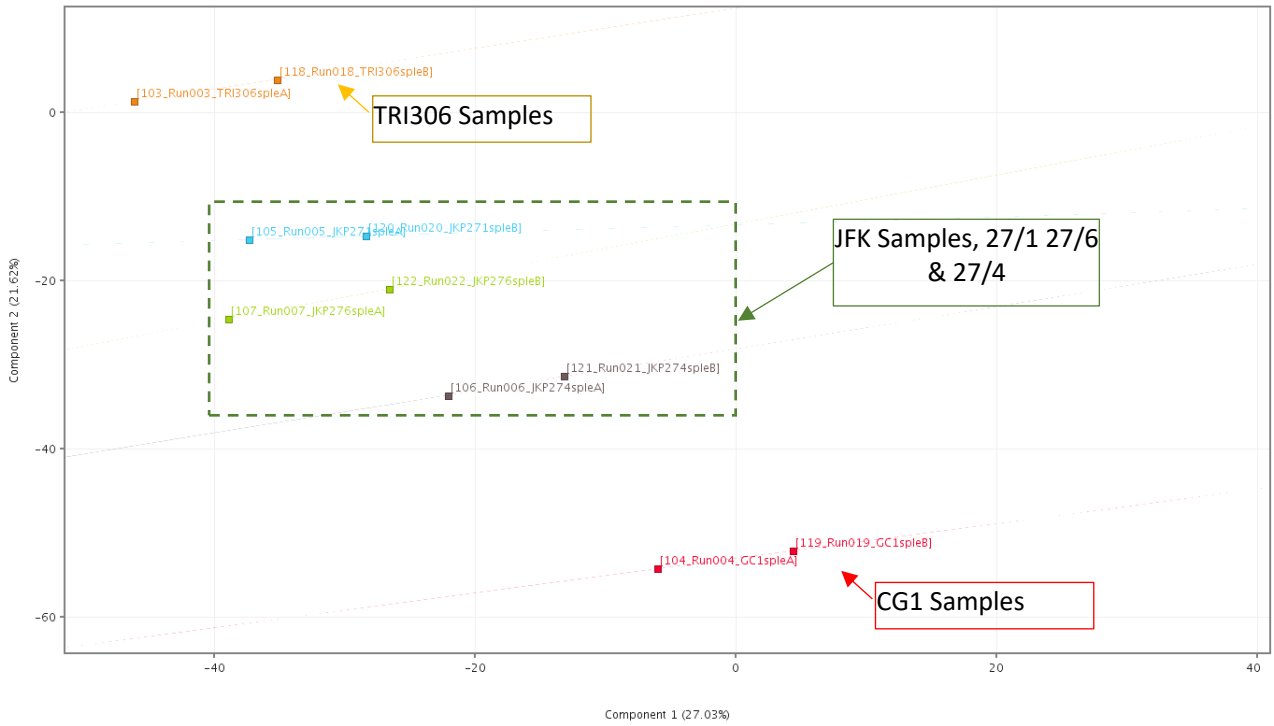

Figure 2-10. MPP 2D PCA plot zoom on only the purple samples.

It shows the two extremes along both x and y-axes are CG1 and Tri306 meaning they're the most different to each other.

Figure 2.10 focuses only on purple tea leaves and shows the two extremes on PC2 are CG1 and Tri306, which are the two parent clones. The JKP samples are blends of these two samples which is why they sit in the middle of the two parents. It also makes sense that the purple extract is most different to the purple tea leaves as due to the difference in sample preparation, they will have different concentrations of the content. The extract is thought to be 4 times more concentrated than the leaf.

Hierarchical Clustering

Cluster analysis is another way to determine which sample groups are most similar to each other. **Figure 2.11** shows the overview of the hierarchal clustering.

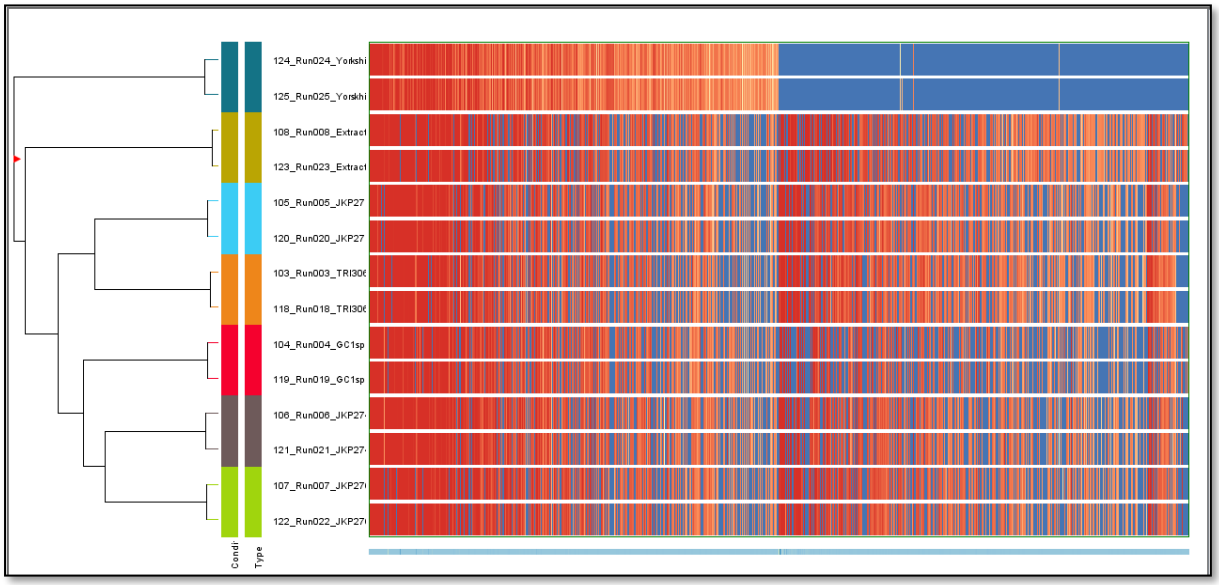

Figure 2-11. MPP Hierarchical Clustering on purple tea – large overview.

The **Figure 2.11** shows the hierarchal clustering of the samples. The red – blue lines on the right hand side each represent a feature, with the red lines showing high abundance and blue lines showing absence of the feature. It can be seen that the Yorkshire extract has about half of the features absent that is present in all the other samples, which emphasizes the difference between the black and purple tea samples.

Figure 2.12 shows a zoom of the clustering tree on the left-hand side of the hierarchal cluster plot.

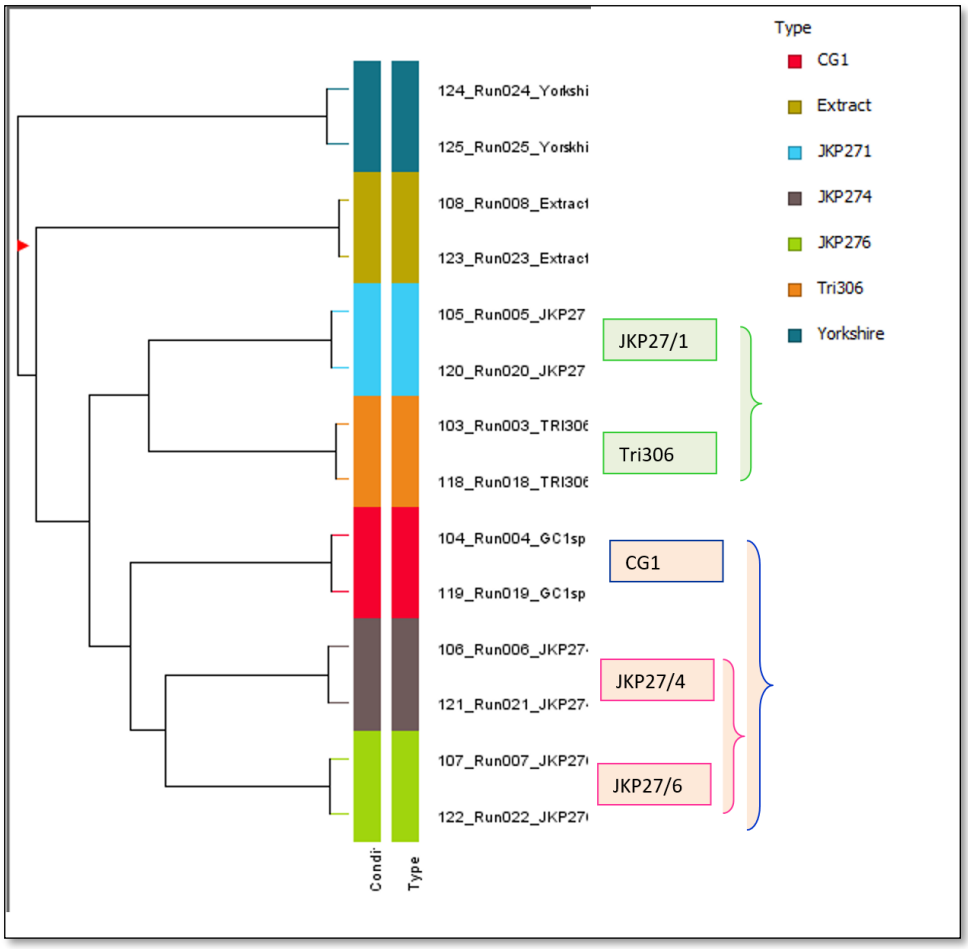

Figure 2-12. MPP Hierarchical Clustering on purple tea.

This shows that the JKP27/1 and Tri306 samples are clustered, the JKP27/4 and JKP27/6 samples are clustered and the CG1 sample is closer related to the JKP27/4 and JKP27/6 samples than to the others.

The clustering in **Figure 2.12** (which is a zoom of the cluster tree from **Figure 2.11**) can be useful to use to back up the PCA plot analysis on the similarities on sample grouping and therefore, you can be more confident in the conclusions of the analysis. This cluster plot in **Figure 2.12** does not give any extra information than what the PCA plots have given. It confirms that the samples JKP27/4 and JKP27/6 are most like each other, and then cluster together with CG1. This leaves JKP27/1 and Tri306 to be grouped together. This clustering is the same as they cluster in the PCA plots also. The Yorkshire tea and extract are furthest from the other samples, as also shown on the PCA plots. This matches the assumptions made that the differences are due to the Yorkshire tea being a black tea

where-as the others are all purple, and the extract is more spread as it will have different concentrations of components to the tea leaves.

Venn Diagram of Unique Compounds

When there are more than 4 sample groups, the Venn diagram for unique compounds turns into a chart. The basis is still the same, though it is not as visual. Below in **Figure 2.13** it can be seen that the extract has most unique features, then Yorkshire tea and Tri306. The other sample groups are fairly similar to each other with 5 or less unique features.

Condition	Count
Entities occurring in exactly 1 condition	
[Extract]	41
[GC1]	3
[JKP271]	2
[JKP274]	4
[JKP276]	5
[Tri306]	15
[Yorkshire tea]	28

Figure 2-13. Chart for Unique Compounds to Each Sample Group

Feature Finding Graphs

Figure 2.14 shows a feature finding graph on significant features with a *p*-value less than or equal to 0.01, and a fold change above 30.

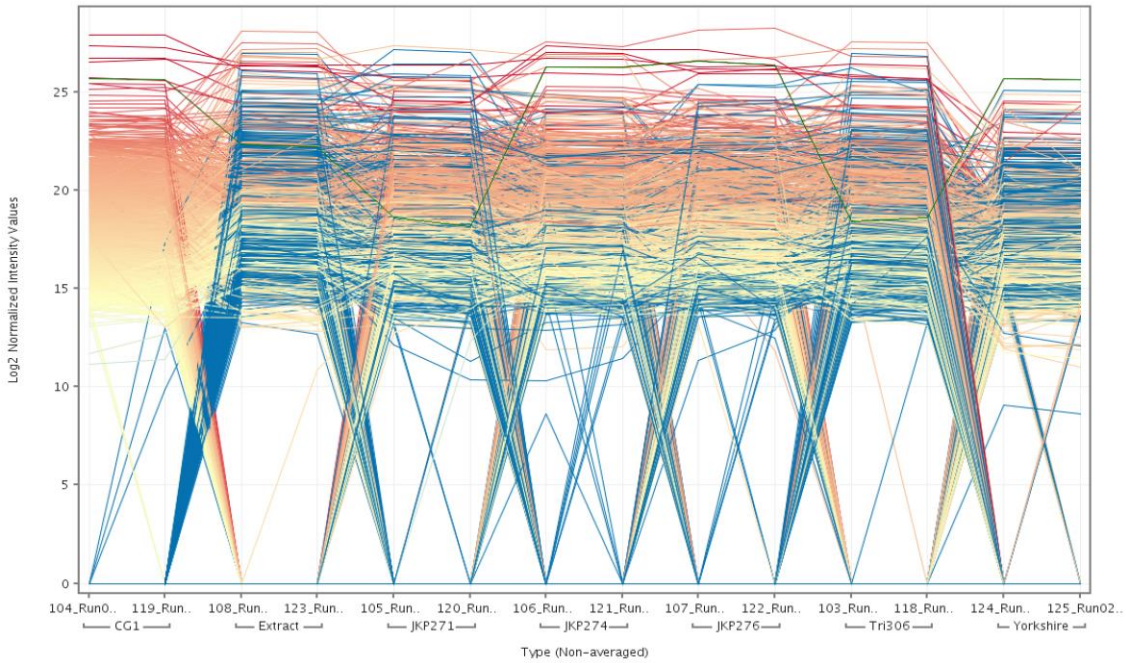

*Figure 2-14. MPP Finding Feature Graphs on all features with a p-*value *≤ 0.01 and fold change > 30. There are >1000 features shown on the graph and this will make it difficult to retrieve useful information from this chart.*

As shown in the graph in **Figure 2.14**, there are many features that fit into this category and it looks messy and hard to interpret. The features can be further filtered by abundance where only the features above a normalised abundance of 25 are included. The most abundant feature in this analysis has a normalised abundance of 28.5 and therefore this is a highly filtered selection.

Figure 2.15 shows the feature finding graph but only with features that have a maximum normalised abundance of 25.

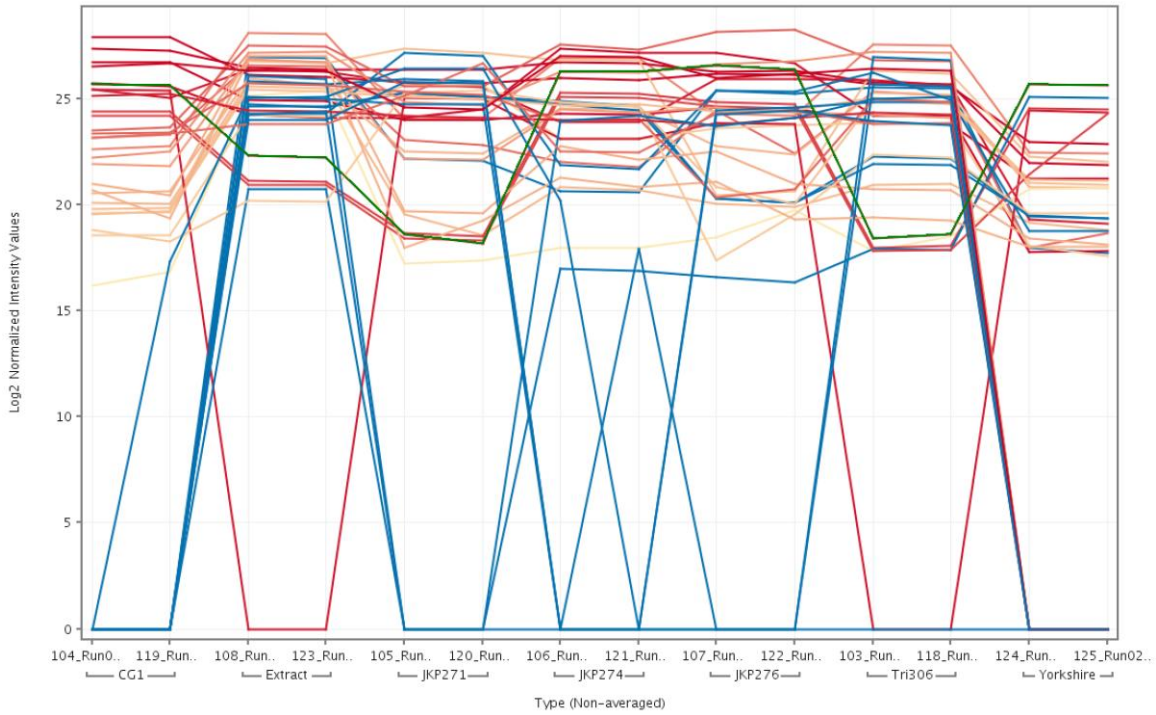

*Figure 2-15. MPP Finding Feature Graphs of features with a p-*value *≤ 0.01 and fold change > 30 AND a maximum normalised abundance of 25. This is a lot more simplified than Figure 2.15 and is better to see useful trends in the data.*

Though this is a highly filtered selection of the significant features, it is clearer to see patterns in the data. The red lines are those that are of highest abundance in at least one group, and there is one particular red line that can be seen to be in high abundance in all sample groups though it is not present in the purple extract and Tri306. Hovering over this feature gives the mass, RT and the identification assigned, of which is shown in **Figure 2.16**. Unfortunately, there are no better ways to visualise what feature the line represents than this; the next best way to determine which features are of significance is exporting the data as a feature table. The identification has originated from the ID browser identification where the Metlin 8.0 database was used. This graph is a useful tool for an overview of trends in data; however it is still difficult to see the full trends with so many lines on the same plot.

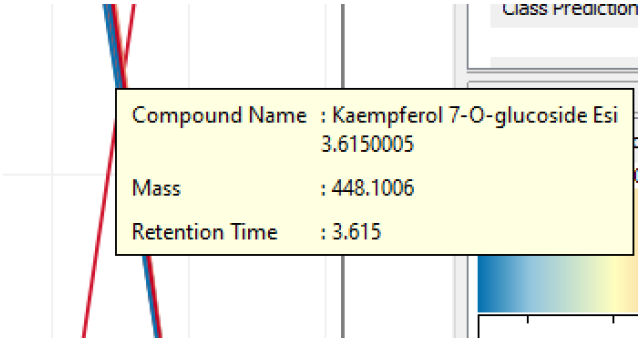

Figure 2-16. MPP Feature identification of selected line from Figure 2.52.

This feature is present in all samples except Tri306 and the purple tea extract.

The MPP software can also give a table of the features from the selected plot in **Figure 2.15**. For example, each line on the graph that has a positive identification in **Figure 2.15** is represented in **Table 2.4**. The limitation is that the table does not show feature abundances in the different groups, and therefore it is only saying that these features are significantly different between the sample groups and not in which group it is most significant in. It also means that the features cannot be ordered by significance.

Table 2.4 shows a list of features that have been annotated by Metlin 8.0 database using MS only results with a Metabolomics Standard Initiative (MSI) reporting level of 2. These features are those that show significant difference between the sample groups. For identifications, standards would be needed for confirmation.

Table 2.4. MPP Annotations on filtered compounds in Figure 2.15. PubChem Compound CID is given for each annotation.

Putatively Annotated Compounds	Compound CID	Retention Time (min)	Mass (Da)
Quercetagitrin	5320826	7.918	480.0902
Dihydroferulic acid 4-O-glucuronide	190069	3.349	372.1055
3'-Galloylprodelphinidin B2	15593122	2.582	762.1442
Styrene	7501	1.139	104.0624
Myricetin	5281672	7.916	318.0376
Pelargonidin 3-(2glu glucosylrutinoside)	131751479	12.774	740.2158
Leucodelphinidin 3-O-alpha-L-rhamnopyranoside	44257158	6.229	468.1245
(-)-Epigallocatechin 3-(4-methyl-gallate)	401129	8.135	472.1000
Delphinidin 3-glucoside	443650	2.563	464.0953
Kaempferol	5280863	15.308	286.0470
Coronarian	441560	2.222	382.0877
Kaempferol 3-O-β-D-glucosyl-(1-2)-β-D-glucoside	6325460	10.196	610.1541
Kaempferol 3-[2''-(6'''-coumaroylglucosyl)-rhamnoside] 7-glucoside	131752764	21.302	902.2471
Kaempferol 7-O-glucoside	10095180	4.379	448.1003
Kaempferol 3-[2''-(6'''-coumaroylglucosyl)-rhamnoside] 7-glucoside	131752764	21.302	902.2469
Kaempferol 7-O-glucoside	10095180	3.615	448.1006
3-Caffeoylpelargonidin 5-glucoside	131752292	14.059	594.1372
Epigallocatechin gallate	65064	6.437	458.0842
3-Caffeoylpelargonidin 5-glucoside	131752292	15.776	594.1365
3,3'-Di-O-galloylprodelphinidin B5	13270037	4.649	914.1534
3-Caffeoylpelargonidin 5-glucoside	131752292	12.335	594.1369

Fold-Change Differences

The major differences are between CG1 and TRI306 as these are the two parent clones. The top 10 differences with the largest fold change are shown below with some possible identifications. These differences are done via an export of the fold change features of Tri306 vs CG1, using those features with a fold change greater than 30 only. These fold change features have then been exported into ID browser to search for possible identifications using Metlin 8.0 database. These possible identifications can be seen alongside the neutral mass and RT in **Table 2.5 – 2.6** below.

Table 2.5 shows compounds up regulated in Tri306 sample group:

Table 2.5. MPP Compounds up regulated in Tri306 compared to CG1. PubChem Compound CID is given for each annotation.

Up Regulated in Tri306	Putatively Annotated Compounds	Formula	Compound CID
616.0695@6.855	-	-	-
448.1003@4.379	Kaempferol 7-O-glucoside	$C_{21}H_{20}O_{11}$	10095180
372.1055@3.349	Dihydroferulic acid 4-O-glucuronide	$C_{16}H_{20}O_{10}$	190069
594.1369@12.335	-	-	-
594.1365@15.776	-	-	-
658.0768@6.044	-	-	-
456.1045@12.546	Epicatechin 3-O-(4-methylgallate)	$C_{23}H_{20}O_{10}$	467297
302.0428@13.261	Melanoxetin	$C_{15}H_{10}O_7$	15560442
610.1298@14.327	Theasinensin C	$C_{30}H_{26}O_{14}$	467317
288.0634@8.134	Fustin	$C_{15}H_{12}O_6$	5317435

Table 2.6 shows compounds up regulated in CG1 sample group:

Table 2.6. MPP Compounds up regulated in CG1 compared to Tri306. PubChem Compound CID is given for each annotation.

Up regulated in CG1	Putatively Annotated Compounds	Formula	Compound CID
610.1541@10.196	Kaempferol 3-O-β-D-glucosyl-(1->2)-β-D-glucoside	$C_{27}H_{30}O_{16}$	10652679
1066.1957@6.694	-	-	-
470.2119@21.369	-	-	-
1064.1833@2.966	-	-	-
430.162@10.015	8-Acetoxy-4'-methoxypinoresinol	$C_{23}H_{26}O_8$	73830447
388.169@3.148	Asn Gln Gln	$C_{14}H_{24}N_6O_7$	145454081
286.0475@12.769	Kaempferol	$C_{15}H_{10}O_6$	5280863
450.1147@6.023	-	-	-
302.0423@11.556	Quercetin	$C_{15}H_{10}O_7$	5280343
320.0514@3.459	2,2',3-Trihydroxy-3'-methoxy-5,5'-dicarboxybiphenyl	$C_{15}H_{12}O_8$	15608168

Compounds that are of interest are anthocyanins or polyphenols, particularly flavonoids such as quercetin, myricetin and kaempferol. Parent anthocyanin compounds are those such as pelargonidin and delphinidin, with the anthocyanidin glucosides/galactosides being a known component of tea. These classes of compounds are of interest due to their extensive research on their health benefits. The following polyphenol compounds in **Table 2.7 – 2.8** are some of those that significantly different between Tri306 and CG1.

Table 2.7 shows compounds up regulated in CG1 sample group:

Table 2.7. MPP Compounds up regulated in CG1 compared to Tri306

Putatively Annotated Compounds	[Tri306] avg peak area	[CG1] avg peak area	Log FC
Kaempferol 3-O-β-D-glucosyl-(1-2)-β-D-glucoside	0	44925584	-25.4
Kaempferol	0	6256529	-22.6
Quercetin	0	5397569	-22.4
Catechin 7-O-gallate	0	3780085	-21.8
Myricetin 3-galactoside	0	3137302	-21.6
3'-Galloylprodelphinidin B2	0	2620951	-21.3
Quercetin	0	1088238	-20.1
Rhamnocitrin 3-glucosyl-(1-2)-galactoside	0	643951	-19.3
(-)-Epicatechin 7-O-glucuronide	0	559682	-19.1
Pelargonidin 3-(6-malonylglucoside)-7-glucoside	0	403666	-18.6
Quercetin	0	369446	-18.5
Isotheaflavin 3'-gallate	0	340725	-18.4
Quercetin 3-sulfate-7-alpha-arabinopyranoside	0	280951	-18.1
Myricetin 3-sambubioside	0	222831	-17.8
Ellagic acid	0	173595	-17.4
Delphinidin 3-(6''-malonylglucoside) 5-glucoside	0	146480	-17.2
Quercetin 3-(3'',6''-di-*p*-coumarylglucoside)	0	132249	-17.0
Luteolin 7-O-glucuronide	0	100685	-16.6
Theasinensin C	0	73513	-16.2
Theaflavin digallate	0	63372	-16.0
Pelargonidin 3-(2glu glucosylrutinoside)	375981	52910792	-7.1

Table 2.8 shows compounds up regulated in Tri306 sample group:

Table 2.8. MPP Compounds up regulated in Tri306 compared to CG1

Putatively Annotated Compounds	[Tri306] avg peak area	[CG1] avg peak area	Log FC
Kaempferol 7-O-glucoside	50297004	0	25.6
Dihydroferulic acid 4-O-glucuronide	47968340	0	25.5
Epicatechin 3-O-(4-methylgallate)	17845418	0	24.1
6-Hydroxydelphinidin 3-glucoside	4643779	0	22.1
Myricetin 3,7,3',4'-tetramethyl ether	1797311	0	20.8
Quercetin 3-rutinoside-3'-apioside	1107615	0	20.1
Prodelphinidin A2 3'-gallate	1007481	0	19.9
Quercetin 3-rhamnoside-3'-sulfate	855103	0	19.7
Quercetin 3-(2''-*p*-hydroxybenzoyl-4''-*p*-coumarylrhamnoside)	832081	0	19.7
Catechin 4'-O-gallate	524921	0	19.0
Quinoline	508374	0	19.0
Procyanidin C1 3,3',3''-tri-O-gallate	397035	0	18.6
Gallic acid	384691	0	18.6
Epicatechin-(2beta-5,4beta-6)-ent-epicatechin	317048	0	18.3
Kaempferol 3-(2''-rhamnosylgalactoside) 7-rhamnoside	256849	0	18.0
Urolithin C	230221	0	17.8
Petunidin 3-(6''-*p*-coumarylglucoside)-5-glucoside	225994	0	17.8
Myricetin 3-(2''-*p*-hydroxybenzoylrhamnoside)	224016	0	17.8
3,5-Digalloylepicatechin	217149	0	17.7
Quercetin 3-rhamnoside-3'-sulfate	203495	0	17.6
Quercetin 3-(2G-rhamnosylgentiobioside)	166468	0	17.3
Kaempferol 3-glucuronide-7-sulfate	164815	0	17.3
Petunidin-3,5-diglucoside	141336	0	17.1
Quercetin 3-(2'''-feruloylsophoroside)	115477	0	16.8
Malvidin 3-(6-coumaroylglucoside) 5-glucoside	114577	0	16.8
Myricetin 3,7-diglucuronide	111044	0	16.8
Malvidin 3-(6-acetylglucoside)	99990	0	16.6
Quercetin 3-(2''-galloyl-alpha-L-arabinopyranoside)	92251	0	16.5
Delphinidin 3-(2''-galloylgalactoside)	69593	0	16.1
Quercetagetin 7-methyl ether 3-(2'''-caffeoylglucosyl)-(1-2)-glucuronide	48895	0	15.6
Epigallocatechin 3-O-(4-hydroxybenzoate)	40195	0	15.3
Peonidin 3-(6''-acetylglucoside)	34560	0	15.1
Quercetagetin 4'-methyl ether 7-(6-(E)-caffeylglucoside)	162823	2469	6.0
Pelargonidin 3-(6-*p*-coumaroyl)glucoside	34483204	798276	5.4

Overall, it seems that the Tri306 has a larger concentration of interesting polyphenols as there is a larger variation of the different anthocyanidin glucosides and catechins etc. Other analysis confirms

that Tri306 has more beneficial 'tea compounds', hence why the JKP clones have more Tri306 than CG1, explaining why they sit closer to Tri306 on the cluster analysis and PCA plots.

2.5.3.2 Mass Profinder

Mass Profinder is great for seeing visual changes in the samples. The table can be useful to see features present in only one sample group. Some examples of compounds that are showing significant differences throughout the samples are shown below throughout **Figures 2.17 – 2.19**.

Figure 2.17 shows EIC (373.1131) - neutral Mass (372.1055) & RT (3.339min) – Absent from CG1 only.

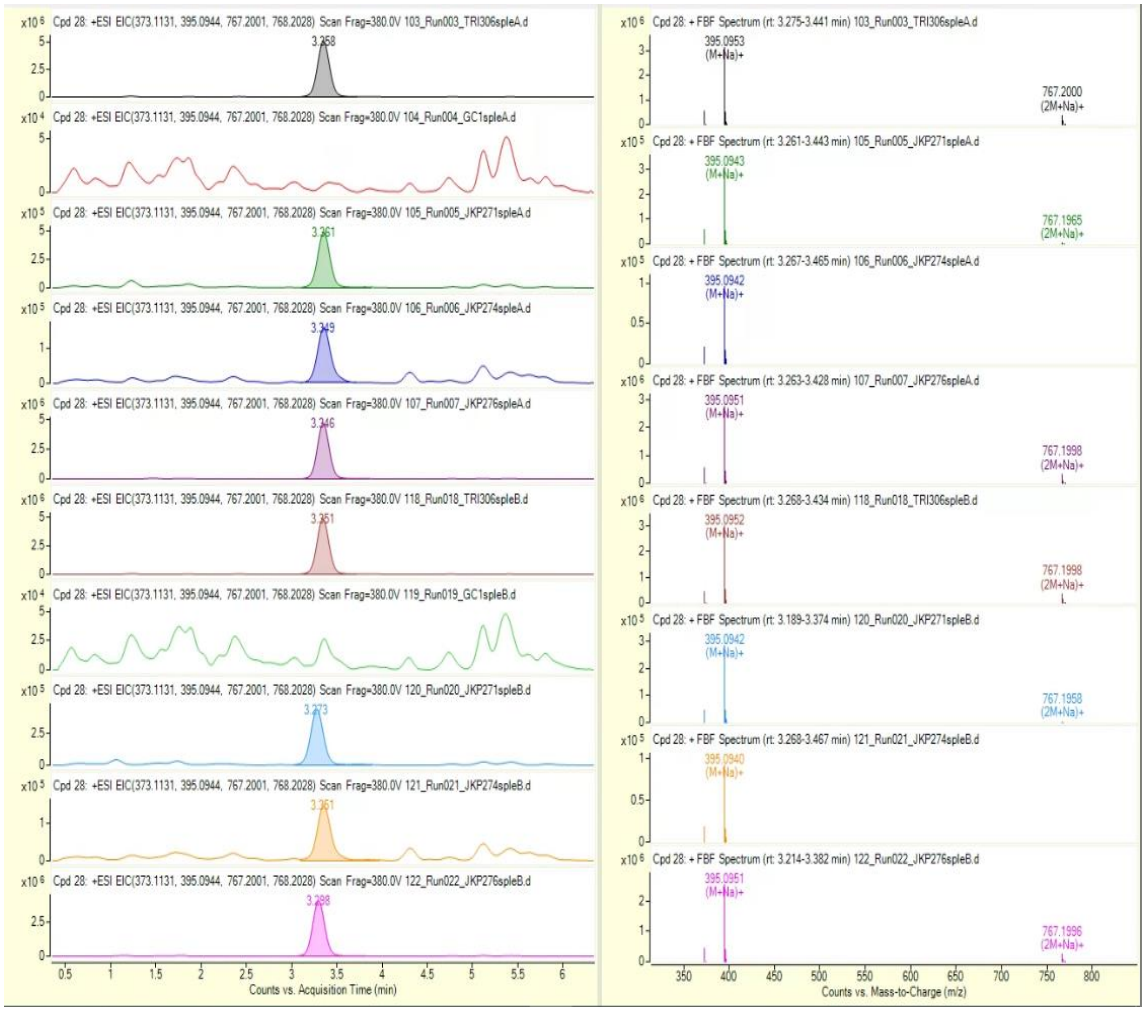

Figure 2-17. Profinder EIC (373.1131) example plot

This **Figure 2.17** shows that the feature with a neutral mass of 372.1055 at 3.3 minutes is present in all samples except CG1 repeats. The mass spectra with the protonated, sodiated mass and corresponding dimer masses, are seen on the right-hand side of the plot.

Figure 2.18 shows EIC (315.0129) - neutral Mass (314.0058) & RT (13.668 min) – Absent from JKP27/1 and JKP27/6

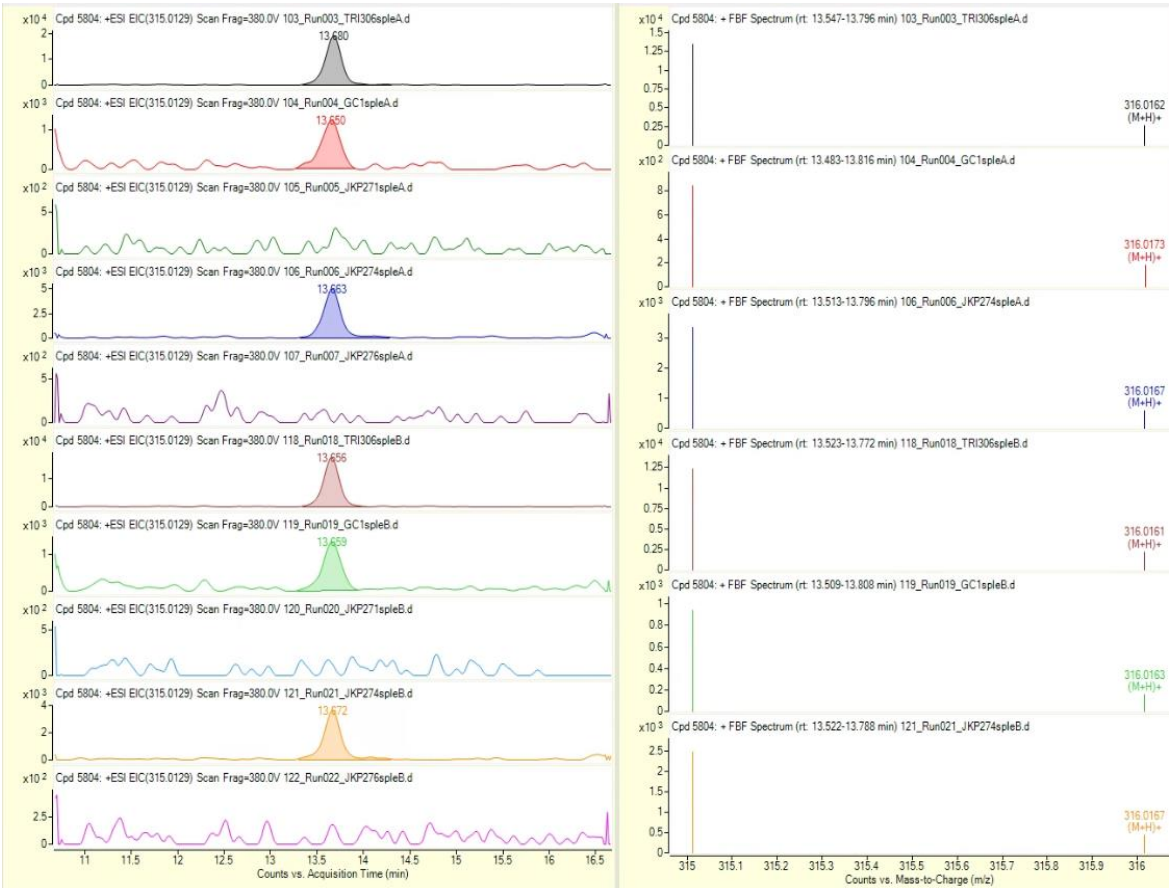

Figure 2-18. Profinder EIC (315.0129) example plot

This **Figure 2.18** shows that the feature with a neutral mass of 314.0058 at 13.7 minutes is present in all samples except JKP27/1 and JKP27/6 repeats. The mass spectra with the assumed protonated mass isotopes can be seen on the right-hand side of the plot.

Figure 2.19 shows EIC (171.0284) - neutral Mass (170.0212) & RT (6.011 min) – Absent from CG1, JKP27/4 and JKP27/6

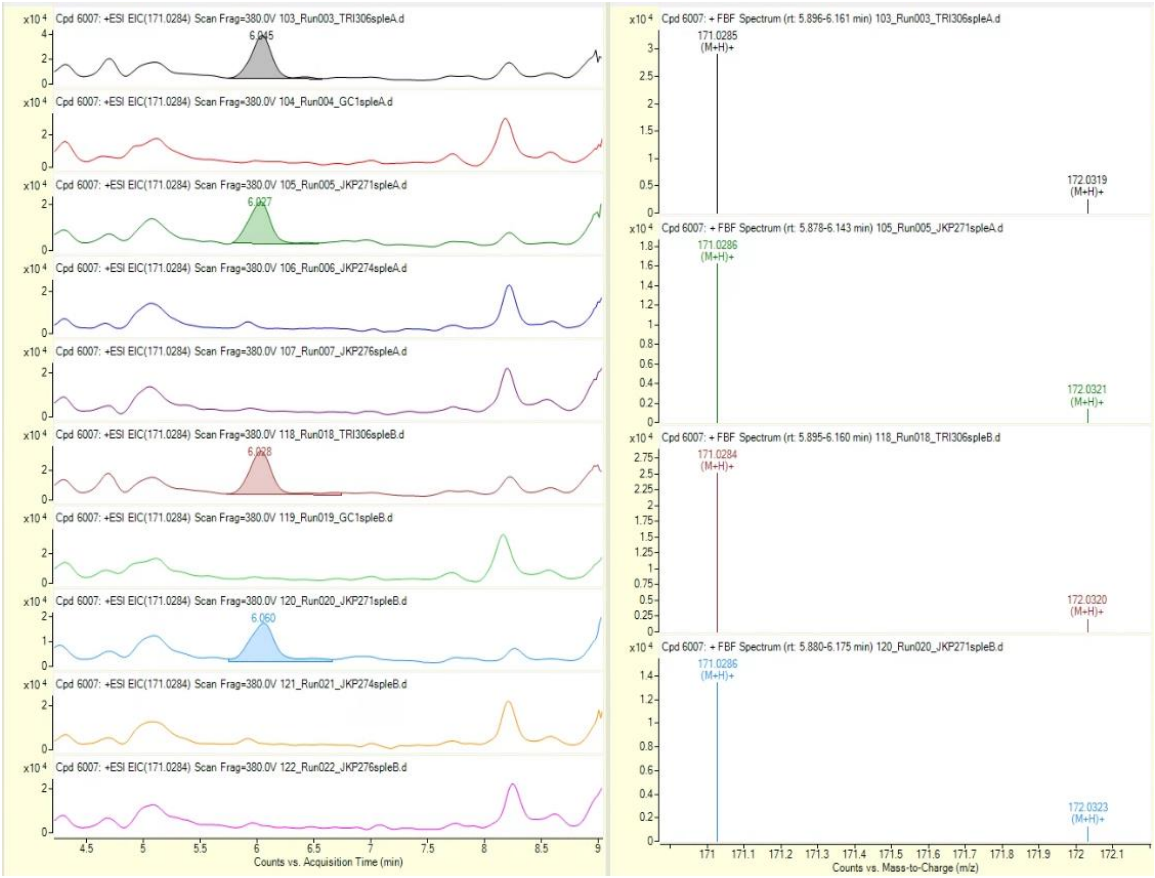

Figure 2-19. Profinder EIC (171.0284) example plot

This **Figure 2.19** shows that the feature with a neutral mass of 170.0212 at 6.0 minutes is present in all samples except CG1, JKP27/4 and JKP27/6 repeats. The mass spectra with the assumed protonated mass isotopes can be seen on the right-hand side of the plot.

Without a target, comparison of all 5 purple samples is difficult in Mass Profinder using the table of results. This is because of the large amount of data present and that no statistical analysis can be conducted, only the abundance of each feature in each sample is given. This means that when 1 vs 1 sample group is analysed, it is easy to get an average abundance within each sample group and determine which features are present or absent, however with 5 sample groups it means that the data analysis is very heavy without a specific target aim.

2.5.3.3 XCMS Online

XCMS used different deconvolution to the data analysis given by Mass Profinder and MPP as XCMS conducted the deconvolution by Isotopologue Parameters Optimization (IPO) processing. Multiple different parameters were tested and those closest to the ones chosen by Mass Profinder were selected where possible [23].

PCA Plot

Figure 2.20 shows the PCA scores plot from XCMS online.

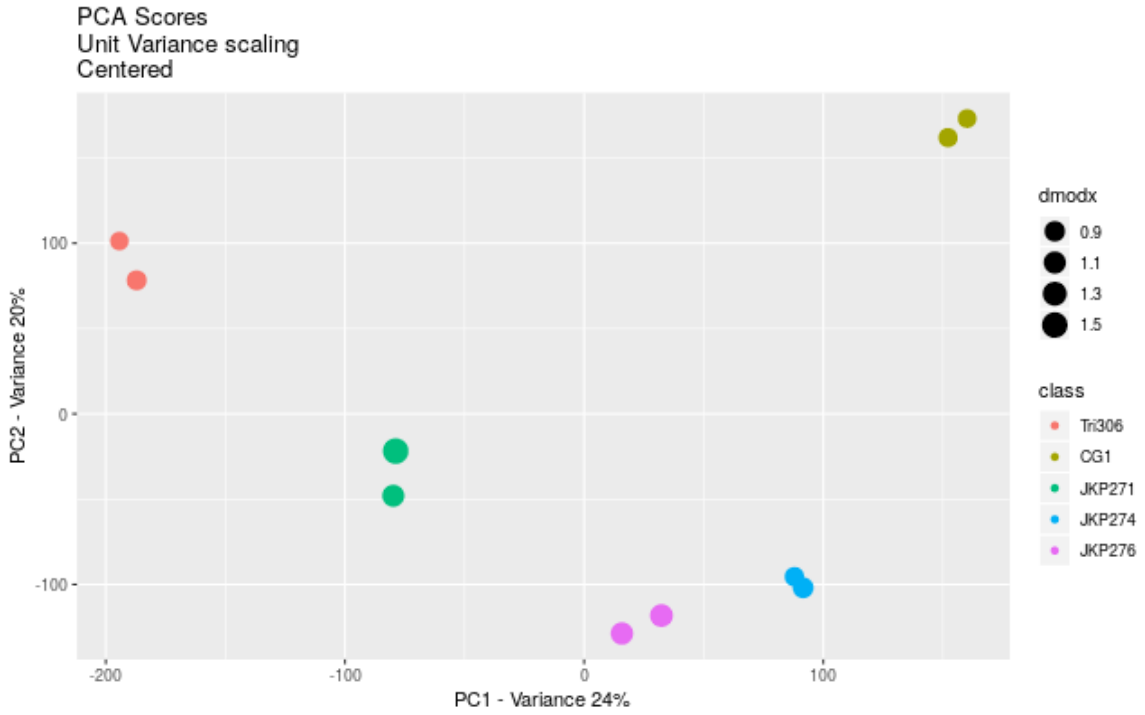

Figure 2-20. PCA Scores plot XCMS online on all purple tea leaf samples.

This shows that the largest difference is between the Tri306 and CG1 samples along the x-axis representing 24% change and 20% is shown up the y-axis that represents the difference between JKP276 and CG1. Each repeat sample is tightly clustered within each sample group.

The PCA scores plot in **Figure 2.20** shows that the repeat samples are all repeatable, therefore the instrument and extraction parameters have worked reliably. There is no real clustering within the sample types, though the closest samples are JKP27/4 and JKP27/6 on the x-axis which represents 24% difference. The two JKP samples are closest to JKP27/1, and all sit in between Tri306 and CG1 at the two extremes. The y-axis represents the second biggest change, which is 20%. The two extremes on this second principal component are between the JKP27/6 and CG1 samples. The two samples

that sit furthest away are the CG1 dots, suggesting that the Tri306 samples are closest related to the JKP samples. There are no obvious changes in the diameter of the dots, representing the distance to model ratio. There seems to be some slight possible changes, hinting that the repeat samples may have slight variation on what would be the third dimension, though this is not a significant difference and is barely visible. According to the statistical analysis within the software, there are no features with a *p*-value ≤ 0.01 and therefore suggesting there is no real difference between the sample groups.

2.6 Discussion

2.6.1 Conclusion of Data Analysis

In conclusion, purple tea is different to the other types of tea that has been compared against in the analysis. This is shown throughout with the differences shown on the PCA plots, with the purple tea sitting in its own cluster when compared to the other types of tea, not sitting within another group of tea samples. This indicates that it will have additional value than the tea that is already on the market as it is different to the yellow and green tea shown. It offers something different with the presence of anthocyanins.

The different clones of purple tea offer different concentrations of features, some of which have been identified. The difference in abundance of polyphenols has been highlighted, with Tri306 looking to have more significant polyphenol content insinuating it holds more value than the CG1 clone.

2.6.2 Software Comparison of Results

The results given in MPP gave possible identifications of compounds that were causing the differences seen on the PCA plots between the samples, particularly between the purple tea leaf samples. These features were extracted using Mass Profinder, then exported as .cef files and

inputting into MPP for statistical analysis. The results given by XCMS stated that there were no significant features found with a *p*-value less than or equal to 0.01, despite these features being discovered in MPP and positively checked in the raw data. The PCA plot in XCMS and MPP were similar showing that the data agrees with each other through the different platforms, despite the statistical analysis being different. The deconvolution parameters were slightly different but the main parameters were the same and should've been picking the same main features throughout the different packages. The difference in statistical analysis shows that the same data files give different results depending on the software and despite the deconvolution parameters been similar, the analysis cannot be repeated.

2.6.2.1 Summary

Overall, the PCA plots have helped distinguish between the differences of the different types of tea samples, along with the knowledge that the clones of purple tea are different to each other. MPP helped decipher the compound masses that were up regulated in certain sample groups, as did the Mass Profinder plots. Despite the PCA plots giving the same conclusions, the statistical analysis provided by XCMS and MPP do not agree with one another. MPP gave the most positive results as the data analysis compounds stated to be providing a difference between the samples, did show to be different in the raw data files. This suggests that the XCMS online results are false, providing all false negatives for the statistical analysis provided.

2.6.3 Limitations

The limitations of this project include the lack of QC samples used throughout. This was because of the low sample numbers, for some parts of this project only 5 samples were run at one time. For the larger project comparing all different types of tea colours, a QC sample could have been beneficial if the intention was to find markers for each tea, however the overall aim was to just compare the samples and so the randomised worklist, combined with the repeat injections would have highlighted any problems throughout the run time as they would not be grouped tightly on the PCA

plots. The instrument is also calibrated before and after each run and there is a reference solution used continuously, which enables the re-calibration of the whole run to some selected masses. For example, re-calibrating it all to 121.0509 *m/z* and 922.0098 *m/z* in positive ionisation mode would correct for any calibration problems that possibly occur throughout the run.

There was a lack of samples used for parts of the project, a larger sample set from a larger range of samples may have added benefits such as a wider range of possible differences within the same clone of tea. This may show that some differences are because of the different lighting and shade two of the same tea plants are grown in rather than differences within the actual tea clones. There was also a lack of repeat injection in some part of the project, the analysis would benefit from three repeats per sample. The samples were only analysed using a C18 column therefore the analysis may benefit from a different column or phase, such as HILIC. This would enable some different types of compounds to be analysed that are not visible on the RPLC phase used.

2.7 Conclusion

Overall, using this untargeted LCMS metabolomics workflow, the conclusion can be made that different types of tea are significantly different to each other, and further differences can be seen within the clones of the same type of tea. The statistical approach used was able to discover markers that are statistically different between groups. These markers can be used in future analysis should a marker for each type of tea clone be of interest in the future. The multivariate statistics showed most successful throughout this study, where all groups were clustered into distinct groups. Therefore, this technique could be used to identify which clone or tea type sample an unknown sample is. The different experiments conducted have all provided different learning interpretations.

This experiment investigated whether there were any differences between the types of tea, green, purple, yellow etc. A further investigation was conducted into the different clones of purple tea leaf to determine whether there are any differences between them that can be seen by the chemometric

software. Overall, it was determined that Tri306 purple tea leaf has the highest number of compounds that have the best health benefits, such as polyphenols/anthocyanins. Purple tea leaf is different to green and albino tea leaves, with it being most significantly different to green tea leaves. There are large differences between these clones that are shown on the PCA plots. The objective was to get a list of compounds causing the differences within the purple tea. The use of MPP and Profinder enabled this accurate list to be compiled, further showing that the combination of MPP and Profinder is the most preferable for this metabolomic workflow.

This untargeted metabolomic workflow using Agilent's MPP and Profinder enabled each objective to be accomplished, something that could not be done without the comparison software. The other pieces of software were not as concise and accurate, with XCMS online showing particularly limited results.

2.7.1 Future Work

For future work, there are other pieces of software that are free to use and may provide results as accurate as Profinder and MPP which could be tested. This would enable others to use the comparison software where money may be an issue and cannot purchase a license for MPP. It is also expected that more samples will arrive in the future that need to be compared and therefore the same workflow can be repeated, using the more streamlined version with only Profinder and MPP. Should markers for each type of clone or type/colour of tea be of interest in the future, this untargeted workflow can be used for the investigation.

2.8 References

1 Khan N, Mukhtar H. Tea polyphenols for health promotion. Life Sciences. 2007;81(7):519-533.

2 Khan N, Mukhtar H. Tea Polyphenols in Promotion of Human Health. Nutrients. 2018;11(1):39.

3 Hilal Y, Engelhardt U. Characterisation of white tea – Comparison to green and black tea. Journal for Verbraucherschutz and Lebensmittelsicherheit. 2007;2(4):414-421.

4 Xu J, Wang M, Zhao J, Wang Y, Tang Q, Khan I. Yellow tea (Camellia sinensis L .), a promising Chinese tea: Processing, chemical constituents and health benefits. Food Research International. 2018; 107:567-577.

5 Yang C, Wang Z. Tea and Cancer. JNCI Journal of the National Cancer Institute. 1993;85(13):1038-1049.

6 Mukhtar H, Ahmad N. Cancer Chemoprevention: Future Holds in Multiple Agents. Toxicology and Applied Pharmacology. 1999;158(3):207-210.

7 Mukhtar H, Ahmad N. Tea polyphenols: prevention of cancer and optimizing health. The American Journal of Clinical Nutrition. 2000;71(6):1698S-1702S.

8 Khan N, Afaq F, Mukhtar H. Cancer Chemoprevention Through Dietary Antioxidants: Progress and Promise. Antioxidants & Redox Signaling. 2008;10(3):475-510.

9 Crozier A, Jaganath I, Clifford M. Dietary phenolics: chemistry, bioavailability and effects on health. Natural Product Reports. 2009;26(8):1001.

10 Landete J. Ellagitannins, ellagic acid and their derived metabolites: A review about source, metabolism, functions and health. Food Research International. 2011;44(5):1150-1160.

11 Zhao W, Wang Y, Hao W, Yang H, Song X, Zhao M et al. Preparative isolation and purification of urolithins from the intestinal metabolites of pomegranate ellagitannins by high-speed counter-current chromatography. Journal of Chromatography B. 2015;990:111-117.

12 Furlanetto V, Zagotto G, Pasquale R, Moro S, Gatto B. Ellagic Acid and Polyhydroxylated Urolithins Are Potent Catalytic Inhibitors of Human Topoisomerase II: An in Vitro Study. Journal of Agricultural and Food Chemistry. 2012;60(36):9162-9170.

13 Cerdá B, Llorach R, Cerán J, Espin J, Tomás-Barberán F. Evaluation of the bioavailability and metabolism in the rat of punicalagin, an antioxidant polyphenol from pomegranate juice. European Journal of Nutrition. 2003;42(1):18-28.

14 Cerdá B, Tomás-Barberán F, Espín J. Metabolism of Antioxidant and Chemopreventive Ellagitannins from Strawberries, Raspberries, Walnuts, and Oak-Aged Wine in Humans: Identification of Biomarkers and Individual Variability. Journal of Agricultural and Food Chemistry. 2004;53(2):227-235.

15 Espín J, González-Barrio R, Cerdá B, López-Bote C, Rey A, Tomás-Barberán F. Iberian Pig as a Model To Clarify Obscure Points in the Bioavailability and Metabolism of Ellagitannins in Humans. Journal of Agricultural and Food Chemistry. 2007;55(25):10476-10485.

16 Doyle B, Griffiths L. The metabolism of ellagic acid in the rat. Xenobiotica. 1980;10(4):247-256.

17 Tomás-Barberán F, González-Sarrías A, García-Villalba R, Núñez-Sánchez M, Selma M, García-Conesa M et al. Urolithins, the rescue of “old” metabolites to understand a “new” concept: Metabotypes as a nexus among phenolic metabolism, microbiota dysbiosis, and host health status. Molecular Nutrition & Food Research. 2016;61(1):1500901.

18 Seeram N, Aronson W, Zhang Y, Henning S, Moro A, Lee R et al. Pomegranate Ellagitannin-Derived Metabolites Inhibit Prostate Cancer Growth and Localize to the Mouse Prostate Gland. Journal of Agricultural and Food Chemistry. 2007;55(19):7732-7737.

19 Cerda B, Espin J, Parra S, Martinez P, Tomas-Barberan F. The potent in vitro antioxidant ellagitannins from pomegranate juice are metabolised into bioavailable but poor antioxidant hydroxy-6H-dibenzopyran-6-one derivatives by the colonic microflora of healthy humans. European Journal of Nutrition. 2004;43(4).

20 Larrosa M, González-Sarrías A, Yáñez-Gascón M, Selma M, Azorín-Ortuño M, Toti S et al. Anti-inflammatory properties of a pomegranate extract and its metabolite urolithin-A in a colitis rat model and the effect of colon inflammation on phenolic metabolism☆. The Journal of Nutritional Biochemistry. 2010;21(8):717-725.

21 Saha R, Takahashi T, Kurebayashi Y, Fukushima K, Minami A, Kinbara N et al. Antiviral effect of strictinin on influenza virus replication. Antiviral Research. 2010;88(1):10-18.

22 Ryu D, Mouchiroud L, Andreux P, Katsyuba E, Moullan N, Nicolet-dit-Félix A et al. Urolithin A induces mitophagy and prolongs lifespan in C. elegans and increases muscle function in rodents. Nature Medicine. 2016;22(8):879-888.

23 1. Albóniga OE, González O, Alonso RM, Xu Y, Goodacre R. Optimization of XCMS parameters for LC–Ms Metabolomics: An assessment of automated versus manual tuning and its effect on the final results. Metabolomics. 2020;16(1). doi:10.1007/s11306-020-1636-9.

3 - Evaluation of Statistical Analysis Through the Detection of Possible Biomarkers for Bleached Human Hair

3.1 Aims

The aim of this experiment was to produce a reliable and accurate list of potential biomarker masses that are unique to bleached hair sample groups through the use of liquid chromatography-mass spectrometry and comparison software with univariate and multivariate statistical analysis. The different software techniques should all produce the same potential biomarker masses if the same deconvolution and multivariate statistics parameters are used. Using these human hair samples, the untargeted comparison software techniques were evaluated to compare whether the same biomarkers are determined throughout.

3.2 Introduction

Industrial companies are in need a confirmative measure to determine whether hair has been cosmetically altered with bleach. By using a metabolomic workflow, a biomarker exclusively for bleach in hair would confirm whether the hair has been purposely altered.

Hair from a range of volunteers of different hair colours, sex and ethnicities, that have never had their hair dyed or bleached, were to be collected. This produced various bleached samples, with the corresponding control (un-bleached) samples so the data analysis can be focused on the direct comparison between the bleached and control samples only. Any potential biomarkers will only be present in the bleached samples and through further investigation, identification of the compound could lead to a commercially available standard being purchased and used as a biomarker for cosmetically bleached hair.

The intention is to use these potential markers in routine Drugs of Abuse (DoA) hair analysis so it can be declared whether any drug results may be misleading due to the presence of bleach. To find any possible altered, stable compounds that are of significantly different concentrations between the untreated and oxidated hair samples, the untargeted hair metabolic approach was taken.

3.2.1 The Current Problem

Hair analysis is becoming more popular in forensic toxicology for the analysis of alcohol and Drugs of Abuse (DoA) markers since it can determine the of concentration of drug marker in each section of hair. This gives retrospective consumption information as to whether the subject is potentially a drug or alcohol abuser and is therefore routinely used for assessment of DoA or alcohol abstinence, child custody cases or workplace drug testing [1-2]. However, these DoA and alcohol marker concentrations can be significantly affected through the use of cosmetic hair treatments, with the most effective being oxidative hair bleaching with hydrogen peroxide [3-4]. These affect the concentration of the markers so significantly that it may potentially result in a false negative drug and/or alcohol test result [5]. Unfortunately, this makes it a popular adulteration avenue for those looking to avoid a positive result. Objective markers used to improve confidence that the hair has being manipulated with bleach, particularly when in court would be useful as so far only visual inspection of the hair gives rise to suspicion of adulteration attempts.

3.2.2 Hair Anatomy

Hair is often seen as an expression of a person's personality; however, it also has functional purposes. The hair on the scalp helps to keep sun rays off the scalp, the eyelashes and eyebrows keep dust and sweat out of our eyes, the hairs in our nose and ears keep germs out and the body hair helps regulate temperature [6].

Each hair has a hair shaft and a root, with the shaft being the visible part that sticks out above the skin [7]. The hair root extends down into the layers of skin and is surrounded by the hair follicle, which

is a tube-like pore that surrounds each root and strand of hair. Each follicle is connected to a sebaceous gland, with a lot of nerve endings surrounding the follicle [8]. At the hair root, there is the 'dermal papilla' which supplies each hair bulb with blood [6,9]. This anatomy is shown in **Figure 3.1**.

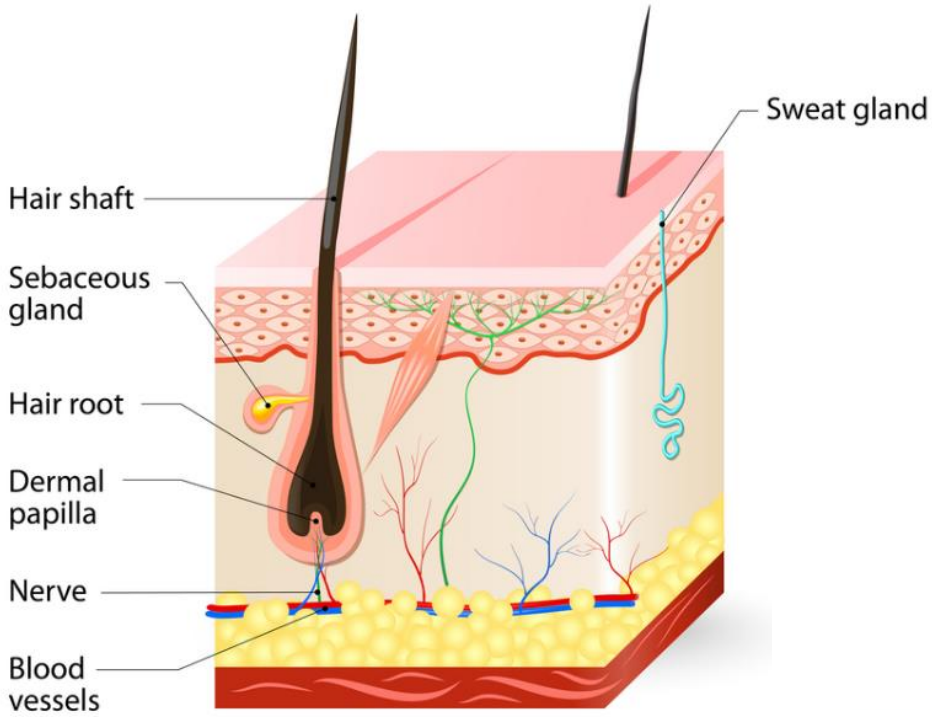

Figure 3-1. Labelled diagram of the outer hair structure.

Figure adapted from 'Rejuvenate Hair Transplant Centre - Hair Structure, Everything You Need to Know' [10]

When looking at the cross section of a hair strand, several layers made up of various substances make up the hair anatomy. The outer layer is the cuticle, which makes up about 10% of the structure. It has a protective function formed of 'keratin scales' and a layered cell structure, therefore, the smoothness of this layer determines the appearance of the hair [10-11]. The middle of the hair is the cortex, which dictates how thick the hair is. It is made entirely of keratin and contains a colouring pigment called melanin, which is what determines the hair colour [10]. This section makes up 85% of the hair anatomy [9]. The core of the hair is named the medulla, which is present is long head hair. This part makes up only 5% of the anatomy [7,10]. This cross-sectional analysis of the hair is shown in **Figure 3.2**.

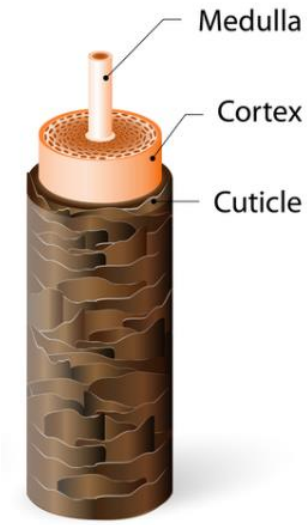

Figure 3-2. Labelled diagram of the cross section of the hair structure.

Figure adapted from 'Rejuvenate Hair Transplant Centre - Hair Structure, Everything You Need to Know' [10]

Within the cortex, it is the two different types of melanin (eumelanin and pheomelanin) that cause the differences in pigmentation of hair colour [12]. Generally, the more melanin present, the darker the hair colour and vice versa. **Table 3.1** shows the types of melanin with the persons hair colour.

Table 3.1. Shows relation of hair colour to presence of melanin types.

Hair Colour Appearance	**Type of Melanin**
Black	Large amount of eumelanin
Brown	Moderate amount of eumelanin
Blonde	Small amount of eumelanin
Strawberry Blonde	A mixture of pheomelanin and eumelanin
Red	Mostly pheomelanin with a small amount of eumelanin
Grey	Absence of both pheomelanin and eumelanin

The natural toning of the hair is due to the ratio of black/brown eumelanin to yellow/red pheomelanin [12]. It is possible to have more than one colour hair follicle on someone's head and it is also possible for one's hair colour to change overtime due to varying levels of melanin throughout their lifetime. Melanin goes through a pigmentation change and therefore a blonde-haired child may become a brunette in their teen/early adult years however, the change of colour may also be affected by external factors such as toxins, pollutants, and climate also. Darker hair tends to be more resistant to UV rays and decay than lighter coloured hair due to the lower photostability of pheomelanin compared to eumelanin. Melanin also plays part in protecting the hair against high

levels of sun exposure consequences, such as drying out and brittleness [7]. Albino hair is where there is no melanin present at all in either form, hence the white colour. Geographical regions and/or certain ethnicities are often associated with a particular hair colour due to the higher frequency of observed hair colours within that region, e.g., straight, dark hair in East Asians; curly, dark hair with Africans but a large variety of dark/light, curly/wavy/straight amongst Europeans [7,11]. Grey/white hair is not caused by a grey/white pigment but from a lack of pigmentation and melanin [12]. The light bouncing off the hair causes it to look the certain shade of grey or white, depending on the natural hair colour. It is often associated with growing older but can be caused by various factors other than age including stress, thyroid and vitamin B12 deficiencies. As the high levels of melanin in the hair protects it against it drying out and becoming brittle, grey hair often has a dry, brittle texture due to the absence of melanin. The different levels of melanin in hair can cause different enhancements of compounds in the hair shaft, therefore someone with dark hair will have different incorporation of compounds in the hair to someone with light hair.

3.2.3 Hair Analysis Background

The reason why hair analysis has become increasingly popular over the past few years when assessing alcohol and drug abstinence, is due to the advantage of its stability over other analytical specimens and easy transportation. The usual forensic analytical specimens for metabolomics, such as urine and blood, are highly dynamic leading to variable compositions dependent on daily activities, diet changes and stress, to name a few [13]. Hair analysis not only has greater stability over these specimens, but also has a non-invasive collection, easy-storage, and a long detection window for chemical substances therefore, making it possible for retrospective analysis for the months previous, depending on the length of the hair. Chemicals from the blood are distributed into the hair from the capillaries to the hair follicle during hair formation [14-15]. These substances are then retained in the matrix and cannot be easily removed by general day to day hygiene practises, such as washing and brushing [15]. Substances of exogenous origin are also incorporated into the hair shaft [13]. Despite the differences in each person's hair, it is assumed to grow at a rate of around 1 cm per month and so

segmental analysis of the hair can pinpoint the consumption, or exposure, of the target chemicals [15-16]. This also means that the longer the section of hair taken for analysis, the further back the analysis can go. For example, a 6 cm section of hair can look at approximately 6 months of history. The hair furthest from the scalp is the oldest, and therefore looks at the history furthest back in time.

3.2.4 Bleaching Hair

People often use bleach to lighten their hair colour. The bleach and melanin react, removing the colour through an irreversible chemical reaction; the bleach oxidizes the melanin molecule. The melanin is still present, but the oxidised molecule is colourless even though bleached hair tends to have a yellow tint to it. This is due to the structural protein in hair, keratin, having a naturally yellow colour to it. Bleach also reacts more readily with eumelanin than with pheomelanin, so some gold/red colour tints may remain after the bleaching procedure. The different levels of eumelanin and pheomelanin in people's hair is what makes each bleaching procedure give a slightly different finishing colour amongst different people. Hydrogen peroxide is commonly used as a lightening agent for hair, the peroxide in an alkaline solution opens the hair shaft to allow the peroxide bleach to react with the melanin. As the hair shaft is opened up, it is probable that this will also adversely affect the incorporation of drugs, alcohol and other compounds in the hair by decreasing their concentration significantly.

3.2.5 Drug Testing

Hair analysis is fairly new in the analytical industry, though it has recently emerged as a valuable analytical technique for retrospective analysis [14]. Hair analysis is often used in workplace drug testing and childhood custody cases to determine whether drugs of abuse (DOA) or alcohol has been ingested however, this can be difficult to determine an accurate concentration as the differences in hair colours and types lead to different uptake of drugs/alcohols in the hair. This is then made even more difficult as these concentrations can be significantly altered in by cosmetic hair treatments [1]. The treatments include hair straightening, dyeing, tinting, bleaching and "detox shampoos" [1,16-21].

The most effective treatment for altering the drug/alcohol concentrations are those treatments that contain hydrogen peroxide (H_2O_2) under alkaline conditions, which are most often what is used in permanent hair dye and bleach [3]. The use of such treatments can alter the concentrations of alcohol or DOA markers to the extent that they may potentially result in a false negative test, making it a popular technique for those looking to actively avoid a positive result [5]. The Society of Hair Testing (SoHT) has a different lower limit for each drug in which they use as the cut off limit; anything above this level will be classed as a "drug abuser" and anything below this level will be classed as "not detected (N.D.)", even if it has a peak on the LC-MS. The diagram in **Figure 3.3** shows the possibility of a drug falling below the cut off limit when the hair is cosmetically altered and therefore declared as not detected, where-as, if the hair was natural, it would be of significant concentration and they would be classed as a drug abuser.

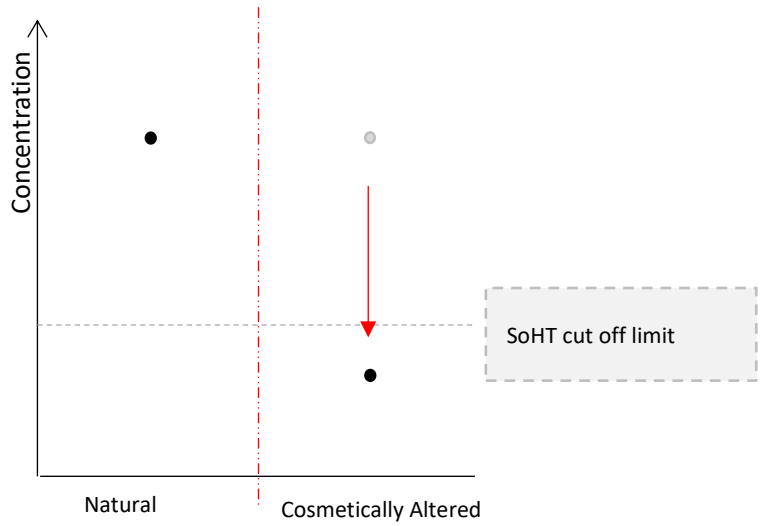

Figure 3-3. Diagram showing the decrease in concentration of drug in dyed hair.

When this is compared to natural hair levels, the presence of dye causes the drug to be declared as not detected, even though it was in fact present at a high concentration before alteration.

At present, the only way to determine any dyeing or bleaching treatments is by visual inspection, often looking for lines on the hair that could indicate a cosmetic change. Confidence on these assumptions can be improved by the presence of chemical markers and can be used in court to show that the levels of any drugs that present may be under-estimated. Such markers are not yet available to be detected through LC-MS yet, though this would be ideal as it can be processed at the same time as the routine procedure for the drugs. In a previous study, oxidation compound cysteic acid,

formed by oxidation of cysteine when bleached, was found to be a marker for hair damage by bleach [22]. However, this is also a compound that is present in control hair samples, therefore it is not indicative of bleached hair only and tedious cut off Experiments would have to be conducted [22-25]. In another previous study, 1H-pyrrole-2,3,5-tricarboxylic acid (PTCA) was described as a marker for oxidative hair treatments (bleaching) however the need cut-off values for PTCA are essential due to the presence in other products and in natural hair also [26]. Therefore, the presence of PTCA does not solely indicate a presence of bleach. A study also showed the varied levels of PTCA in the hair if the hair is repeatedly treated or washed differently. These limitations make it more difficult to determine whether treatment has occurred and therefore, something more exclusive to bleach is desired. During the course of this project, a paper was published where 1H-Pyrrole-2,3,4,5-tetracarboxylic acid (PTeCA) has been discovered as a compound that is more exclusively formed by oxidative hair treatments [27]. As it is exclusive to oxidative treatments, the cut-off values are not necessary however, its melanin dependency still hinders the current use of it in routine settings, alongside the lack of commercially available reference standards [2,27]. There is the possibility of using the oxidated product of each DOA, however this would require further evaluation of every drug due to the lack of knowledge on what each product would be, and the impact of treatment on each one. Therefore, there is a need for further reliable, drug-independent marker(s) to indicate the use of cosmetic hair treatment. A study has looked at the effects bleach has on amino acids, lipids and proteins in the hair however these studies were only on a small selection of compounds therefore no biomarkers were determined, only that bleaching had a significant impact on each studied compound [22-23].

Applying an untargeted metabolomics approach would widen the number of potential differences in the treated/untreated hair samples and detect possible biomarkers that indicate hair bleaching. Any biomarkers found can be combined with the routine screening of hair samples and therefore providing increased confidence as to whether the hair has been purposely manipulated: it will be no longer left to human interpretation. With further studies, the confirmed presence of bleach and/or dye could be used in to determine what the “real” concentrations of drugs should be in the hair.

3.3 General Workflow

Figure 3.4 shows the generalised workflow for each stage of the hair analysis. Each stage is important and is explained further in each section below, 3.3.1 to 3.3.6.

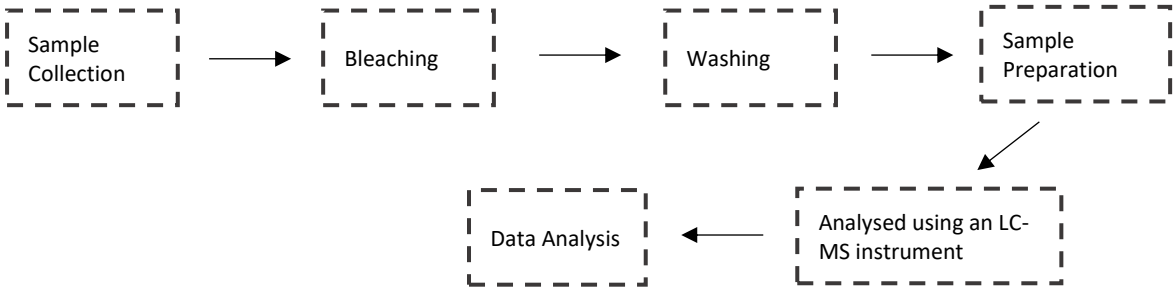

Figure 3-4. Generalised workflow for this hair analysis

3.3.1 Sample Collection

Ethical approval was sought from 'Keele University FNS Non-psychology Faculty Research Ethics Committee' and approved under reference NS-200085. Hair samples were to be cut from volunteers 0.5 cm from the scalp to ensure that no DNA could be retrieved from the samples. These samples were received by volunteers who had never previously dyed or bleached their hair and is therefore known as 'blank' hair and can be used as control samples. Sectioning each of these samples into two means that half can be bleached (diseased samples) and the rest can be kept as blank (control samples). Each sample was anonymised and therefore was given a number. After the samples were prepared, they could then be referred to as MRS_*n* and MRS_ *n*C, where the 'C' is the control sample from each person numbered *n*. For example, MRS_1 would be the bleached sample and MRS_1C is the control sample, both from person 1.

3.3.2 Bleaching

Bleaching of the hair was conducted in a manner similar to what someone may do themselves at home to try replicate the real samples that may occur in drug testing. Only one type of bleach was used throughout the course of this Experiment; 'Jerome Russell B-Blonde High Lift Powder Bleach' which was mixed with 'Jerome Russell 12% cream peroxide', both purchased from Boots UK. The instructions from the box were followed. Only one type of product was tested at this time as it was difficult to come across this product available to buy in the shops, without needing to be a trained hairdresser. Looking into the different bleach available in the shops whilst considering it is the hydrogen peroxide that causes the chemical change, the highest % peroxide solution was chosen to enable a noticeable change in the hair physically and chemically. The typical products available range from 2 – 12% peroxide and therefore the product chosen falls within the u*p*-most range for this study [10,28].

3.3.3 Washing

The purpose of the washing procedure is to decontaminate the samples from surface oils, like sweat, sebum, dirt, residue shampoo/conditioner etc., to ensure it does not lead to misinterpretation of results or interfere with any analytical procedures [1]. It improves the recovery of any metabolites or drugs incorporated in the hair and reduce the analytical background noise on the instrumentation. The washing procedure also removes potential external contaminants that may mislead any drug interpretations in the results [24]. Different solvents are effective at removing different compounds, for example organic solvents are largely effective for removing THC contaminants whereas aqueous solvents are better at removing ionisable drugs such as cocaine. It needs to be noted; however, not all hair decontamination procedures are 100% effective at removing external contamination to less than the reportable levels for all drug analytes, reported by the SoHT [30].

3.3.4 Sample Preparation

The sample preparation, including wash procedure, for this Experiment followed the approved Standard Operating Procedure (SOP) that a hair drug testing lab uses for their general routine screening procedure. The samples were all washed with 3 solvents to remove any potential interference from products, such as conditioner, hair gel or contamination from the air. Firstly, they were washed with water to remove ionisable contaminants, then with methanol to remove other potential contaminants and finally with dichloromethane. The DCM removes other contaminants, but the low boiling point also ensures the hair is left dry and clean.

The samples were then ready for sectioning into approximately 1 cm sections to cover a month per section, with the proximal end of the hair covering the most recent time, which is closest to the scalp. The distal end covers the oldest section of hair, with this being the furthest from the scalp. For this project, the distance from the scalp was not questioned, only knowledge that it is over 1 cm away from the scalp to ensure no DNA collection is possible. These 1 cm sections were then further chopped into small *ca.* 1mm sections. This gives the hair as much surface area as possible to allow the solvents to get into the hair for metabolite extraction. The solvent used is '30:30:40 MeOH: ACN: 0.1% Formic Acid in Water', which has been found to be the best overall solvent for metabolite extraction with least preparation time and cleanest MS background spectra. The samples were extracted whilst shaking at heat and the supernatant is analysed by MS.

To note, the sample preparation methods need to be kept as close to the routine procedures used by the drugs lab as possible. It would be un-productive and time consuming to need a different sample preparation and mass spectrometry method, in addition to the ones they already perform and therefore, the potential biomarkers need to integrate with their current sample workflow. It is of interest to see what additional information that could be obtained by keeping the sample preparation as close to this routine sample preparation and method.

3.3.5 Run on LC-MS Instrument

For this untargeted screening project, the samples were screened on the QTOF using LC conditions that are like the final MS method that will be used routinely, with the same mobile phases and column. The QTOF used was an Agilent 1290 Infinity II UHPLC + 6550 Q-ToF. These mobile phases, along with the chosen 'Agilent poroshell phenyl hexyl' column, have been shown to give good peak resolution and separation, without too much peak interference from the matrix. The gradient was broad to allow full separation of compounds in the preliminary Experiments. The samples were all run in positive and negative mode for initial screening.

Full parameters of the LC-MS instrument and method can be seen in **Table 3.2** below.

Table 3.2. LC-MS parameters for QTOF Hair Screening Analysis

HPLC-MS					
Instrument	Agilent 1290 Infinity II UHPLC + 6550 Q-ToF with iFunnel				
Column	Agilent Poroshell Phenyl Hexyl 2.1 x 100mm, 2.7µm				
Oven (ºC)	40°C				
Pump	Mobile Phase A	5mM Ammonium Formate 0.02% Formic Acid in Water			
	Mobile Phase B	5mM Ammonium Formate 0.05% Formic Acid in Methanol			
	Flow rate	0.4 mL/min			
	Gradient	Time / (min)	%A	%B	
		0.0	99	1	
		2.0	99	1	
		11.0	1	99	
		12.0	1	99	
		12.1	98	1	
		14.5	98	1	
	Runtime (min)	14.5			
Injector	Volume (µL)	3			
MS	QTOF/QQQ/TOF	Mass Spec Type:	QTOF	Mode	+ve (-ve)
	Source –Dual Jet Stream ESI	Gas temp	200°C	Gas flow	11 L/min
		Sheath Gas temp	350°C	Sheath gas flow	12 L/min
		Nebuliser pressure	207 kPa	Nozzle Voltage	100 V (1000 V in -ve)
		VCap	3500 V	Fragmentor	380 V

3.3.6 Data Analysis

Different pieces of software will be used to determine whether there may be a possible biomarker indicating the use of bleach. Each one will have different functions, some more useful than others but the result should be the same, and ideally each one should give the same possible list of biomarkers available. This would give confidence in the data analysis, and the potential biomarkers chosen. The different pieces of software to be used are the following:

- Mass Profiler
- Mass Profinder
- Mass Profiler Professional
- XCMS Online
- MetaboAnalyst

In Chapter 2, the analysis was conducted using XCMS online. However, the conclusions were that this data analysis did not seem to be the most accurate and reliable. For this reason, a further piece of data analysis software was used for this Chapter analysis, called MetaboAnalyst, as well as XCMS Online. It is available online for free and can both perform statistical analysis as well as give visual analysis is required. However, this software cannot perform deconvolution and the deconvoluted data from Profinder was exported as csv file and inputted into MetaboAnalyst.

3.4 Materials

Water, methanol (MeOH), acetonitrile (ACN), formic acid and ammonium formate of optima LCMS grade were obtained from Fisher Scientific (Loughborough, U.K.). Dichloromethane (DCM) of HPLC grade was also obtained from Fisher Scientific (Loughborough, U.K).ESI-L low concentration tuning mix and reference peak markers were purchased from Agilent Technologies (California, USA). The bleach used was a mix of ‘Jerome Russell B-Blonde High Lift Powder Bleach’ and ‘Jerome Russell 12% cream peroxide’, both purchased from Boots UK. The analysis was performed using Agilent

Technologies 1290 Infinity II UHPLC system, coupled to an Agilent Technologies 6550 Accurate-Mass Quadrupole-Time-of-Flight mass spectrometer with iFunnel.

3.5 Experimental

The sample preparation methods need to be kept as close to the drug testing lab's routine procedures as possible. It would be un-productive and time consuming to need a different sample preparation and mass spectrometry method, in addition to the ones they already perform and therefore, the potential biomarkers need to integrate with their current sample workflow. Preliminary Experiments were conducted to determine the best way to prepare the bleached samples and ensure that the method is successful.

3.5.1 Preliminary Experiment 1 – Investigation into Bleaching Procedure

Seven samples of hair were taken and each one was duplicated in a glass vial. One of each sample was taken and chopped into small sections. The bleach was mixed in a glass beaker, as per instructions on the box, 1 part liquid to 1 part powder bleach. Vials were labelled up with chosen Experimental initials of MRS, followed by numbers 1 – 7 corresponding to the original hair sample bag. One of each duplicate vial were labelled 'C' for control, leaving seven vials labelled MRS_1 – MRS_7, and seven vials labelled MRS_1C – MRS_7C. An aliquot was coated onto the hair in each vial that has labels MRS_1 to MRS_7 *only*. The bleach was left for 45 minutes, then washed off with water. The samples were filtered using filter paper and funnel, then washed again with water and left to dry for 24 hours at room temperature. All of the remaining samples, control and bleached, were then taken and washed with roughly 1 mL of water then decanted off. The process of washing then decanting off was repeated with methanol and finally with dichloromethane. The samples were then left to dry for 24 hours at room temperature.

3.5.2 Preliminary Experiment 2 – Investigation into Sample Preparation

Three samples were chosen for this Experiment. The hair was kept at length before bleaching, all with the proximal end of the hair held at the top, no chopping, split into two sections and one portion was bleached. The same bleach as Experiment 1 was used and it was also left on the hair for 45 minutes. It was then washed twice with water and left in the large 28 mL vial to dry for 24 hours. All samples, control and bleached, were washed with a 1 mL portion of solvent which were each decanted off; the solvents used were water, methanol then dichloromethane in order. The samples were left to dry for 24 hours at room temperature to ensure all solvents were removed. Each sample was then finely chopped into roughly 1 mm pieces and approximately 20mg of each sample was weighed out into a small 7mL vial, exact weights can be seen in **Table 3.3**. The bleached sample was prepared in duplicate (*x* and *x*a), but only single preparation for the control sample was conducted (xC), therefore resulting in 9 samples in total. 1 mL of '30:30:40 MeOH/ACN/0.01% formic acid in water' was added to each sample and then placed in the incubator at 60°C for 1 hour, whilst shaking at 200 rpm. The solvent was then transferred to a centrifuge tube, spun at 9800 RCF for 5 minutes then the supernatant was transferred to an LC vial for analysis.

Table 3.3. Table of sample weights for Experiment 2.
Repeats are shown by 'a' and 'b' where-as 'C' refers to a Control sample.

Sample		Weight (mg)	
1a	1b	19.20	21.12
1C		20.66	
2a	2b	19.80	21.48
2C		19.86	
3a	3b	20.92	20.46
3C		20.76	

3.5.3 Experiment 3 – Discovery of Potential Biomarkers

3.5.3.1 Overview

This experiment was conducted with more concentrated extracts so that the screening is more accurate, and the masses of interest are not lost in the noise. Once potential biomarkers are determined, the identification of the compound can be deduced, and a standard can be obtained.

The method can then be transferred over QqQ (MS/MS) where the standard will be run to determine its fragments and overall, the QqQ analysis will give more sensitive detection of the target compound(s) than QToF, therefore determining whether the compound is present within the samples or not by looking at the parent mass along with the transitions. The QqQ is more sensitive due to the increased time spent on the specific retention time and mass and therefore when the samples are less concentrated in the routine analysis, the biomarker may still be able to be detected at low levels. The potential biomarker(s) will be added into the routine drugs panel to give a conclusive result as to whether any potential drug concentrations may have been adulterated by the presence of bleach.

3.5.3.2 Sample Preparation

The same samples, control and bleached, from Experiment 2 were washed with a 1 mL portion of solvent which were each decanted off; the solvents used were water, methanol then dichloromethane respectively. Into a vial, 40 mg (±2 mg) of each sample were weighed out and the bleached samples were weighed in duplicate (*x* and *x*a), as seen in **Table 3.4**. 750 μL of '30:30:40 MeOH/ACN/0.01% formic acid in water' was added to each sample and then placed in the incubator at 60°C for 1 hour, whilst shaking at 200 rpm. The solvent was then transferred to a centrifuge tube, spun at 9800 RCF for 5 minutes then the supernatant was transferred to an LC vial for analysis. The analysis method used can be found in **Table 3.2**.

Table 3.4. Sample Weight Table for Experiment 3.
Repeats are shown by 'a' and 'b' where-as 'C' refers to a Control sample.

Sample		Weight (mg)	
1a	1b	39.61	40.33
1C		40.77	
2a	2b	40.01	40.79
2C		40.85	
3a	3b	40.63	39.55
3C		40.02	

Sample MRS1 was injected twice to test for repeatability of the instrument, as well as duplicate bleached hair preparation. Samples MRS2 and MRS3 bleached hair samples were prepared in duplicate, but the controls were only prepared once, with MRS1 injected twice hence why it appears as 001 and 002.

3.5.4 Experiment 4 – Secondary Check of Potential Biomarker Masses

3.5.4.1 Sample Preparation

More samples were prepared and tested, to ensure that the previously selected biomarkers are also present in other bleached hair samples of different starting colours, the hair colours of each sample can be seen in **Table 3.5**. Each hair sample was also re-bleached to ensure that repeat analysis can be conducted, giving confidence that the bleaching procedure gives the same biomarkers each time.

Table 3.5. Hair colour of samples used in Experiment 4.

Sample	Hair Colour
MRS4_1	Warm, light brown
MRS4_2	Warm, mid-brown
MRS4_3	Warm, dark blonde
MRS4_4	Grey/White/Black mix
MRS4_5	Grey and black mix
MRS4_6	Ashy light brown and grey mix

Using the original collected hair at the start of the project, 6 of these were taken and a portion was put into a vial in duplicate. One of each duplicate was taken and bleached using the same brand of bleach as Experiment 1. The hair was not chopped before bleaching, leaving the strands as long as possible. The bleach was left on for 45 minutes, then washed thoroughly with water and left to dry at room temperature for at least 24 hours. The samples, control and bleached, were washed with a 1 mL portion of solvent which were each decanted off; the solvents used were water, methanol then dichloromethane respectively. The hair was chopped, and 40 mg was weighed into small glass vials in duplicate (all samples in duplicate resulting in 12 samples for the control and 12 for the bleached),

the weights can be seen in **Table 3.6**. 750 μL of '30:30:40 MeOH/ACN/0.01% formic acid in water' was added to each sample and then placed in the incubator at 60°C for 1 hour, whilst shaking at 200 rpm. The solvent was then transferred to a centrifuge tube, spun at 9800 RCF for 5 minutes then the supernatant was transferred to an LC vial for analysis. The analysis method can be seen in **Table 3.2**.

Table 3.6. Sample Weights Table for Experiment 4.
The columns of 'a' and 'b' represents repeat preparations. The first 6 samples are bleached, where-as those with sample names ending with 'C' show control samples (non-bleached).

Sample	Sample Weights (mg)	
	a	b
MRS4_1	39.75	41.51
MRS4_2	39.09	41.12
MRS4_3	40.27	39.92
MRS4_4	40.50	38.85
MRS4_5	38.95	38.25
MRS4_6	39.92	39.58
MRS4_1C	40.90	41.01
MRS4_2C	41.38	39.03
MRS4_3C	40.25	40.22
MRS4_4C	38.16	40.35
MRS4_5C	38.38	40.19
MRS4_6C	40.54	41.54

3.6 Results

3.6.1 Experiment One – Investigation into Bleaching Procedure

Four of the samples were found to be pre-chopped too small and so they were dissolved by the bleaching process. All samples were discarded, and the Experiment was conducted again but with an adopted bleaching procedure.

3.6.2 Experiment Two – Investigation into Sample Preparation

The sample analysis worked fine, however the concentration of the peaks looked weak and this made it difficult to discover unique compounds at low concentration levels. For this reason, it was decided that the analysis needed to be repeated using more concentrated samples for the discovery of biomarkers. Once the biomarkers have been discovered, the analysis can then be repeated on less concentrated extracts to ensure that they are still detectable at a decent abundance. With higher

concentration of the potential biomarkers, more accurate masses can be determined hence making it easier to gather more structural information to determine possible identifications.

3.6.3 Experiment Three – Discovery of Potential Biomarkers

Despite the bleached samples being prepared in duplicate, MRS3a was injected twice. During the incubation stage of the sample preparation, the vial for MRS3b cracked, resulting in the loss of all the sample and solvent. Therefore, MRS3a was injected twice instead.

3.6.3.1 Traditional Analysis

In positive mode, the raw data TICs overlay perfectly and show no differences. On inspection of raw negative mode data, there are peaks in the bleached samples that are not as prominent in the control samples. The TIC overlay of all samples in **Figure 3.5** shows this difference, with an annotated zoom in **Figure 3.6.**

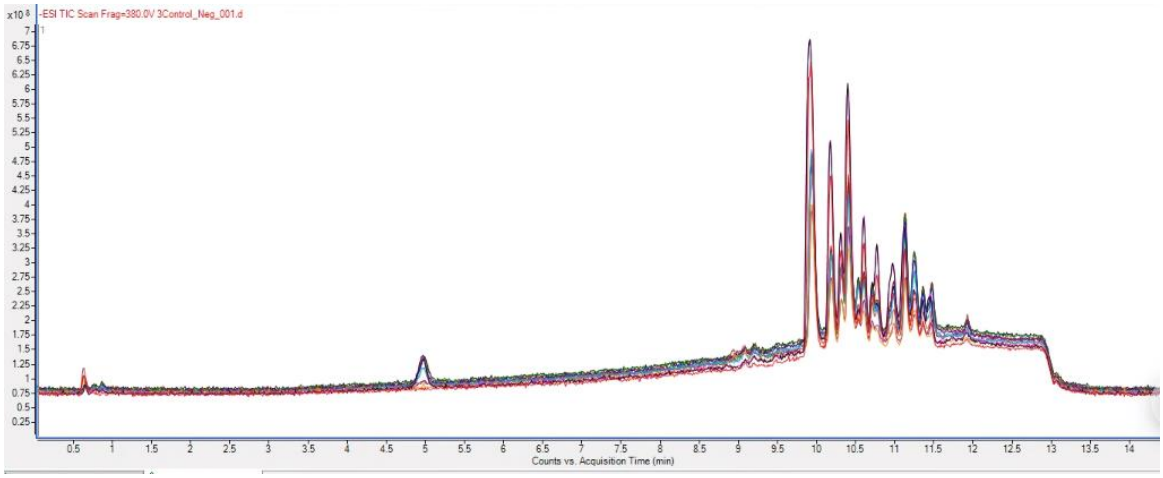

Figure 3-5. TIC Overlay of all samples Negative Mode.
One peak can be seen at around 5 minutes that appears different in some samples than others.

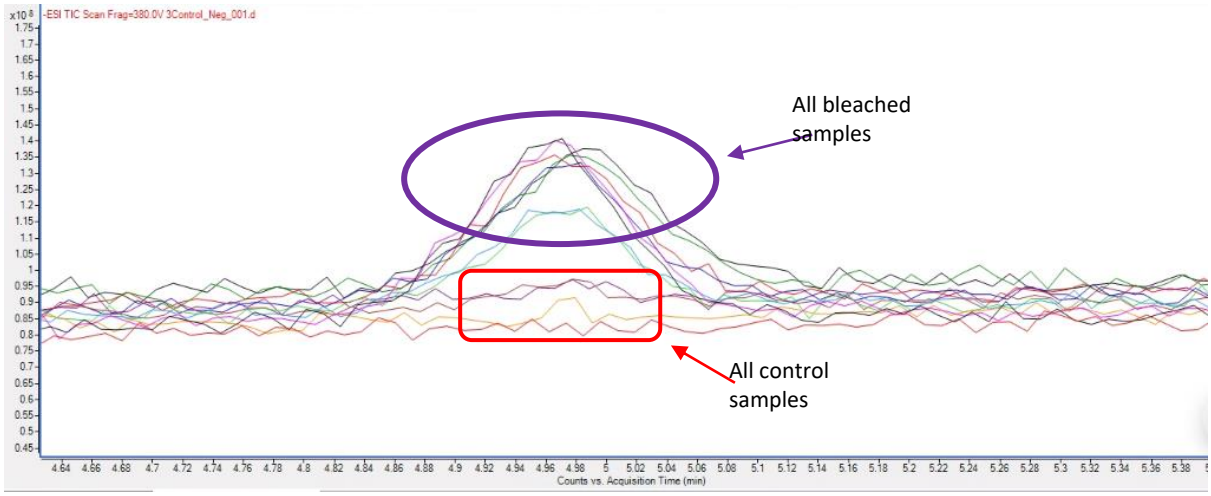

Figure 3-6. TIC Overlay zoom of difference in abundance Negative Mode.

Control samples do not show a peak present at the same retention time that the bleached samples show a peak.

The mass was extracted from this peak and showed to have an accurate *m/z* of 297.0379. This was then extracted in all samples to see if it is present in any of the control samples. The EICs can be seen in **Figures 3.7 - 3.8** for all samples overlaid on one plot.

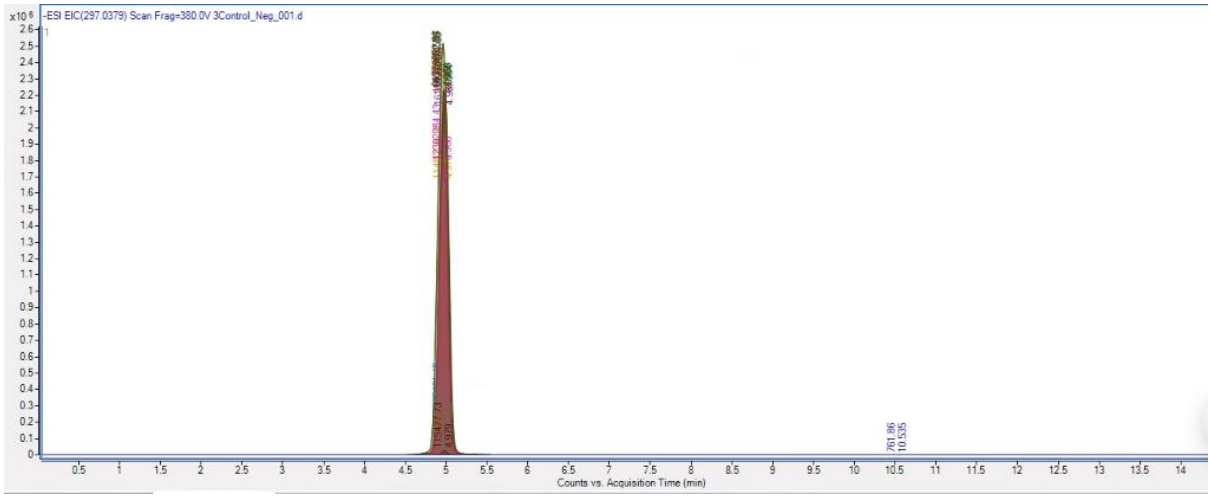

Figure 3-7. EIC (297.0379) m/z in all samples Negative Mode

Figure 3.8 shows a zoom on peak of interest with the bleached and control samples annotated:

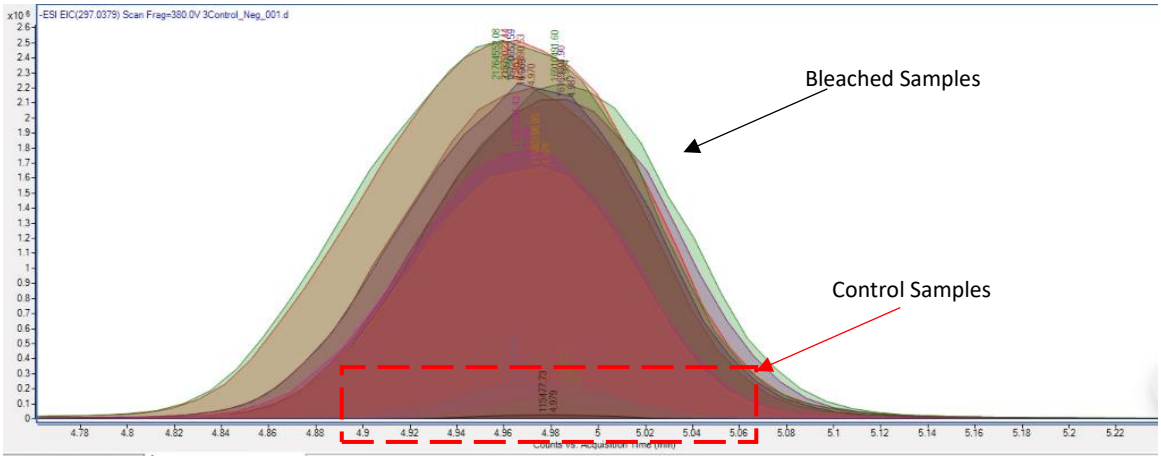

Figure 3-8. Zoom on EIC (297.0379) m/z from Figure 3.7.

Bleached samples show a large peak extracted. A small peak is also present in some control samples.

Figure 3.9 shows the EIC (297.0379) *m/z* in the control samples only. The blue and green peaks are the sample MRS1 repeats, with the black peak representing sample MRS2. Sample MRS3 control does not have the peak present. **Figure 3.10** shows the EIC (297.0379) *m/z* in all bleached samples. The abundance of the peak differs slightly throughout the samples; however, it is still a peak in every bleached sample at a substantially higher concentration than the control samples.

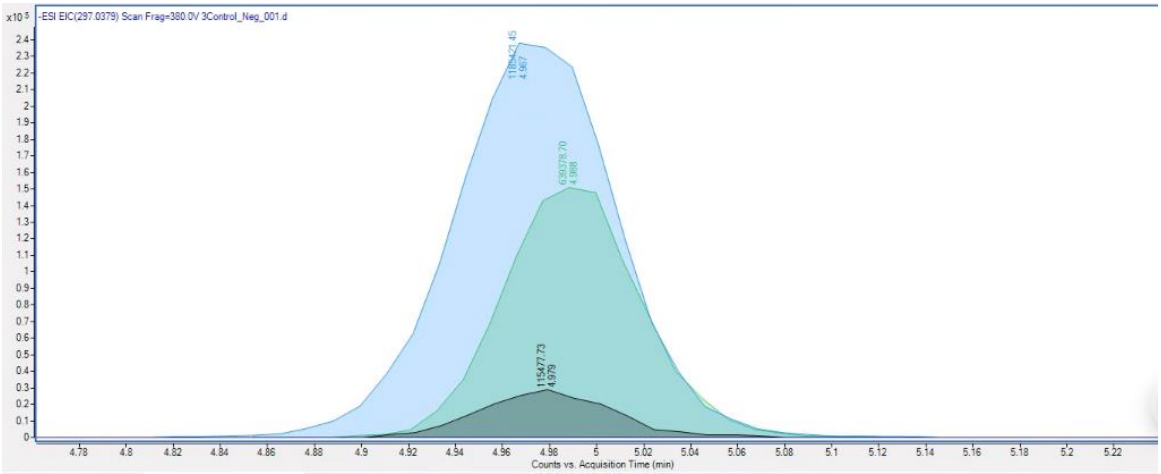

Figure 3-9. Only control Samples EIC (297.0379) m/z Negative Mode.

Positive peak present in MRS1 control and MRS2 control. Lack of a peak extracted in MRS3 control.

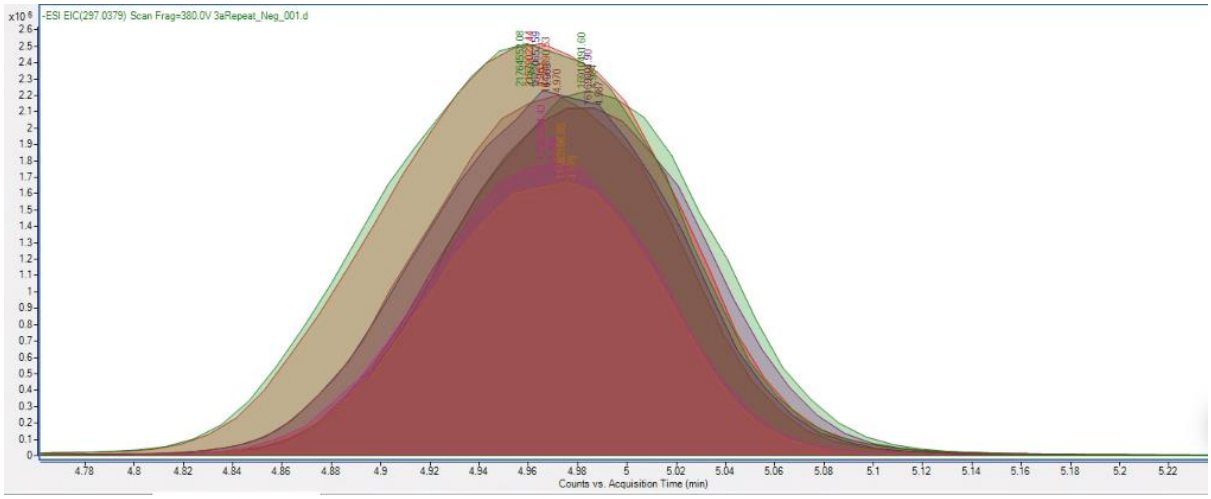

Figure 3-10. Only Bleached Samples EIC (297.0379) m/z Negative Mode.

Peak extracted in all bleached samples.

This compound has a retention time of 4.973 minutes and a neutral mass of 298.0452 Da. No identifications can be made using the 'Metlin 8.0' database, nor the 'Metlin Metabolites' PCDL database. An empirical formula of $C_{12}H_6N_6O_4$ has been suggested from the neutral mass in MassHunter Qualitative Analysis software.

The abundances of this compound can be seen in **Table 3.7** across all samples, 'a' is a duplicate preparation of the sample (i.e., 1 and 1a are duplicate preparations corresponding to the 1C - control sample). This peak would not be an ideal biomarker since it is present in some control samples, though there is a large difference between the two sample sets.

Table 3.7 shows the abundance of the peak across the samples, with an average of 1a, 1b and 1control.

Table 3.7. Table of peak areas of EIC (297.0379) m/z Negative Mode.

*The average abundance of each sample group, fold change and p-*values *are also included.*

Sample	Peak Area of EIC (297.0379) *m/z*
1a	16910492
1b	16169902
1Control	639379
2a	11480197
2b	12302284
2Control	115478
3a	21579022
3b	21764552
3Control	0
Median Bleached Abundance	16540197
Median Control Abundance	115478
Fold Change	143.23
***p*-value (Two-Tailed)**	0.0291

Using the median values, the fold change has been calculated. This FC value is 143 showing that there is a significant difference between the two groups. Using t-test, the *p*-value was calculated using two-tailed test and can be found to be 0.0291, which is larger than the significant *p*-value of 0.01 that is used to test significance. Therefore, this shows that there is not a significant difference between the sample groups when a *p*-value of 0.01 is used. This shows that the compound is not a potential biomarker, despite the traditional analysis looking different to the eye on the extracted ion chromatograms. This shows the need for software that can look further into the chromatograms to determine if there are any differences not visible on the TIC.

3.6.3.2 Mass Profinder

Mass Profinder was used to follow the same type of metabolomic workflow that has previously shown promise in Chapter 2. The visual aspect in this software allows checking of the EICs quickly, without having to check them in the raw data like Profiler requires which should reduce the overall analysis time, whilst maintaining the required level of accuracy. The exported tables give this data in

an excel format that can be easily read to determine the most abundant features that are potentially unique to one sample group.

3.6.3.2.1 Positive Mode

There are 227 features that Mass Profinder highlights as unique to bleached hair. **Table 3.8** shows the top 10 markers identified in positive mode that were found to be unique to the bleached samples.

Table 3.8. Mass Profinder – Top features unique to bleached

Top 10	Mass (neutral Da)	RT (min)	Ave peak area in Bleached
1	523.3513	10.920	36565286
2	431.2876	7.770	12587497
3	519.3193	10.570	12241327
4	435.3192	8.180	9308439
5	177.1264	5.080	5362560
6	365.2892	10.150	4815890
7	431.2884	7.620	4296880
8	433.3045	8.920	4061533
9	379.2708	10.050	2572259
10	241.1672	6.610	2444113

The top feature in **Table 3.8** has an *m/z* of 524.3584 at 10.9 min and **Figure 3.11** shows the Profinder results where the peak integrated correlates to the bleached samples only.

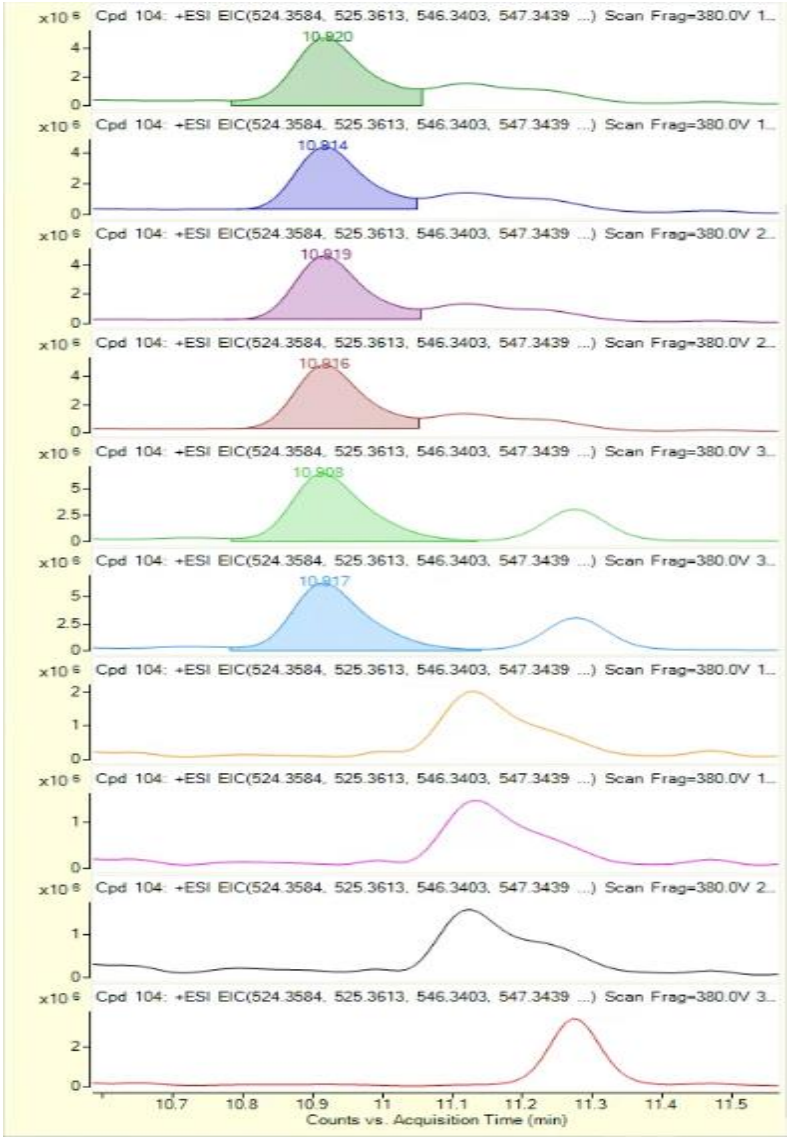

Figure 3-11. EIC (524.3584) screenshot from Mass Profinder.

It shows the presence of an integrated peak in the bleached samples only. The control samples do not have a peak at the correct retention time with the correct m/z.

The raw data confirms the Mass Profinder results, which can be seen in **Figure 3.12** (bleached samples showing presence the peak at the correct retention time) and **Figure 3.13** (control samples showing absence of a peak).

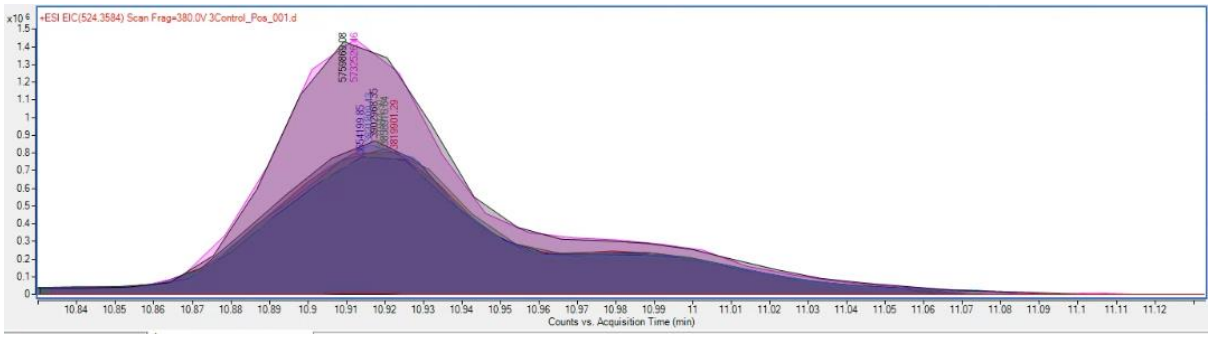

Figure 3-12. EIC (524.3584) in bleached samples from raw data.

Integrated peak present in all bleached samples.

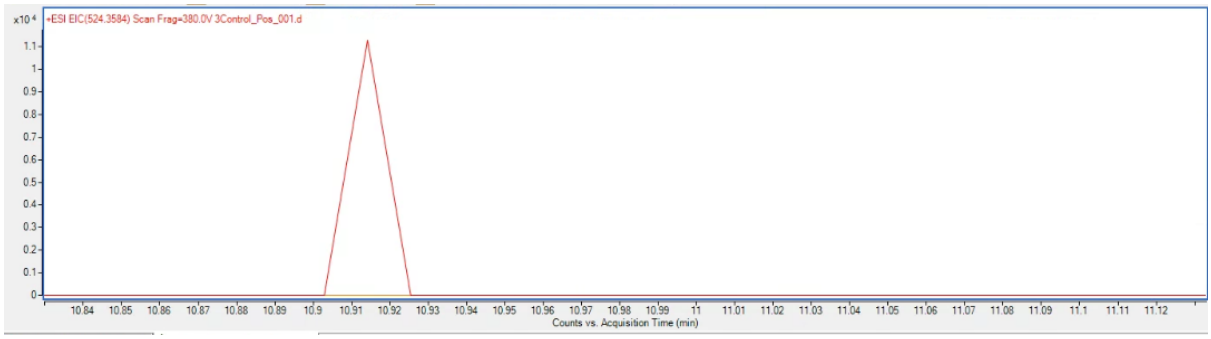

Figure 3-13. EIC (524.3584) in control samples from raw data.

No integrated peaks present in any control samples.

The raw data shows the presence of a compound with the *m/z* 524.3584 at 10.91 minutes in the bleached samples (**Figure 3.12**) however, in the control samples there are no peaks except for the outline of one red trace in one control sample, MRS3 control, as seen in **Figure 3.13**. This slight trace may be the presence of a compound too low in concentration to be detected, or it may be noise coincidentally at the same retention time.

3.6.3.2.2 Negative Mode

There are 91 features that Mass Profinder highlights as unique to bleached hair. **Table 3.9** shows the top 10 markers identified in negative mode that were found to be unique to the bleached samples.

Table 3.9. Profinder – Top features unique to bleached

Top 10	Mass (neutral Da)	RT (min)	Ave peak area in bleached
1	531.3003	9.200	1803687
2	519.3184	10.550	1312605
3	417.3075	8.860	830533
4	563.2897	6.790	690126
5	529.2842	8.450	494262
6	417.3074	8.760	482833
7	433.3022	7.620	426369
8	480.3604	11.490	388337
9	549.3104	8.180	373278
10	553.8638	0.660	362564

Table 3.9 shows the most abundant feature at 9.20 min with a *m/z* of 530.2937 and can clearly be seen to be present only in bleached samples. **Figure 3.14** shows the Profinder chromatograms where an integrated peak can be seen in only the bleached samples (the top 8 chromatograms).

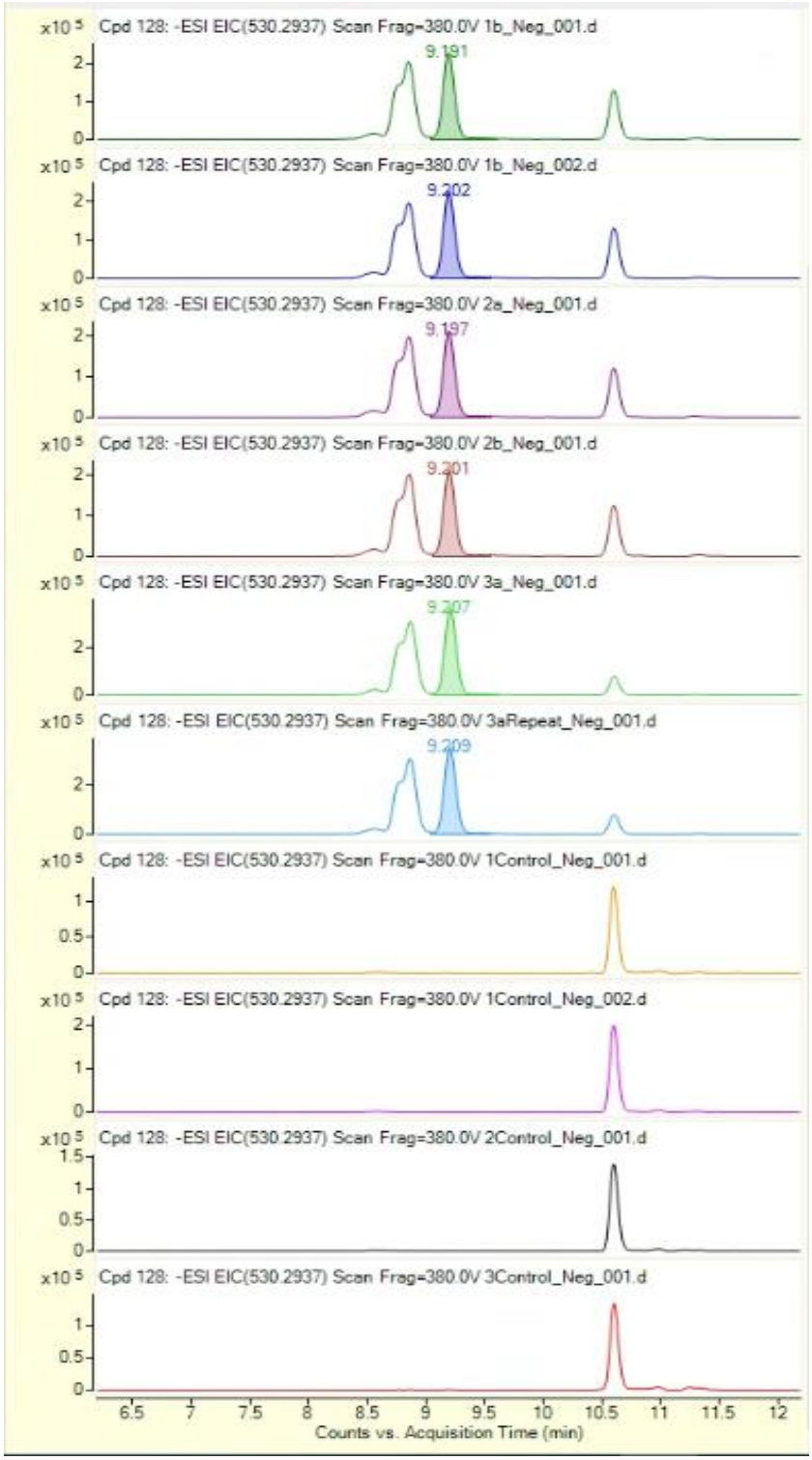

Figure 3-14. EIC (530.2937) in Mass Profinder.

It shows an integrated peak in each of the bleached samples, yet the absence of a peak in the control samples at the same RT.

Looking at the raw data analysed in MassHunter Qualitative, the bleached samples showed presence of a peak in each sample however, there were no peaks in the control samples. The **Figure 3.15**

shows the bleached samples only, with 530.2937 *m/z* extracted, and **Figure 3.16** shows the control samples with same mass extracted yet there are no peaks present, only traces relating to noise.

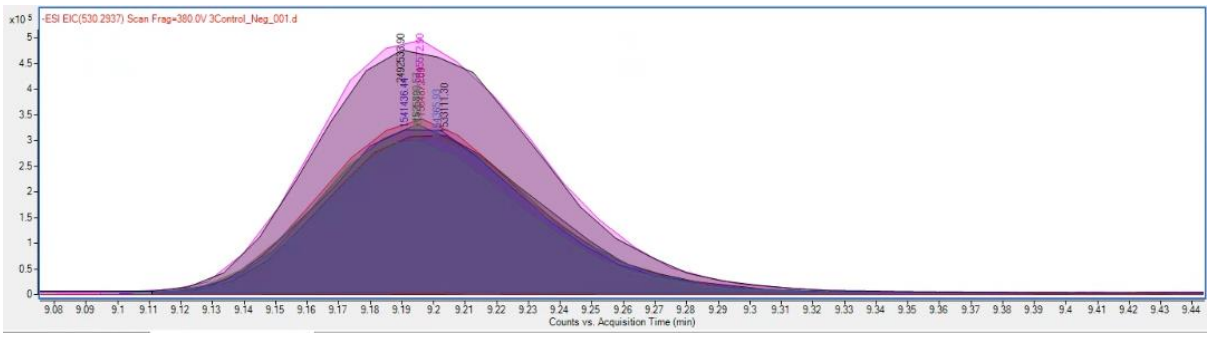

Figure 3-15. EIC (530.2937) extracted in bleached samples, raw data.

Integrated peaks present in all bleached samples.

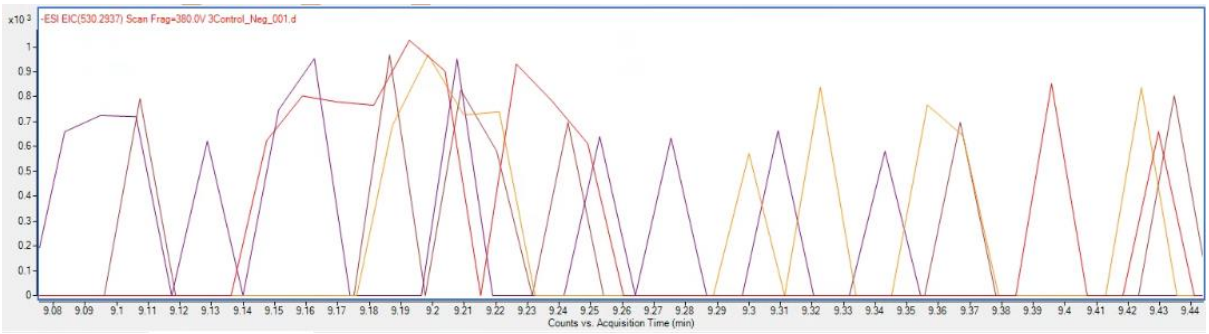

Figure 3-16. EIC (530.2937) extracted in control samples, raw data.

No integrated peaks present in any control samples.

Other Compounds

The compound PTCA was found at 0.8 minutes, present in all samples though the area of the peaks varies throughout. There is a significant difference seen between the bleached and control samples, with a further difference seen between each different person which is highlighted in **Figure 3.17**. This is not an ideal marker due to the presence in the control samples, but also due to the large difference seen within the samples group due to the different hair colours. The peak areas can be seen in **Table 3.10** where the grey boxes show the control samples. These peak areas are lower than those in the white boxes which are the bleached samples. Sample 3 has an overall higher concentration than samples 1 and 2, showing inconsistent concentrations of PTCA between the different hair colours and/or types.

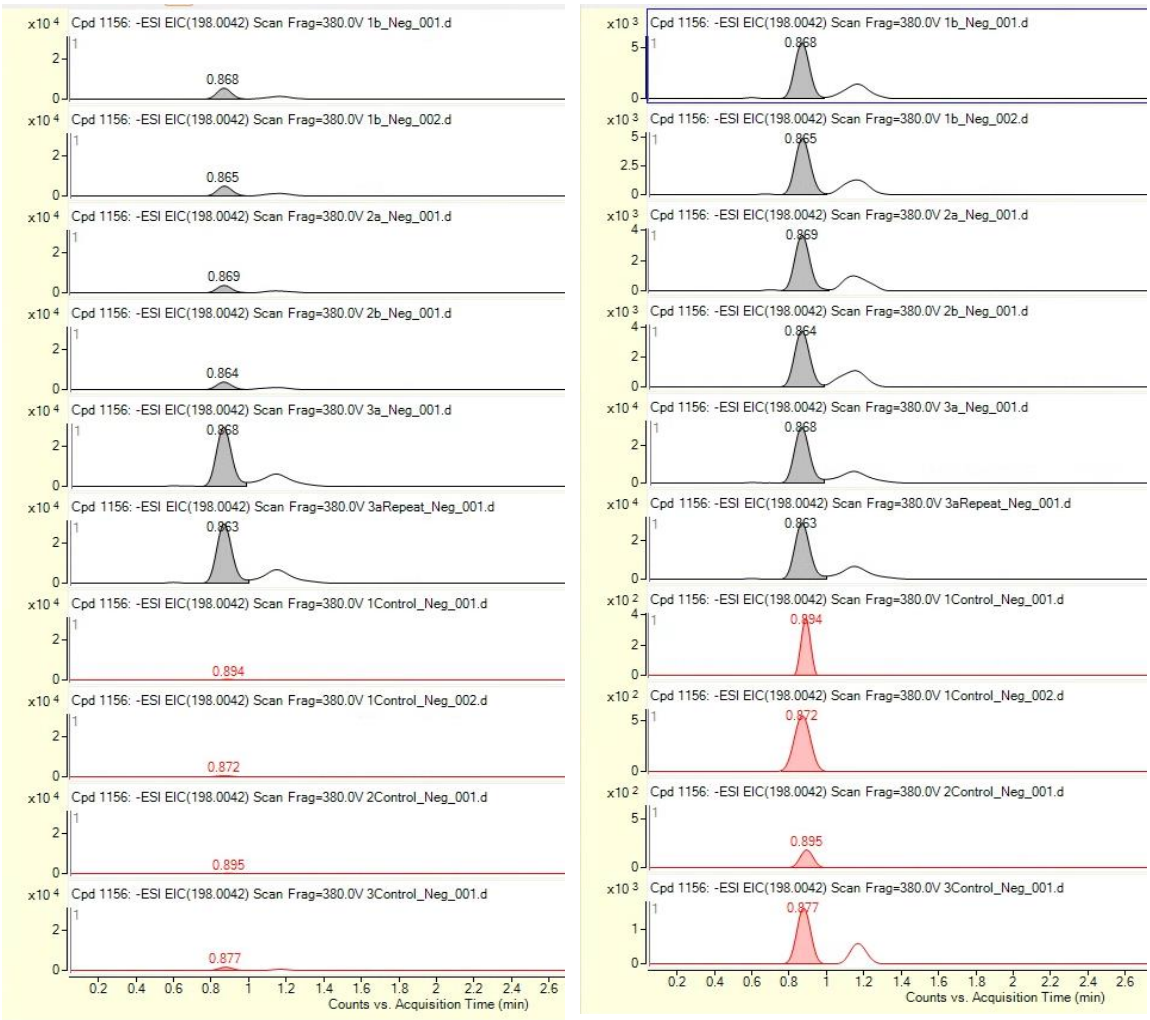

Figure 3-17. PTCA peak in Mass Profinder.

The left-hand side shows the peak with the y-axis linked therefore showing the significantly higher peak in sample MRS3. Right-hand side shows the zoom on the peak, with the y-axis not linked and so showing the varying abundance in more detail.

Table 3.10 shows the table of PTCA areas from Mass Profinder, along with the fold change and *p*-values associated with the bleached and control sample groups.

Table 3.10. Table of PTCA peak area in Profinder 3.6.3 Negative Mode

Sample	Peak Area [PTCA]
1a_001	30024
1a_002	30481
1b_001	30770
1b_002	28520
2a_001	21172
2b_001	21744
3a_001	168925
3a_002	170047
1C_001	1343
1C_002	3421
2C_001	923
3C_001	8734
Median Bleached Abundance	29949
Median Control Abundance	2382
Fold Change	12.6
p-value (Two-Tailed)	0.2665

The fold change is only 12.6, therefore is lower than the value of 30 that is typically used to show significant difference between the two sample groups. The *p*-value in **Table 3.10** is 0.2665 which is higher than the 0.01 value used for *p*-value significance. Therefore, this shows that PTCA is not a potential biomarker according to this analysis using Profinder.

3.6.3.3 Mass Profiler

Mass Profiler compares two datasets directly against each other only, giving the possibility to view plots of features that are unique to one dataset only. It deconvolutes the data and gives full statistical analysis based on the '.d' raw data file. It gives the results in a table, with the RT, neutral mass and abundance, therefore allowing features that are unique to one dataset to be determined by a table.

In Mass Profiler, the data was inputted as raw data file and the software conducted Molecular Feature Extraction. A method was created to set parameters for feature finding, alignment, normalisation, statistics and filtering. For this work, the default values are already optimised for small organic molecules and so were kept to only those ions with peak intensity above 600 counts and a charge state of 1-1 as only small molecules are of interest at this stage. The RT tolerance was +/- 0.1% min and a mass tolerance of +/- 20 ppm. Only features occurring in 100% of at least one group would be shown if the feature has a score above 70 (score is set by the software based on expected/observed mass and isotope ratios). A group difference with a fold change of more than 4 was also set.

In positive ionisation mode, out of 561 features that Mass Profiler identifies as unique to bleached hair, 92 of them are the same as those identified in Mass Profinder. Majority of the features, 469, are different. Some of these are due to the deconvolution of peaks where it can be seen that some peaks have not been deconvoluted efficiently giving two features the same mass and retention time of 0.01 min difference, and some are completely different masses and retention times than those in the original Profinder analysis.

In negative ionisation mode, out of 289 features that Mass Profiler identifies as unique to bleached hair, 56 of them are the same as those identified in Mass Profinder (19.4%). Majority of the features, 233, are different. As with positive ionisation mode, some of these are due to the deconvolution of peaks where it can be seen that some peaks have not been deconvoluted efficiently giving two features the same mass and retention time of 0.01 min difference, and some are completely different masses and retention times than those in the original Profinder analysis.

3.6.3.4 MPP

3.6.3.4.1 Positive Mode

PCA Plots

In the PCA plot, MRS3 samples all sit separately to MRS1 and MRS2. This could be due to the colour of the natural hair as MRS1 and MRS2 are dark blonde where-as MRS3 is black in colour and this difference is known to incorporate different compounds in the hair due to the melanin differences. The first component shows a 36.46% variance for the PCA plot shown in **Figure 3.18**. The second component has a 29.30% variance. The samples are distinctly grouped into bleached and non-bleached samples, even before filtering the features by *p*-value and fold change.

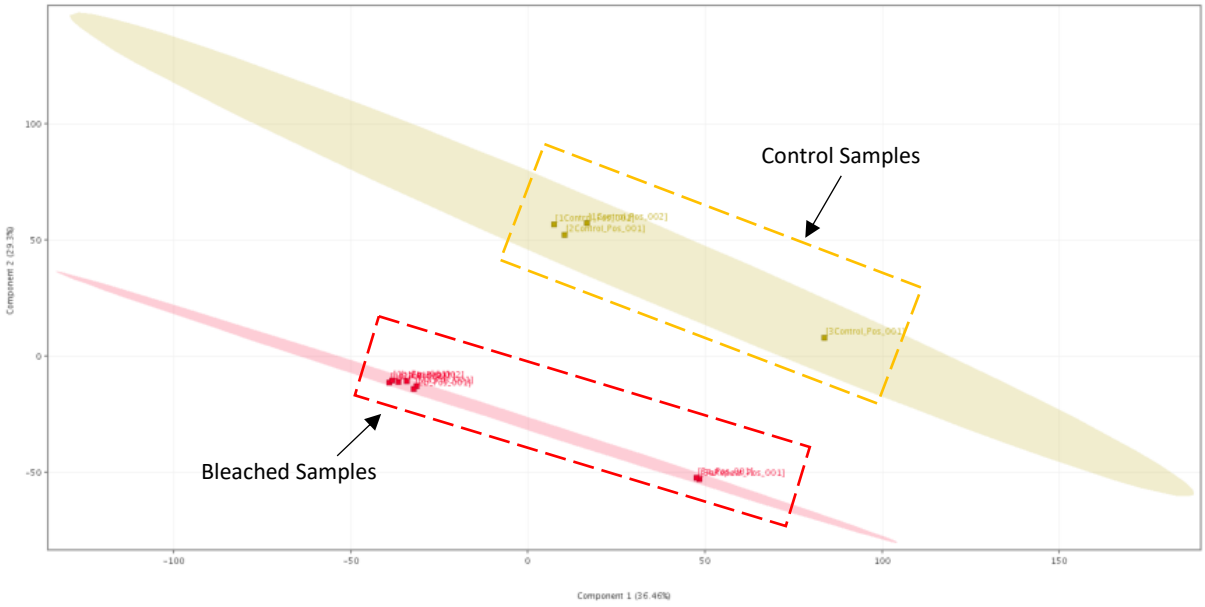

Figure 3-18. MPP 2D PCA plot on all features.

Red represents bleached samples, yellow represents control samples. This shows that the samples are separated by sample group, and further separation within the groups, likely to be due to natural hair colour differences.

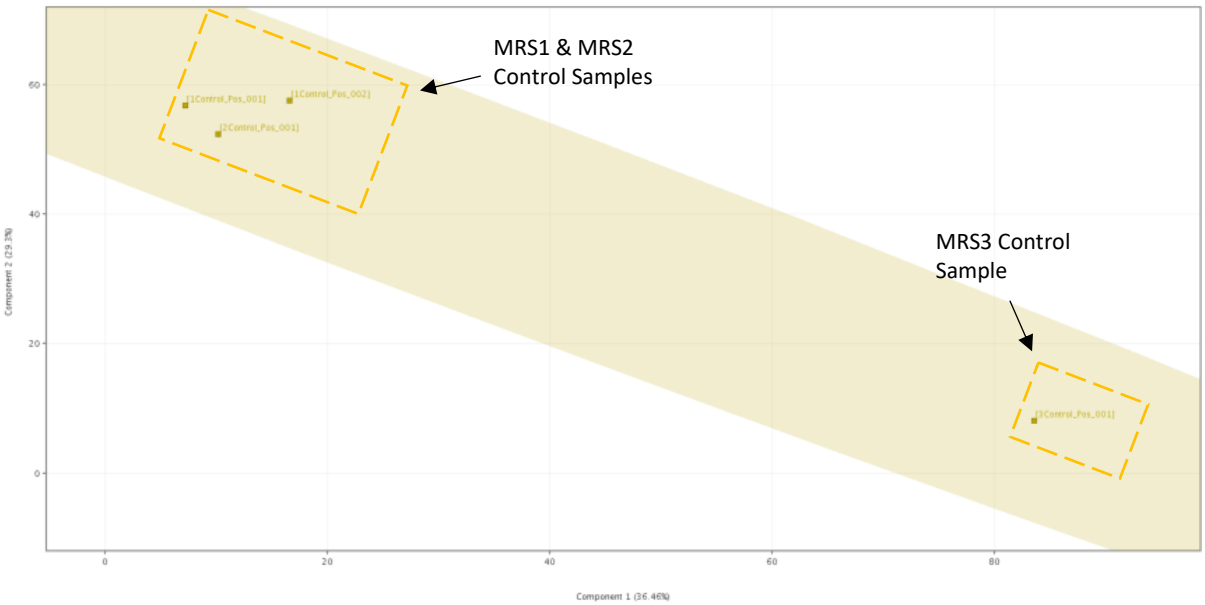

Figure 3-19. MPP Zoom on control samples in 2D PCA from Figure 3.27.

The separation between the MRS1, MRS2 control and MRS3 control samples can be seen spread along the x-axis (1st principal component) therefore representing a large variance in the samples.

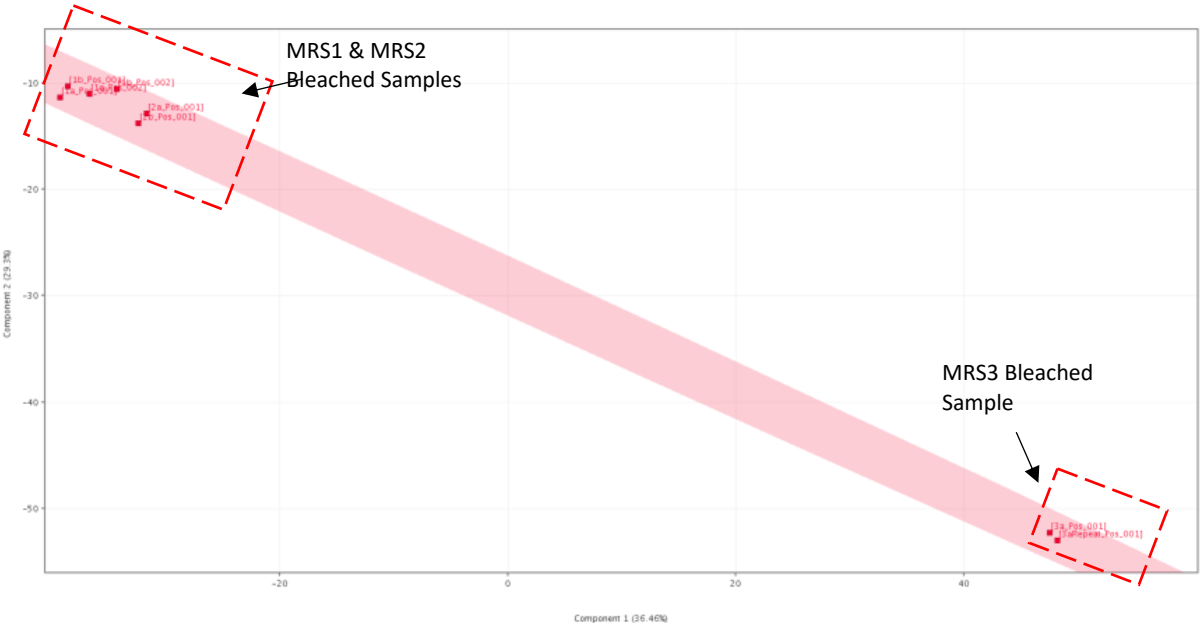

Figure 3-20. MPP Zoom on bleached samples in 2D PCA plot Figure 3.27.

The separation between the MRS1, MRS2 bleached and MRS3 bleached samples can be seen spread along the x-axis (1st principal component) therefore representing a large difference in the samples.

The PCA plot shown in **Figure 3.18** and the coloured 95% confidence ellipses shows that the samples are grouped distinctly into bleached and control samples. However, the sample groups have further clustering within these groups, which can be seen in **Figure 3.19** for the control samples and **Figure 3.20** for the bleached samples. The zooms in **Figure 3.19-3.20** both show that the sample MRS3 is

different to the MRS1 and MRS2 samples, in both bleached and control, and this difference is a significant difference as they're spread along both the x and y axes. This difference is most likely due to the alternate hair colours or types as MRS1&MRS2 are both dark blonde/light brown of European ethnicity, where-as MRS3 is black in colour and of Asian ethnicity. The difference between the control and bleached group within each hair sample (i.e., between MRS1 bleached and MRS1 control) is spread along the y-axis mainly, but also the x-axis showing that there is still a significant difference between the groups.

Overall, it can be deduced that there is a difference between bleached and control samples, but also a large difference within the samples potentially due to hair colour and ethnicity. This is further represented by the negative ionisation mode giving the same clustering of samples and groups.

Feature Finding Graph

Data filtering was conducted on all the features so that only features of significance are shown, which means they have a *p*-value less than or equal to 0.01 and a fold change greater than 30. For ease, the log fold change (log FC) is shown in the tables.

The Feature Finding Graph in **Figure 3.21** shows the significant features with a significant difference between the control and bleached samples:

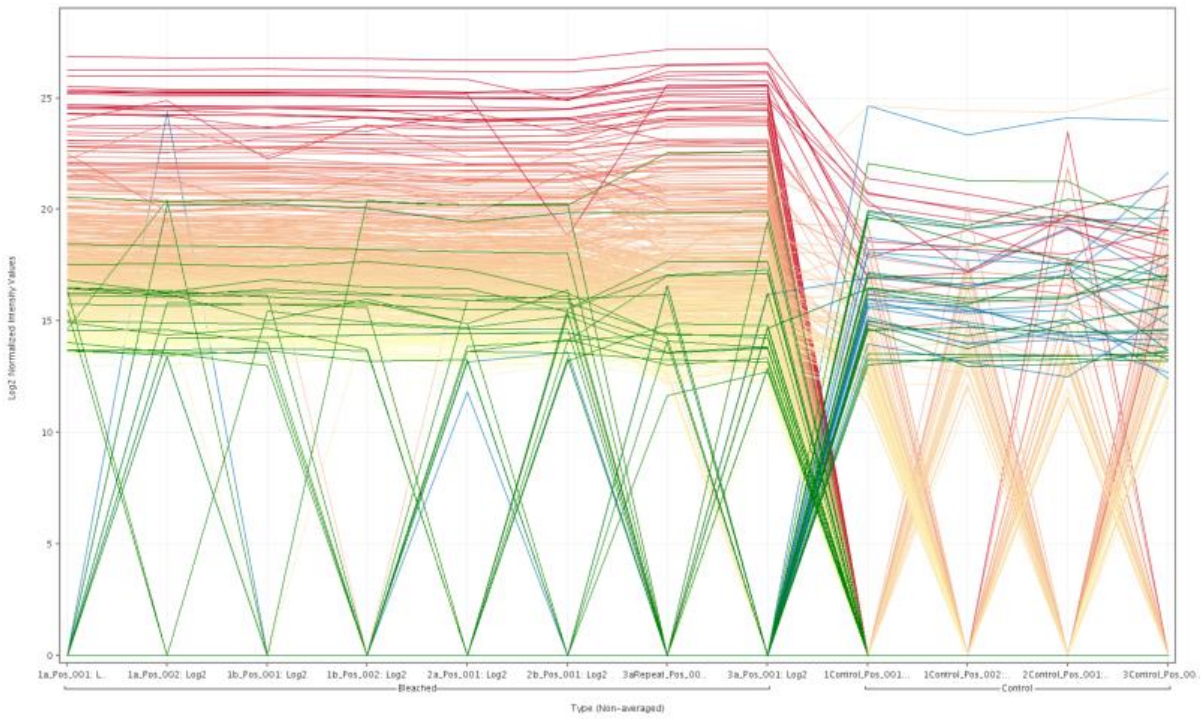

*Figure 3-21. MPP Feature Finding Graph on significant features with filtering by p-*value *and fold change.*

*Each line represents a feature that has a p-*value *≤ 0.01 and fold change greater than 30. The red lines have the largest abundance in bleached samples and the blue lines have the lowest abundance in bleached samples. The colours fade from red, orange, yellow, green to blue with decreasing abundance.*

There is some variance of abundances in each sample within each sample group, though the red lines are features of most significance which show the features of significant change between the groups, with the largest abundance in the bleached samples. The red lines are those showing the highest abundance in bleached samples, the gradient of colour to blue shows decreasing of abundance. The orange – yellow lines represent features of lower abundance than the red features, and these seem to be present in some of the control samples as well as in all the bleached samples. The green lines on the left are features present in only some bleached samples, though present in 100% of control samples.

The raw data are deconvoluted in Mass Profinder and exported as .cef files. Data are then re-analysed by MPP where this MPP significance plot gives 94 features that are the same as those given by Profinder out of 336 total features deemed unique to bleached hair.

In negative ionisation mode gives 58 features that are the same as those given by Profinder out of 276 total features deemed unique to bleached hair. This means 21% of the features are the exact same as those given by Profinder but majority are different.

Venn Diagram of Unique Compounds

The Venn diagram in **Figure 3.22** shows those which are unique to each group. It also shows the number of features that are common to both. For this biomarker discovery, ideally the potential biomarker needs to only be present in the bleached samples.

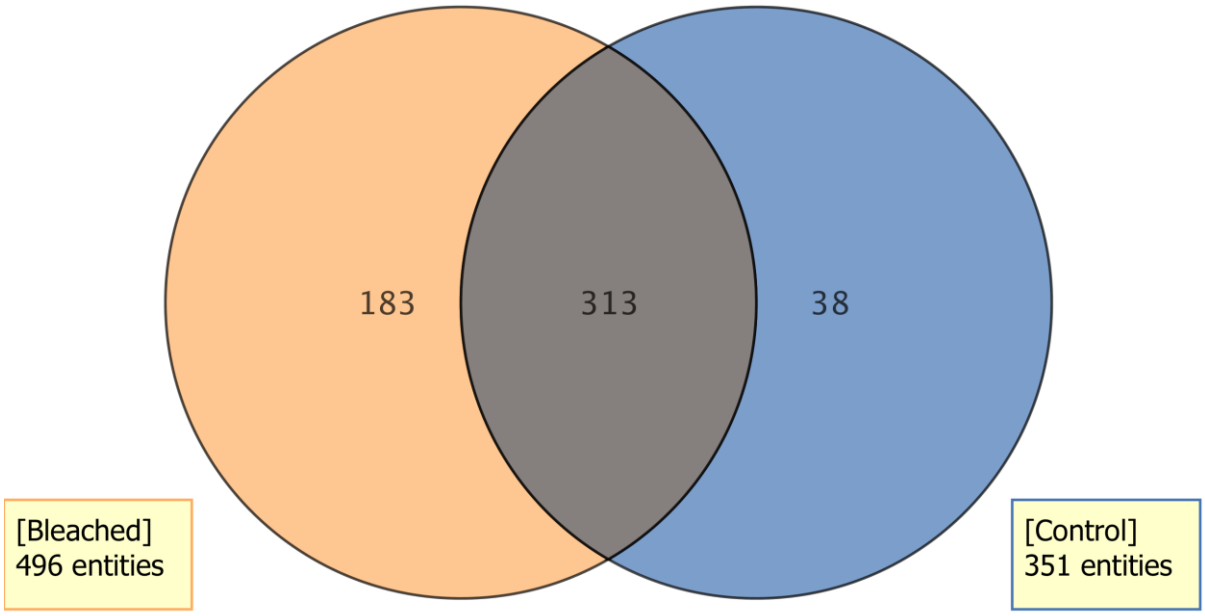

Figure 3-22. MPP Venn Diagram on Unique Compounds in Positive Mode.

There are 183 entities that are unique to bleached samples, and 38 that are unique to control samples.

In positive mode, out of the 183 features the Venn diagram says are unique, only 112 of these features are real features that give peaks in the mass spectrometry data. Of these 112, 21 of them are the same as those masses given in Mass Profinder, leaving 91 masses (81.25%) that it deems to be unique that Mass Profinder has not picked out.

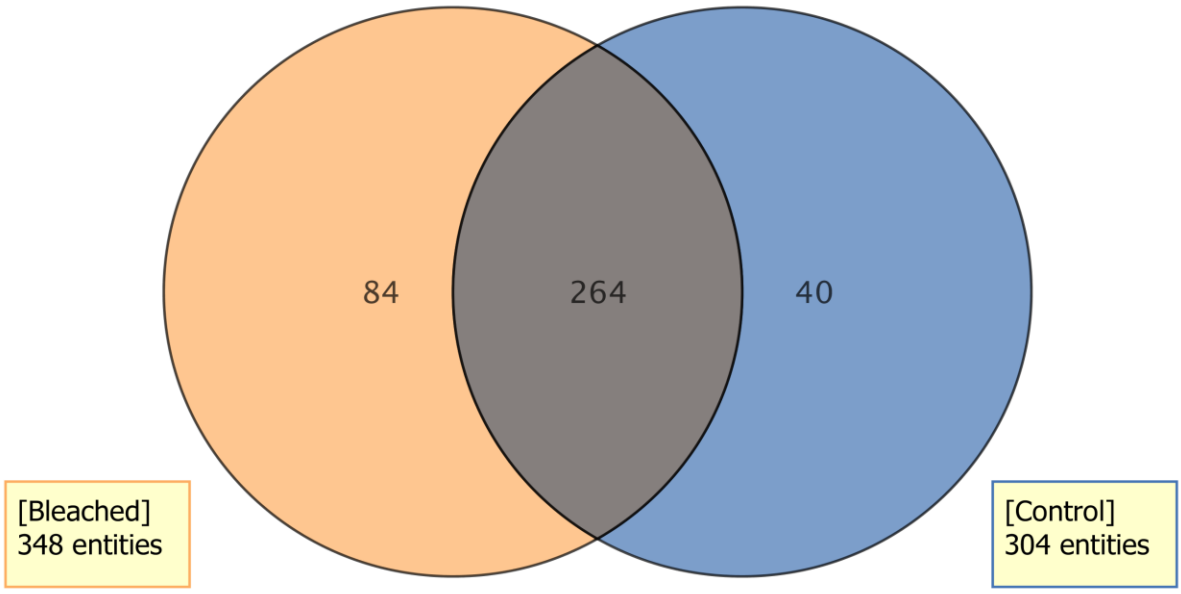

Figure 3-23. MPP Venn Diagrams of Unique Compounds in Negative Mode. Out of 264 features, 84 are found to be unique to bleached hair samples.

In negative ionisation mode, out of the 84 features the Venn diagram says are unique in **Figure 2.23**, only 50 of these features are real features that give peaks in the mass spectrometry data. Of these 50, only 2 of them are the same as those masses given in Mass Profinder. This is highly significant that majority of the compounds are different, even though the deconvolution only occurred once in Mass Profinder. This shows that analysis of the same compounds in different pieces of software, or even in the same software just by different techniques, gives different results.

3.6.3.5 XCMS Online

XCMS Online is a different type of software that is accessible online. The same parameters were used, as close as possible to those done in Agilent's software so that the end results are directly comparable though deconvolution is done by XCMS online rather than in Agilent software. The parameters have not been optimised to each data file, and have been left to the default settings that XCMS claim to have optimised themselves for general QTOF data files.

The features are those with a fold change greater than 30, with a *p*-value less than, or equal to 0.01. Only features after 1 minute are of significance in order not to be analysing those compounds in the solvent front.

In positive ionisation mode, 303 features are exported but give only 212 real features. A questionable 2 features are the same as those found by Mass Profinder, leaving 210 features that have been found unique to bleached hair by XCMS that are not present in Profinder data. A lot of these features can be seen to be a result of bad deconvolution in the software as there are 18 features that are all between 471.3577 – 471.4878 *m/z,* with retention time of either 6.8 or 7.2 minutes.

The negative ionisation mode showed better analysis results than positive mode ionisation, though majority of features are still different from those found by Mass Profinder. Out of 126 analytes that XCMS posed to be unique to bleached hair samples, 21 of these compounds are the same as those from Mass Profinder. The deconvolution parameters are tried to be the same where possible, yet only 16.7% of compounds are shown to be the same for features that are unique to bleached hair.

3.6.3.5.1 MetaboAnalyst

A new different website was discovered online which seems to give the same type of analysis that is provided by MPP. MetaboAnalyst, however, is free and therefore offers benefits to companies who will not have access to MPP, if it gives the accurate results. The data analysis gives the feature (neutral mass and RT), fold change (FC), log fold change (log2(FC), raw *p*-value (raw.*p*val) and the log value of the *p*-value (log10(*p*)).

Data can be inputted as a product of Mass Profinder and therefore the deconvolution has already been conducted. MetaboAnalyst re-analyses the data to give the features it believes are truly unique to bleached hair.

Out of 146 features, 36 features are present in the 'unique to bleached hair' category that was also found to be unique by Mass Profinder. This leaves nearly 75% of the features different between the different software packages despite the same deconvolution being conducted on the data. Performing analysis using the different software packages would give different results to the user, with even the top hits from both software packages being different.

This can be seen in both positive and negative ionisation modes, with only 49 features out of 192 unique to bleached hair being the same in negative ionisation mode. This means that 25.5% of features are the same, but the majority of features are different despite the deconvolution occurring only once in Mass Profinder. The differences occur in the second analysis conducted by MetaboAnalyst which has filtered out many false positives, but also seems to have filtered out some true positives as well.

3.6.4 Overall Results within Experiment Three

A compilation of the biomarkers that were determined to be present in more than one piece of software were taken, and a table of compounds for both positive and negative mode were made. The list of compounds was examined, and the top 15 were selected based upon the mass and RT. A wide range of masses were selected (between 100 – 800 Da) and the RT was between 0.8 – 11 min to stay clear of the solvent front and wash regions. Those compounds with the largest abundances were chosen first, then the gaps between the masses were filled with the next most abundant compounds.

3.6.4.1.1 Positive Mode:

	Mass Profinder	Mass Profiler	XCMS Online	MetaboAnalyst	MPP Venn	MPP Significance
Total Features	150	561	268	146	112	336
Features same as Mass Profinder	-	92	2	36	21	94
% Same	-	16.4	0.7	24.7	18.8	28.0
% of 150 features from Profinder	-	61.3	1.3	24.0	14.0	62.7

The significance test in MPP proves to have the greatest number of features the same in Profinder and MPP, giving 62.7% of the same features. XCMS Online has different deconvolution parameters and is done by a different software provider and only gives 0.7% of the total features the same as

Profinder proving to be the most dissimilar. Mass Profiler is also good, giving 61.3% of features that are the same as the 150 features given by Mass Profinder, and given that different deconvolution has occurred in the different software packages, it shows that these features are of high reliability to be truly unique features for bleached hair.

Of all the features that are present in Mass Profinder and at least one other software package, **Table 3.11** shows a list of the top 15 compounds that Experiment 3 determines to be possible biomarkers in positive mode. These have been selected based upon the highest abundance throughout the samples and the best spread of variability of mass and retention time.

Table 3.11. Potential biomarkers found in Experiment 3 Positive Mode

RT (min)	Mass (Da)	Predominant *m/z*
6.839	132.0683	133.0754
2.600	145.1467	146.1545
5.889	177.1262	178.1340
5.090	211.1307	212.1383
10.659	222.0647	223.0743
10.662	240.0784	241.0855
10.667	283.0838	284.0905
6.631	299.1808	300.1967
0.897	315.1466	316.1625 / 338.1417
9.019	399.2977	400.3033
9.244	421.2903	422.3058
7.662	431.2874	432.2965
3.331	544.2704	545.2782
4.027	588.2959	589.3034
5.285	646.3201	647.3279

3.6.4.1.2 Negative Mode:

	Mass Profinder	Mass Profiler	XCMS Online	MetaboAnalyst	MPP Venn	MPP Significance
Total Features	91	289	126	192	50	276
Features same as Mass Profinder	-	56	21	49	2	58
% Same	-	19.4	16.7	25.5	4.0	21.0
% of 91 features from Profinder	-	61.5	23.1	53.8	2.2	63.7

Table 3.12 shows a list of the top 15 compounds that Experiment 3 determines to be possible biomarkers in negative mode. The mass 243.0016 was selected due to its known presence as a

potential biomarker for bleached hair in the literature, PTeCA, despite its absence from the results in most of the software types.

Table 3.12. Potential biomarkers found in Experiment 3 Negative Mode

RT (min)	Mass (Da)	Predominant *m/z*
6.825	160.0636	159.0561
4.851	243.0016	241.9938
4.970	320.0131	319.0049
4.982	374.0389	373.0311
9.909	403.3286	402.3205 / 448.3254
9.937	411.3147	410.3068
8.751	417.3069	416.2994 / 462.3051
9.393	447.3178	446.3099
1.975	459.1556	458.3269
8.570	463.3125	462.305
10.545	519.3184	518.3099
8.857	531.3003	530.2925
6.843	563.2899	562.2819
4.971	594.0312	593.0230
9.928	669.4841	668.4763

3.6.5 Experiment Four – Secondary Check of Potential Biomarker Masses

The biomarkers determined in Experiment 3 were extracted in this dataset to investigate whether they are also present in the repeat dataset, and present across a larger range of hair colours/ethnicities/genders. These were extracted using MassHunter Qualitative Analysis within the raw data and the results can be seen in **Table 3.13 - 3.14** for positive and negative ionisation mode, respectively.

3.6.5.1 Positive Mode:

Table 3.13 shows the potential biomarker masses for positive ionisation mode with the results from Experiment 4.

Table 3.13. Comments on Potential Biomarker Masses Experiment 4 Positive Mode

RT (min)	Mass (Da)	Predominant *m/z*	Comments
6.839	132.0683	133.0754	Present
2.600	145.1467	146.1545	Not present in bleached samples MRS5 or MRS6
5.889	177.1262	178.1340	Present
5.090	211.1307	212.1383	Present
10.659	222.0647	223.0743	Present
10.662	240.0784	241.0855	Present
10.667	283.0838	284.0905	Present
6.631	299.1808	300.1967	Present
0.897	315.1466	316.1625 / 338.1417	Present
9.019	399.2977	400.3033	Present
9.244	421.2903	422.3058	Present
7.662	431.2874	432.2965	Present in grey control hair too
3.331	544.2704	545.2782	Not present in bleached samples MRS5 or MRS6
4.027	588.2959	589.3034	Not Present
5.285	646.3201	647.3279	Not Present

Out of the 15 compounds that were searched for in the positive ionisation mode analysis, only 10 of them were also present only in the bleached samples throughout these 6 hair samples. There were two compounds with neutral mass of 588.2959 Da (4.027 min) and 646.3201 Da (5.285 min) that were not present in any of the bleached samples. Compounds 145.1467 Da (2.600 min) and 544.2704 Da (3.331 min) were not present in MRS5 or MRS6 bleached samples, of which were a mixed colour of grey/black and brown/grey, respectively. Compound 431.2874 Da (7.662 min) showed presence of the peak in samples MRS4, MRS5 and MRS6 control samples, all of which have grey hair mixed with another colour. This suggests that the compound is present in grey hair as well as bleached hair.

3.6.5.2 Negative Mode:

Table 3.14 shows the potential biomarker masses for negative ionisation mode with the results from Experiment 4.

Table 3.14. Comments on Potential Biomarker Masses Experiment 4 Negative Mode

RT (min)	Mass (Da)	Predominant *m/z*	Comments
6.825	160.0636	159.0561	Present
4.851	243.0016	241.9938	Present
4.970	320.0131	319.0049	Not Present
4.982	374.0389	373.0311	Not Present
9.909	403.3286	402.3205 / 448.3254	Present
9.937	411.3147	410.3068	Present
8.751	417.3069	416.2994 / 462.3051	Present
9.393	447.3178	446.3099	Present
1.975	459.1556	458.3269	Not Present
8.570	463.3125	462.3050	Present
10.545	519.3184	518.3099	Present
8.857	531.3003	530.2925	Present in grey control hair
6.843	563.2899	562.2819	Not Present
4.971	594.0312	593.0230	Not Present
9.928	669.4841	668.4763	Present

3.7 Discussion

3.7.1 Discussion on Experiment 1 – Investigation into Bleaching Procedure

The bleach dissolved the hair samples, therefore leaving little, to no bleached samples left for weighing out. The bleaching procedure needed to be re-designed. Not chopping the hair samples before bleaching and leaving them with longer hair strands would be more ideal as it does not leave small sections vulnerable to the bleach.

3.7.2 Discussion on Experiment 2 – Investigation into Sample Preparation

Looking into the preliminary results, the concentration of the compounds present in the hair strands looked weak. There were concerns that it may not be ideal to try and gather structural information on a potential biomarker for bleached hair with samples that are weak. It was thought that more concentrated samples would be more ideal for the determination of biomarkers as more accurate

masses can be determined, enabling structural information to be gathered for accurate identifications.

3.7.3 Discussion on Experiment 3 – Discovery of Potential Biomarkers

Looking at the raw data, the peaks are looking good however they could be narrower and sharper. This may be due to the column particle size and therefore a 1.8uM diameter would give better resolution. The run time is also a little longer than it needs to be, there could be at least a minute saved from the wash at the end as it already reaching equilibrium before the run is completed.

Looking at the data analysis results, there are a few false positives and some that have identified the wrong isotope as the parent ion, leading to the wrong mass declared as the neutral mass with all the software types. There are also several features picked out at the wrong RT and some that seem to have multiple masses at the same RT, however they all have multiple isotopes leading to the conclusion that they are co-eluting compounds rather than fragments of one larger compound. These have been corrected and a list of 15 significant potential biomarkers are listed in **Table 3.15** - **3.16**. These are the most abundant masses that are present in more than one type of software, giving confidence that it is the best choice of biomarker since multiple pieces of software determined the compounds to be unique to bleached samples, minimising the chance that it is a one hit wonder or inaccurate mass or RT. It is of best interest for the biomarkers to elute during the middle of the run, staying away from the solvent front and the washes and so only the range 0.8 – 11.0 minutes have been included. The instrument that the potential biomarkers will be targeted on is used within the range 80 – 1000 *m/z* and therefore only compounds up to 800 Da have been chosen, though most compounds that are present in more than one piece of software seems to be between 400 – 600 Da.

The sample number for this Experiment was limited and so a wider range of hair types, across a larger sample set needed to be examined. The compounds shown in **Tables 3.15 – 3.16** are to be

extracted in this wider sample range, to determine whether they are truly biomarkers for bleached hair samples.

Positive Mode:

Table 3.15. Potential biomarkers found in Experiment 3 Positive Mode

RT (min)	Mass (Da)	Predominant *m/z*
6.839	132.0683	133.0754
2.600	145.1467	146.1545
5.889	177.1262	178.1340
5.090	211.1307	212.1383
10.659	222.0647	223.0743
10.662	240.0784	241.0855
10.667	283.0838	284/0905
6.631	299.1808	300.1967
0.897	315.1466	316.1625 / 338.1417
9.019	399.2977	400.3033
9.244	421.2903	422.3058
7.662	431.2874	432.2965
3.331	544.2704	545.2782
4.027	588.2959	589.3034
5.285	646.3201	647.3279

Negative Mode:

Table 3.16. Potential biomarkers found in Experiment 3 Negative Mode

RT (min)	Mass (Da)	Predominant *m/z*
6.825	160.0636	159.0561
4.851	243.0016	241.9938
4.970	320.0131	319.0049
4.982	374.0389	373.0311
9.909	403.3286	402.3205 / 448.3254
9.937	411.3147	410.3068
8.751	417.3069	416.2994 / 462.3051
9.393	447.3178	446.3099
1.975	459.1556	458.3269
8.57	463.3125	462.305
10.545	519.3184	518.3099
8.857	531.3003	530.2925
6.843	563.2899	562.2819
4.971	594.0312	593.0230
9.928	669.4841	668.4763

3.7.4 Discussion on Experiment 4 – Secondary Check of Potential Biomarker Masses

The results show that there are compounds that have potential as biomarkers in Experiment 3 but lacked potential when analysed in Experiment 4. This could be for various reasons, though a major reason is the presence of the compound in the grey control hair, or the lack of a potential biomarker in the hair colours that weren't used in Experiment 3 (mostly grey). There are some other compounds that are not present in Experiment 4 bleached samples, despite their strong presence in samples from Experiment 3. This shows that there is a need for repeat Experiments as these may be from instrument and/or solvent contamination.

A compilation of the masses that are present in both Experiment 3 and Experiment 4 have been listed in **Tables 3.17** & **3.18**. This is a great start to determine possible biomarkers as these are shown to be present in a range of different hair types, only in the bleached samples and not in any control samples.

The following compounds in **Table 3.17** are potential biomarkers determined in positive ionisation mode.

Table 3.17. Biomarkers in Experiment 3 & 4 Positive Mode

RT (min)	Mass (Da)	Predominant *m/z*
6.839	132.0683	133.0754
5.889	177.1262	178.1340
5.090	211.1307	212.1383
10.659	222.0647	223.0743
10.662	240.0784	241.0855
10.667	283.0838	284/0905
6.631	299.1808	300.1967
0.897	315.1466	316.1625 / 338.1417
9.019	399.2977	400.3033
9.244	421.2903	422.3058

The following compounds in **Table 3.18** are potential biomarkers determined in negative ionisation mode.

Table 3.18. Biomarkers in Experiment 3 & 4 Negative Mode

RT (min)	Mass (Da)	Predominant *m/z*
6.825	160.0636	159.0561
4.851	243.0016	241.9938
9.909	403.3286	402.3205 / 448.3254
9.937	411.3147	410.3068
8.751	417.3069	416.2994 / 462.3051
9.393	447.3178	446.3099
8.570	463.3125	462.3050
10.545	519.3184	518.3099
9.928	669.4841	668.4763

The compounds in **Tables 3.17 – 3.18** are to be taken forward for further investigation in possible future work to determine structural information and identifications. Using the current data, any possible annotations that have been found by using the Metlin 8.0 database have been listed below in **Tables 3.19** & **3.20**, however confirmation of these identifications has not been completed. This would require the standards of each to be ordered and MS/MS analysis to be completed by determining transitions of each compound and checking these create the same fragmentation

patterns in the spectrum. If there is no identification possible, then the empirical formula suggested in MassHunter Qualitative Analysis has been given instead.

Table 3.19 shows possible annotations of the potential biomarkers determined in positive ionisation mode. Where annotations couldn't be determined, a possible formula has been suggested from the accurate mass.

Table 3.19. Identifications of biomarkers in Experiment 3 & 4 Positive Mode

RT (min)	Mass (Da)	Putatively Annotated Compounds or Formula	PubChem Compound Identifier (CID)
6.839	132.0683	4-Methylpyrrolo[1,2-a]pyrazine	583433
5.889	177.1262	$C_{10}H_{15}N_3$	-
5.090	211.1307	$C_7H_{20}ClN_4O$	-
10.659	222.0647	Threoninyl-Cysteine	18218245
10.662	240.0784	Chrysophanic acid 9-anthrone	68111
10.667	283.0838	(E)-Avenanthramide D	15607909
6.631	299.1808	$C_{12}H_{25}N_7S$	-
0.897	315.1466	$C_{10}H_{19}N_8O_4$	-
9.019	399.2977	$C_{20}H_{39}N_4O_4$	-
9.244	421.2903	$C_{22}H_{46}ClN_2OS$	-

Table 3.20 shows possible annotations of the potential biomarkers determined in positive ionisation mode. Where annotations couldn't be determined, a possible formula has been suggested from the accurate mass.

Table 3.20. Identifications of biomarkers in Experiment 3 & 4 Negative Mode

RT (min)	Mass (Da)	Putatively Annotated Compounds or Formula	PubChem Compound Identifier (CID)
6.825	160.0636	3-methyl-2-Quinoxalinone	26384
4.851	243.0016	1H-pyrrole-2,3,4,5-tetracarboxylic acid (PTeCA)	19377973
9.909	403.3286	$C_{20}H_{43}N_4O_4$	-
9.937	411.3147	Cyclopamine	442972
8.751	417.3069	$C_{18}H_{39}N_7O_4$	-
9.393	447.3178	$C_{21}H_{43}N_4O_6$	-
8.570	463.3125	$C_{20}H_{37}N_{11}O_2$	-
10.545	519.3184	$C_{33}H_{47}NP_2$	-
9.928	669.4841	$C_{30}H_{67}N_7O_7S$	-

3.7.5 Summary of Experiments

Overall, there are several compounds in both positive and negative ionisation mode that have great potential as biomarkers for bleach in hair. These have been present throughout different samples, different sample preparations, different LC runs on different days etc., which shows that they are consistent so far. There are some possible identifications for these masses, though these are only preliminary from the neutral mass; no standards or MS/MS Experiments have been run/conducted.

3.7.6 Limitations

There are many limitations when determining possible biomarkers that need to be evaluated. Some of these limitations have been considered but were limited in this project due to the availability of resources and volunteers.

The types and colours of hair available were limited to the range of volunteers. This means that despite the types of hair including fine, medium, thick, wavy, coiled and straight, only a small range were tested, with no samples of coiled or thick hair included. Fine, straight hair described most of the samples used; thick coiled hair may not include some of the potential biomarkers determined. Hair types are determined with genetics. The different type of hair may affect the uptake of bleach, and they will have different compounds in the hair meaning that the compounds present in the hair after cosmetic alteration may be different depending on the hair type. Hence, a range of different hair types was tested, though a wider range should be tested with multiple repeats of each. Ethnicity also plays a part in hair type, but it is unknown as to whether the ethnicity of a person may affect the potential biomarkers. Only three different ethnicities were tested during this project, which is clearly a very limited range and other volunteers should be sought, where possible.

Similarly, the natural hair colour of a person affects the bleach biomarker availability. This is known through the discovery that some potential biomarkers that were discovered during Experiment Three in **Section 3.6.3** were found to be present in grey control hair in **Section 3.6.4**. The differing levels of

melanin amongst the different hair colours mean that the bleach uptake is variable. Therefore, different concentrations of compounds will be present after bleaching different coloured hair. For example, light blonde hair will go extremely light blonde/white where-as black hair will only lighten to an orange/yellow colour. Therefore, a range of hair colours were tested, although a more extensive range is needed. Hair colours like auburn, strawberry blonde and albino were not tested which limits the reliability since these markers may not be present on these hair colours.

It is known that compounds are present in different levels up and down the hair shaft. For example, PTCA is present throughout the hair but in varying concentrations therefore it may fall below a cut off value at one section of the hair, where-as it will be in significant concentration at another section. This is due to the different exposure time the hair has had to the atmosphere and UV light especially. Other compounds are likely to be affected like this also and therefore, the test needs to be conducted on a range of hair lengths with testing at each section from the proximal end to the distal end of a long section of hair. In this project, a variety of different hair lengths were used however it was unknown how far from the scalp the sections were on and therefore a more in-depth Experiment would need to be conducted to determine whether the concentrations are significantly different along the hair shaft.

It has not been tested to see what happens chemically if someone was to dye their hair, then bleach it. This experiment was conducted only with completely untreated control hair, which was then bleached. A further Experiment with a wide selection of hair dye before bleaching could also be conducted. It is interesting to determine whether the markers are the same with both bleach and dye, or whether the presence of dye alters the biomarkers. It is also known that cosmetic procedures such as heat styling, affects the chemical compounds in hair and therefore it would need to be testing that these potential biomarkers are still present after bleaching and repeated heat styling. Effects of how the cosmetic bleaching procedure is conducted may alter the concentration also. The biomarkers need to be present in a range of different bleaching products, including products from

the high street shops and hair salons. When hair is bleached in a salon, a cosmetic solution called a 'toner' is often used after bleaching to cover the yellow created by the bleach. It is unknown whether the toner affects the biomarker presence, and/or concentration, and so this needs to be investigated.

A study on how any potential biomarker behaves over time would need to be conducted. It may decrease in concentration throughout the lifetime, or increase. Any changes need to be determined, which may be natural decrease from UV light. This also includes the effects of pollution and repeated washing of the hair throughout the lifetime. Certain brands of shampoo may remove the biomarker quicker than another and if it does, it is interesting to know whether this rate of decrease is comparable to the rate that drug incorporation in the hair shaft also decreases.

Another thing to consider is that any biomarkers declared would need to be unique to bleached hair, so it needs to be confirmed that it is not present naturally in anything else, such as certain foods or drinks. It cannot be prescribed for any reasons in legal drugs or produced in illegal drugs as a bi-product. There shouldn't be any other reason that the biomarker can be in the hair other than bleach itself.

Some other limitations of this project have made parts more difficult to determine a biomarker. The aim was to determine what additional information could be obtained whilst keeping the sample preparation and LC-MS analysis as closely related to the SoHT approved procedures as possible. Alternate potential biomarkers may have come to light if using different solvents, columns, different phases, such as HILIC based methods or ionic exchange. The databases that are accessible are also limited and therefore not many identifications have been possible. There was also no QC sample used which could've improved the Experiment as this would've given more confidence that the instrument was working at the same level consistently and no issues occurred mid analysis. The instrument is monitored daily with independent standards to ensure that the sensitivity and

resolution is maintained. This is run before and after each run so that any deterioration within the instrument parameters is detected and records the overall performance over time.

3.8 Software Comparison and Discussion

Data analysis of untargeted metabolomics studies presents a key challenge as it requires extensive processing of thousands of features from the raw MS data. There are multiple pieces of software that have been developed to handle this data processing, however it has not yet been studied whether the different pieces of software give the same results, or which one is best in terms of feature detection and unique compound analysis. It's also interesting to discover which piece of software is the easiest to use, giving the most accurate results in the most simple, quickest way.

The different pieces of software compared throughout Chapter 3 are:

- XCMS Online
- Mass Profiler (Agilent)
- Mass Profinder (Agilent)
- Mass Profiler Professional, MPP (Agilent)
- MetaboAnalyst (Online)

Biomarker Detection

In **Tables 3.21 – 3.22**, the presence of each potential biomarker across the different pieces of software is described in both positive and negative ionisation mode.

Table 3.21. Table of biomarkers with description of presence (+) or absence (-) across the different pieces of software, in positive ionisation mode

RT (min)	Mass (Da)	XCMS Online	Mass Profiler	Mass Profinder	MPP	MetaboAnalyst
6.839	132.0683	-	+	+	+	+
5.889	177.1262	+	-	+	+	+
5.090	211.1307	-	+	-	+	-
10.659	222.0647	+	-	-	+	-
10.662	240.0784	-	+	+	+	-
10.667	283.0838	-	+	+	-	-
6.631	299.1808	-	-	-	+	+
0.897	315.1466	-	+	+	-	-
9.019	399.2977	+	+	-	+	+
9.244	421.2903	+	-	-	+	+

Table 3.22. Table of biomarkers with description of presence (+) or absence (-) across the different pieces of software, in negative ionisation mode

RT (min)	Mass (Da)	XCMS Online	Mass Profiler	Mass Profinder	MPP	MetaboAnalyst
6.825	160.0636	+	-	-	-	+
4.851	243.0016	-	+	-	-	-
9.909	403.3286	+	+	-	+	+
9.937	411.3147	-	+	+	+	+
8.751	**417.3069**	+	+	+	+	+
9.393	447.3178	+	+	-	+	+
8.570	463.3125	+	+	-	+	+
10.545	519.3184	+	-	+	+	+
9.928	669.4841	+	-	-	+	+

Looking at the **Tables 3.21 – 3.22** there is one biomarker that is consistently present throughout all the different pieces of software, 417.3069 Da at 8.7 min in negative ionisation mode. There are no other masses that are extracted in all the software analysis results. There is a particular absence of 243.0016 Da in the software results, despite it being published as a potential biomarker. Looking at the raw data, this is an u*p*-regulated compound in bleached hair consistently, though compared to other features in the data it is low abundance. Therefore, it may have fallen through below the cut off level within the parameters in Mass Profinder, hence its absence in MPP and MetaboAnalyst as well.

All these masses in **Tables 3.21 – 3.22** have been extracted in the raw data and give positive results for biomarker potential, showing that each piece of software has some issues since they are showing false negatives, as well as the false positives previously seen. Positive and negative ionisation mode give different reliability results, though overall MPP seems to be the most accurate as it has only missed 2 of the selected biomarker masses in positive mode and 2 masses in negative mode. **Table 3.23** shows the number of biomarkers, across both positive and negative mode, detected out of a possible 19 biomarkers, 10 in positive mode and 9 in negative mode. The table shows MPP gives an overall accuracy of 79% for the selected biomarker masses, which is the highest % overall.

Table 3.23. Number of biomarkers detected overall in each piece of software, compilation of both positive and negative ionisation mode.

Features Detected	XCMS Online	Mass Profiler	Mass Profinder	MPP	MetaboAnalyst
Overall Accuracy (Out of 19)	11	12	8	15	13
Overall Accuracy %	58	63	**42**	**79**	68

Mass Profinder analysis gives the least accurate results of only 42%. It can be assumed that this is due the data only having the one processing method, compared to MPP and MetaboAnalyst which get re-processed in the software after the Mass Profinder results are exported, with further statistical analysis. Looking into the raw tabulated data produced by Mass Profinder, all the biomarker masses from **Tables 3.21 – 3.22** are present, however they also have a significant value in the 'control' column, suggesting that there is a sample with the compound present in the control group, hence it has not been flagged as a unique compound. These are not real peaks in the control samples when looking into the raw data and so the re-processing within MPP and MetaboAnalyst has captured the false positive within the control samples. Hence, they then appear as unique markers present only in bleached samples for the secondary data analysis in MPP and MetaboAnalyst.

Overall Feature Detection

Table 3.24 shows the number of features detected by each piece of software overall. The data are filtered so that the features must be present in 100% of samples, within at least one group, however

this was not possible to filter by this parameter in XCMS online, hence the large difference in features detected.

Table 3.24. Number of features detected overall in each piece of software, with no filtering.

Features Detected	XCMS Online	Mass Profiler	Mass Profinder	(MPP)	(MetaboAnalyst)
Positive Mode	27033	7471	5216	(5216)	(5216)
Negative Mode	8219	4387	3698	(3698)	(3698)

The overall features detected are the same for Mass Profinder, MPP and MetaboAnalyst because the data files were exported as .cef and .csv files from Mass Profinder for the data analysis in MPP and MetaboAnalyst. Therefore, the overall features detected from the raw data were not conducted using MPP and MetaboAnalyst. It is important to note that using the .cef and .csv files from Mass Profinder, the data are then re-analysed in MPP and MetaboAnalyst with statistical analysis. Hence, the results from each are different to each other with different presence of unique features.

When comparing XCMS online, Mass Profiler and Mass Profinder, the most features were detected in XCMS online, then Mass Profiler followed by Mass Profinder. However, XCMS online did not give the ability to filter the features detected by 'present in 100% samples in at least one sample group' as the other pieces of software did. It can also be seen that the features detected in XCMS online are false positives as when the mass is checked in the raw data, a lot of these masses do not agree with the XCMS online results. Some of the peaks that it has picked are not 'real' as only the noise is integrated. The peak picking parameters have already been optimised and so it was determined that the software detects a significant number of false positives. Looking at the paper by Myers et al., there seems to be a several known problems with the peak detection algorithm in XCMS online, *centWave,* which causes many false positives [31-32].

Software Comments

Looking at the individual pieces of software, Mass Profinder is a standout piece of software that would make the analysis more difficult without its use. It gives the ability to conduct peak picking with a wide range of parameters that can be adapted to the user's needs, then exported as various formats (i.e., csv and cef files) that can be taken forward into other pieces of software. The visual details of the EICs and mass spectra given within the software means that the data can be checked internally. This enables the detection of retention time drift or other issues that may be occurring, such as decrease of abundance of one peak over the analysis run time, which suggests potential instability of the samples. The data can be filtered in many ways, such as 'present in only 2 samples' or 'max abundance greater than 50,000', all of which are customisable. The data can be exported and analysed directly in excel, though the results show that it is more beneficial to use another piece of software after Mass Profinder. The platform is free to use for those who purchase the Agilent MassHunter suite.

Mass Profiler Professional (MPP) uses the .cef files generated from Mass Profinder and re-analyses the data with narrower constraints on each parameter, for example a 10 ppm mass error window rather than the 15 ppm window given in Mass Profinder. This enables the abundance of each feature can be checked, allowing further filtering of samples. This checking of the features means that there are some additional features that may be unique, or up/down regulated, to one sample group that was missed in the Mass Profinder analysis, but it also means that some features are filtered out that were previously tagged as potential biomarkers due to the narrower alignment or altered abundance filters. The use of the .cef files does mean that the results are dependent on those provided by Mass Profinder and so if the initial collected data isn't accurate or is too constricted, the MPP results will also be inaccurate and potentially important data will be missed. The visual techniques available in MPP are very useful, especially the PCA plots that give a nice overview of the analysis results. The results tables can be filtered to the user's needs, allowing features of a specified *p*-values or fold changes to be extracted into a separate table. This, along with the Venn diagrams and feature graphs

of unique and up/down regulated compounds, is a useful technique used for the determination of potential biomarkers. From the results, this software is shown to be reliable and accurate with its results, though it does have some false negatives. For academic purposes, MPP comes at a one off charge of £20,500 and updates within the version you buy are provided at no extra cost. If you wish to upgrade to the newest version, it comes at a cost of £11,600 and the license key is moved across so only one version can work at a time. E.g., if you purchase MPP 12, then 12.1, 12.2, 12.3 etc., can be used for free, but an upgrade fee for MPP 15.1 would be required. Given the accurate information that it provides in the simplest ways and the large range of analysis that can be conducted, the cost is justified.

MetaboAnalyst provides the next most accurate results overall, after MPP. It uses the .csv files generated from Mass Profinder and is therefore dependent on the results from this, though there are other ways that the data can be inputted. It is a free website available to anyone online, and therefore is useful for many who do not have the means to cover the cost of MPP. It is possible to do multivariate statistical analysis, along with many other functions provided by MPP however it is not always as detailed with the plots given. It gives the basic plots but it is not able to customise each one, therefore showing it is more basic than MPP which is expected given the cost difference. It may also be more difficult to provide highly complex graphics using an online website, in comparison to the software provided by Agilent for MPP version 15.1. Each graphic and table provided within MetaboAnalyst online can however be downloaded as individual plots, images and pages, or the full analysis can be exported as a pdf report. MetaboAnalyst can also be used in conjunction with R, something that has not been investigated during this project due to time constraints and learning resources available.

Mass Profiler can be used by inputting .cef files, or as the raw .d data files. The raw data files were used in this project and it shows that there are some inconsistencies compared to the other software used, however it gives some accurate results. It is quicker to get the results than it is using Mass

Profinder, however there are less graphics and it is more difficult to assess the results given. It gives a graph of compounds unique to a sample group which is useful, however it requires further checking of the data as a lot of times the results were inaccurate when the raw data was checked in MassHunter Qualitative. The parameters were optimised however, the results still showed a peak as absent in control samples when the raw data showed that it was in fact present. The reasoning behind this is unknown but assumed to be a clash within the chosen parameters not working for each feature selected. The platform is free to use for those who purchase the Agilent MassHunter suite although it is limited in the statistical analysis possible. It can only look at 1 vs 1 sample groups, so QC samples or blanks cannot be assessed at the same time as sample groups which is a limitation. It is useful for a quick overview of the data whilst waiting for more in-depth analysis conducted using Mass Profinder.

XCMS online is the least accurate software used, and it has the most false positive features discovered, despite being the most cited pre-processing tool used in the literature [31]. This has showed to be a problem with the peak picking conducted and potentially someone with more computer programming knowledge may know how to adapt all the parameters to optimise them further than the default settings. It has a variety of techniques and statistical analysis available, though these are basic and limited, with no ability to zoom or customise each one. There is a 'help' function on the website, though this never generates any response from the developers or experts. The website was free to use, though there is a way to pay for XCMS Plus at a yearly cost of approximately £7300 for a company, with renewals costing £3000 per year. Without access to the paid version, it cannot be determined whether it will be any more accurate or reliable than the free version.

Overall, the best software program seems to be Profinder combined with MPP as it has a wide range of opportunities for feature extraction, biomarker discovery with visual features and reliable results. As someone who has access to Profinder and MPP, it is a combination of software techniques that

provide accurate results with visual ways to display the data in a simple way. It can be applied to a various range of samples and is easy to use, once you have acquired knowledge of the best parameters that work for your dataset.

3.9 Conclusion

3.9.1 Summary of Biomarker Results

The data that has been gathered in this work has shown overall to be reproducible and reliable since there are multiple masses that have repeatedly been determined as potential biomarkers. A QC sample would've improved the Experiment and should be considered in the future.

The statistical approach taken in this research has shown to provide a range of potential biomarkers, though also has provided with a significant number of random masses that are not of interest due to the lack of presence in repeat Experiments. The multivariate statistical analysis has shown some group separation however, it shows that the clustering also occurs within the samples taken from a person where the control and bleached samples from one person sometimes cluster closer than the overall bleached and control samples. This could be for a multitude of reasons including the cosmetics, diet, pollution, hair colour or ethnicity of a person being closely related to that of another, though being completely different to another person. The levels of melanin in hair will significantly affect the uptake of bleach and therefore it is more likely that those of similar hair colours will be clustered closer to those of similar natural hair colour than someone who has a totally opposite colour, especially albino or grey hair compared to black or blonde.

Overall, this untargeted study has confirmed that there are differences between bleached and control hair samples in LC-MS analysis and a range of potential biomarkers have been given.

3.9.2 Summary of Software Comparison

More than one processing software should be used to avoid missing the presence of a potential biomarker or missing the overall chemical signature of a sample group. Overall, the best software program pair is Profinder combined with MPP as it has a wide range of opportunities for feature extraction, biomarker discovery with visual features and the most reliable results. The statistical approach used shows a range of potential biomarkers in all the software types, though each one has provided a significant number of false positives, false negatives and several masses that lack consistency throughout repeat Experiments. Some software seems more reliable and accurate than others. The Mass Profinder and MPP results combined are the most accurate with most of them positively checked against the raw data. Mass Profiler provides significantly more false positives than the other platforms used, and Mass Profinder when used as a standalone piece of software also has many false positives.

3.9.3 Future Work

Now that there are some potential masses that could lead to be a biomarker for bleached hair, the next step would be to try and find out the identification of these compounds. The masses also need confirming on an even larger scale, with multiple different hair types. A wider selection of ethnicities and hair colours are needed, including albino and red hair, with a range of ages from children to elderly adults. This is to ensure that the compound isn't present at any stage of a person's life. The study on whether these masses occur in natural hair in any section along the hair shaft would need to be determined to ensure it is truly unique to bleached hair. The compounds would also need to be tested across a wide time scale to see how long it is detectable in the hair for, and if the rate in which it decreases is similar to the rate that the drug concentration also decreases.

3.10 References

1 Pragst F, Balikova M. State of the art in hair analysis for detection of drug and alcohol abuse. Clinica Chimica Acta. 2006;370(1-2):17-49.

2 Eisenbeiss L, Binz T, Baumgartner M, Kraemer T, Steuer A. Cheating on forensic hair testing? Detection of potential biomarkers for cosmetically altered hair samples using untargeted hair metabolomics. The Analyst. 2020;145(20):6586-6599.

3 Van Elsué N, Yegles M. Influence of cosmetic hair treatments on cannabinoids in hair: Bleaching, perming and permanent coloring. Forensic Science International. 2019;297:270-276.

4 Morini L, Zucchella A, Polettini A, Politi L, Groppi A. Effect of bleaching on ethyl glucuronide in hair: An in vitro Experiment. Forensic Science International. 2010;198(1-3):23-27.

5 Petzel-Witt S, Pogoda W, Wunder C, Paulke A, Schubert-Zsilavecz M, Toennes S. Influence of bleaching and coloring on ethyl glucuronide content in human hair. Drug Testing and Analysis. 2017;10(1):177-183.

6 Informed Health Online 2006 - What is the structure of hair and how does it grow?. [Cologne, Germany]: Institute for Quality and Efficiency in Health Care (IQWiG); 2019. Accessed via Available from: https://www.ncbi.nlm.nih.gov/books/NBK546248/ August 2022.

7 Bolognia J, Jorizzo JL, Rapini RP. Biology of Hair and Nails. In: Dermatology. St. Louis, MO: Elsevier; 2012:1075–92.

8 Odya E, Norris M. Hair anatomy. In: Anatomy & Physiology. Hoboken, NJ: John Wiley & Sons, Inc.; 2017.

9 Buffoli B, Rinaldi F, Labanca M et al. The human hair: from anatomy to physiology. Int J Dermatol 2014;53: 331–341.

10 Centre R. Hair Structure - Everything You Need To Know [Internet]. Rejuvenate Admin. 2022 [cited 23 August 2022]. Available from: https://rejuvenatehairtransplant.com/blog/hair-structure/

11 Krause K, Foitzik K. Biology of hair follicle: the basics. Semin Cutan Med Surg 2006;25:2–10.

12 Jones LN. Hair structure anatomy and comparative anatomy. Dermatol Clin 2001;19:95–103.

13 Eisenbeiss L, Steuer AE, Binz TM, Baumgartner MR, Kraemer T. (UN)targeted hair metabolomics: First considerations and systematic evaluation on the impact of sample preparation. Analytical and Bioanalytical Chemistry. 2019;411(17):3963–77.

14 Chang W, Wang P, Chang C, Chen Y, Liao P. Extraction strategies for tackling complete hair metabolome using LC-HRMS-based analysis. Talanta. 2021;223:121708.

15 Barbosa J, Faria J, Carvalho F, Pedro M, Queirós O, Moreira R, et al. Hair as an alternative matrix in Bioanalysis. Bioanalysis. 2013;5(8):895–914.

16 Jurado C, Kintz P, Menéndez M, Repetto M. Influence of the cosmetic treatment of hair on drug testing. International Journal of Legal Medicine. 1997;110(3):159-163.

17 Skopp G, Pötsch L, Moeller M. On cosmetically treated hair — aspects and pitfalls of interpretation. Forensic Science International. 1997;84(1-3):43-52.

18 Kerekes I, Yegles M. Coloring, bleaching, and perming: influence on EtG content in hair. Therapeutic Drug Monitoring. 2013;35(4):527-529.

19 Ettlinger J, Kirchen L, Yegles M. Influence of thermal hair straightening on ethyl glucuronide content in hair. Drug Testing and Analysis. 2014;6(S1):74-77.

20 Binz T, Baumgartner M, Kraemer T. The influence of cleansing shampoos on ethyl glucuronide concentration in hair analyzed with an optimized and validated LC–MS/MS method. Forensic Science International. 2014;244:20-24.

21 Ramírez Fernández M, Wille S, Di Fazio V, Samyn N. Influence of bleaching and thermal straightening on endogenous GHB concentrations in hair: An in vitro Experiment. Forensic Science International. 2019;297:277-283.

22 Joo K, Kim A, Kim S, Kim B, Lee H, Bae S et al. Metabolomic analysis of amino acids and lipids in human hair altered by dyeing, perming and bleaching. Experimental Dermatology. 2016;25(9):729-731.

23 Grosvenor A, Deb-Choudhury S, Middlewood P, Thomas A, Lee E, Vernon J et al. The physical and chemical disruption of human hair after bleaching - studies by transmission electron microscopy and redox proteomics. International Journal of Cosmetic Science. 2018;40(6):536-548.

24 Mantinieks D, Gerostamoulos D, Wright P, Drummer O. The effectiveness of decontamination procedures used in forensic hair analysis. Forensic Science, Medicine and Pathology. 2018;14(3):349-357.

25 Steuer A, Brockbals L, Kraemer T. Untargeted metabolomics approaches to improve casework in clinical and forensic toxicology—"Where are we standing and where are we heading?". WIREs Forensic Science. 2021;4(4):1449.

26 Petzel-Witt S, Meier S, Schubert-Zsilavecz M, Toennes S. PTCA (1H-pyrrole-2,3,5-tricarboxylic acid) as a marker for oxidative hair treatment. Drug Testing and Analysis. 2017;10(4):768-773.

27 Eisenbeiss L, Binz T, Baumgartner M, Steuer A, Kraemer T. A possible new oxidation marker for hair adulteration: Detection of PTeCA (1H-pyrrole-2,3,4,5-tetracarboxylic acid) in bleached hair. Drug Testing and Analysis. 2019;12(2):230-238.

28 Guerra-Tapia A, Gonzalez-Guerra E. Hair Cosmetics: Dyes. Actas Dermo-Sifiliográficas (English Edition). 2014;105(9):833-839.

29 Harrison S, Sinclair R. Hair colouring, permanent styling and hair structure. Journal of Cosmetic Dermatology. 2003;2(3-4):180-185.

30 Cooper G, Kronstrand R, Kintz P. Society of Hair Testing guidelines for drug testing in hair. Forensic Science International. 2012;218(1-3):20-24.

31 Li Z, Lu Y, Guo Y, Cao H, Wang Q, Shui W. Comprehensive evaluation of untargeted metabolomics data processing software in feature detection, quantification and discriminating marker selection. Analytica Chimica Acta. 2018;1029:50-57.

32 Myers O, Sumner S, Li S, Barnes S, Du X. Detailed Investigation and Comparison of the XCMS and MZmine 2 Chromatogram Construction and Chromatographic Peak Detection Methods for Preprocessing Mass Spectrometry Metabolomics Data. Analytical Chemistry. 2017;89(17):8689-8695.

4 – Differences in the Metabolite Profile Between a Control and Diseased Sample Set of Dried Blood Spots

4.1 Aims

This study investigated the difference in metabolic content between a control and diseased group to determine a biomarker for Major Depressive Disorder in people. The objective was to develop a reliable, repeatable and accurate workflow to determine if there are any possible biomarkers. The results obtained by the different types of comparison software packages used were assessed to determine if there are any differences between them; ideally, they should all give the same results. Using these liquid human blood samples, the untargeted comparison software techniques were evaluated to compare whether the same biomarkers are determined throughout.

4.2 Introduction

4.2.1 Use of Dried Blood Spot Analysis

Dried blood spot (DBS) analysis is not over popular within the metabolomics industry with only few people using this technique. Blood based biofluids, such as serum and plasma, tend to have more extensive studies however they require stringent storage conditions and require specialist healthcare professionals to collect the samples. Metabolite degradation can occur when the samples are not stored in the correct conditions which can lead to false results [1].

In 1913, Ivar Bang first described the use of dry blood spots however it wasn't until 1963 that Robert Guthrie described the concept that blood, obtained by pricking the finger or heel and blotting it onto some filter paper, could be used to screen for different metabolic diseases in neonates in Scotland [2]. Following this, dried blood spots have been used routinely since the 1960s for new-born child screening [3]. Screening for phenylketonuria in neonates became a nationwide task in 1969-70 but since then, Guthrie cards have been collected routinely to screen for other disorders in infants [3].

More recently, the sample cards have been collected in over 20 countries to screen for congenital hypothyroidism, sickle cell disorders and HIV infections also [3]. Previously, the use of dried blood spots has been restricted by the lack of sensitivity and specificity given the small volumes of blood however, recent advances in knowledge and instrumentation have overcome many of these problems. Given this, dried blood spot analysis is now a well-known technique used throughout the world [2].

Dry blood spots are collected and stored on filter paper cards and therefore are an attractive alternative when it comes to storage, shipment and analysis of liquid blood samples. Collection of the blood spot sample can be done in various ways, however, one of the most common methods for the preparation is the use of fingerstick lancets. It is minimally invasive in comparison to other sample collection techniques, such as venepuncture where a specially trained healthcare professional (phlebotomist) takes whole blood. DBS collection also is beneficial as it only requires a small amount of blood, which can be collected by the patients themselves by following a set of instructions. This can then be sent back to the labs by regular mail therefore only requiring low cost [4]. The lancet is therefore a great method of choice given that it is easy to sample, painless, cheap and gives a large representation of a person in one small drop of blood. The risk of bacterial contamination is minimal also and the storage time of DBS sample cards is relatively long in comparison to liquid blood samples due to their stability of analytes when dried and kept in a packet with a desiccator. They also can be kept at room temperature (with a desiccant in the packet with the card) without worsening the results of the analysis. Unlike the use of blood plasma, the sample does not need preparing immediately [5]. DBSs are also less dangerous to handle than other biofluids, such as blood and plasma as the dried blood inactivates pathogens, lowering the biohazard risk [5].

However, due to the small volume of blood collected in DBS techniques, the potential target analyte may be of a rather low concentration, potentially less than 1 ng/L. This would require an extremely sensitive method for the detection and quantification of the substance; mass spectrometry is the

most common method used for such analysis [6-8]. Another problem is that it is difficult to know how much blood is collected on the filter card each time, therefore direct comparison with another sample may be difficult. When the blood is spotted, there is a procedure to follow which includes letting the 2nd drop of blood flow onto the card itself with no direct contact of the finger with the card. It is difficult to know whether the same amount of blood drops each time since the diameter of the spot will be different each time and the loading of the spot will be different. The diameter can be overcome by using the same width-sized hole punch each time, but the height of the spot is not controlled. The viscosity of each person's blood will also be different and therefore assessing whether the method is working correctly can be difficult.

It is difficult to obtain accurate assessment of analyte recovery due to factors such as storage conditions and haematocrit levels, which affects the viscosity of blood with resulting effects on the blood droplet diameter and height. This will in turn affect the distribution of analytes on the paper, consequently affecting the potential analyte recovery. Haematocrit (HCT) level is the volume of red blood cells in the blood, higher HCT level results in a small, dried blood sample and a lower HCT level results in a larger dried blood sample. This is a problem as when a hole punch of the same diameter is used on all samples, there is varying amounts of blood being extracted due to the differences in height caused by the HCT level.

Figure 4.1 shows the difference that the HCT level of a person can cause when spotting blood onto the card.

Figure 4-1. Figure showing the differences in blood spot depending on the HCT levels within a person.

High HCT levels means the blood makes a smaller spot with a larger spot height. Low HCT levels creates a larger spot with a smaller spot height.

These HCT issues are partly why DBS analysis isn't as widely accepted in the clinical world, despite the various attempts to overcome the problems [9]. Some of these attempts include:

1. Using volumetric DBS in combination with whole-spot analysis (e.g., 20 µL spots as well)
2. Pre-assessing the donor's haematocrit level before sampling, using calculations post-analysis to counter for the differences
3. Analysis of the relationship between the HCT level in combination with the spreading of DBS sample area.

The use of an internal standard (ISTD) in the analysis can help this if it is sampled at an early stage of the process. The ISTD can help with assessing the recovery of analyte from DBS and the analysis of the analyte by the LC-MS method. Collection cards that are pre-treated with ISTD to ensure that the blood and ISTD are both subjected to the same extraction effects are available. However, this is not feasible in many different studies with the difficulty of getting the cards pre-treated. It is possible to treat the spots with the ISTD at the solution preparation stage, ready for the elution/extraction. In this case, it is still subjected to the same extraction as the metabolites getting eluted from the DBS [5]. The process of sample punching with either manual or semi-automatic punchers on the card can lead to carryover between samples, therefore it is important to ensure that adequate 'cleaning' is done in between samples. It is suggested that punching two blank cards in between each spot prevents any cross-contamination from occurring [10].

Storage conditions affect the samples, with the use of a desiccator and temperatures part of the questions as to how each one effects the metabolite concentration over time. As well as these practical difficulties, there are also several other difficulties to do with the sample contents. These include the possible differences of the metabolome for people with varying sex, age, body mass index (BMI), diet, metabolism pathways. Two people with the same diseased state will have varying compounds in the blood for many factors such as a different diet, and different time passed between their last food and drink intake. Other factors such as disease states each person may already have can also cause differences, such as someone who has diabetes may have differences in the metabolism pathway as well as differences within the blood. All these factors are difficult to control and therefore, only age and gender matched samples tend to be used, with a large sample set that should hopefully cover a range of different personal states.

In recent years, the use of DBS has significantly increased into new, diverse fields of applications such as therapeutic drug monitoring, forensic toxicology, toxic and pharmacokinetic studies and environmental pollution control [11-12].

4.2.2 Use of DBS Analysis with Mental Health Conditions

According to World Health Organisation, Mood disorders affect approximately 400 million people worldwide with bipolar disorder (BD) and MDD being in the top 20 disorders responsible for the loss of years lived, both being characterised by low energy and mood levels [13]. In the general primary care setting, only half of patients are correctly diagnosed with MDD and therefore there is a need to develop a diagnostic test for more accurate, early diagnosis [14-16]. Individuals suffering with MDD often experience debilitating, recurrent symptoms along with a high level of psychiatric and somatic comorbidities resulting in the decreased quality of life and increased mortality [14-16]. Those suffering often find there are further implications of the disease by experiencing a heavy burden in many other areas of their lives and expanding the direct impact of the disease [17]. It is predicted that by 2030, MDD will be the most debilitating disorder across the world and the largest burden of global disease

with associated costs of approximately £85 billion (US $102.9 billion) [13,18]. MDD currently has no objective tests; it relies on the clinical evaluation of self-reported symptoms due to a limited biological understanding. The diagnosis is outlined in formal classification systems, such as the Diagnostic and Statistical Manual of Mental Disorders, 5th edition (DSM-5) [19-20]. The clinical evaluations given by general practitioners (GPs) are usually time restricted and can be subjective; each GP may diagnose something different and therefore MDD is often over-, under-, or misdiagnosed. Based on a study of over 50,000 people, only 47% of MDD patients are correctly diagnosed by GPs [21]. Patients with BD often are miss-diagnosed with MDD due to the initial depressive episodes that can be mistaken for the depressive episodes associated with MDD [22]. There is often an 8-10 year delay in a BD diagnosis, including the initial symptom manifestation until the psychiatric evaluation, until the correct diagnosis [23]. These current issues also overlap with the reluctance a patient may feel to seek help, the shortage on mental health practitioners and limited focus on mental health in primary care [24]. Due to the wrong diagnosis, patients often do not receive the adequate treatment which is detrimental on the following years of their lives. This outlines the clear need for an accurate and reliable test for early diagnosis of MDD, resulting in more effective treatment and care for the patients.

Despite the various studies into biomarkers, at the current time there are no diagnostic biomarkers currently implemented in routine clinical practices for psychiatric disorders [25]. There is an interest into blood-based biomarkers of MDD due to growing evidence that disease-related alterations can be detected by the peripheral system [19]. Hence, various studies have been conducted into looking at the blood-based biomarkers of such conditions, but these have largely relied on the use of serum/plasma. Dried blood spots offer a novel sampling technique given the multiple advantages when it comes to the implementation of a possible diagnostic test, including the ease of home-testing, which is important when thinking about the context of the mental condition the patients will be subject to. Given the complexity of psychiatric disorders, there is a high threshold for identifying and validating any possible biomarkers and therefore, rather than one, a panel of multiple

biomarkers would be preferable. The changes in the protein concentration of the blood have been investigated however the metabolite changes are yet to be explored but DBS sampling for biomarker discovery holds promise for the development of a rapid, cost-effective diagnostic tests, especially given the challenges for patient recruitment in psychiatric disorders [26-27]. The strengths of mass spectrometry make it an ideal technique for psychiatric diagnosis due to the ability to quantify many analytes, the high specificity, high sensitivity and reproducibility.

4.2.2.1 Introduction into the Project

The samples collected in this study were from a project started by a group at the University of Cambridge, with the intention to develop molecular diagnostics for major neuropsychiatric disorders. The idea is to determine the cellular mechanisms that regulate the expression of possible biomarkers which are altered when in individual is subjected to a depressive or uplifted state, with a further plan to develop novel therapeutic strategies. Over the years, this group has conducted many studies relating to different neuropsychiatric disorders and recently focused on trying to differentiate between major depressive disorder (MDD) and bipolar. They have collected a range of dry blood spots (DBSs) of people either diagnosed with either of the two disorders, or completely healthy. There is an interest is trying to differentiate people with low mood from those with MDD and therefore they performed a proteomics study, combined with a digital mental health assessment, on DBS samples collected. The results were greatly promising, with 5 proteins showing important predictors of MDD along with 4 sociodemographic, clinical and personality characteristics [19]. These characteristics were poor self-rated mental health, high BMI, reduced daily experiences with positive emotions and tender-minded [19]. As proteomics is only based on larger molecules requiring heavy sample preparation and running times, a metabolomics study of the same raw samples to determine whether there is a possible change in the blood when the person is subjected to MDD, is of interest.

The study is part of a larger ‘Delta Study’ launched in 2018 by Cambridge Centre for Neuropsychiatric Research (CCNR) in collaboration with Psyomics Ltd. The overall aim of the study was to develop

tests based upon DBS samples and novel digital mental health assessments, to be used to diagnose individual patients presenting with subclinical low mood with either BD or MDD, with a secondary objective of achieving earlier, and more accurate diagnosis of MDD. A total of 5422 participants were recruited through various means, with strict recruitment inclusion criteria. However, after the psychiatric assessment through digital and telephone interviews and DBS sample collection, whilst fasting at the time of sample collection and presenting with at least a low mood, only 495 participants were eligible for the study. The others were excluded were for various reasons such as wrongly diagnosed, no sample collection provided, not fasting during sample collection and/or not presenting with low mood at the time of sampling. Of these, 232 patients had a previous MDD diagnosis, where-as 263 had no previous MDD diagnosis. A further 174 participants were excluded at this point having no major depressive episodes (MDE) within the past 1-6 months. Consequently, 130 participants were then declared to have subclinical low mood, 40 patients with new current MDD (MDE within the past month), 53 with already established current MDD (MDE within the past month) and 72 with established non-current MDD (MDE within the past 6 months, but not within the past 1 month). Within this study, Mann-Whitney *U* tests were conducted to determine whether the use of antidepressant medication were associated with the proteomic biomarkers determined to be indicative of MDD, but with *p*-values of 0.05, these were shown to not be significantly different. A PCA plot also showed minimal separation between the antidepressant users and non-users, showing that the use of these medications does not affect the results.

It is to be noted that this work with dried blood spots was conducted at the University of Cambridge following the guidance from the Human Tissue Authority. This Delta Study was approved by the University of Cambridge Human Biology Research Ethics Committee, approval number 'HBREC 2017.11'. Hep B immunisation was taken by Keele University Occupational Health Service, with completion by August 2019.

4.3 Materials

Water, methanol (MeOH), acetonitrile (ACN), formic acid and ammonia solution (25% in water) of optima LCMS grade were obtained from Fisher Scientific (Loughborough, U.K.). ESI-L low concentration tuning mix and reference peak markers were purchased from Agilent Technologies (California, USA). The analysis was performed using Agilent Technologies 1290 Infinity II UHPLC system, coupled to an Agilent Technologies 6550 Accurate-Mass Quadrupole-Time-of-Flight mass spectrometer with iFunnel source. The samples were obtained by the group at University of Cambridge.

4.4 Method Development

The dried blood spots were collected by Cambridge Centre for Neuropsychiatric Research (CCNR) by getting a fasted participant to prick their fingertip with a lancet. The first drip is wiped away with cotton ball and the following 3 – 5 drips are collected by the supplied DBS filter paper. These are left to dry and then put into a small pouch with a desiccator for storage till they are ready for analysis.

4.4.1 Sample Preparation and Chromatographic Methods

Different sample preparation methods were trialled to determine which is the most effective at extracting small molecules from the DBS samples. Random volunteer-DBS cards were used for the method development as to not waste the study samples. For metabolomics, a short, simple sample preparation method is required as to not change the metabolome and make the procedure as quick as possible. Therefore, four methods were chosen based on a literature search. The blood samples were prepared in duplicate, and a blank card was also prepared to match each preparation.

- Sample Preparation 1

A 3mm hole punch was used to punch a hole in the blood spot (or an empty blood spot card for the blank) which was placed in the centrifuge tube. To this, 40 μL of LCMS grade water was added and placed on the incubator at 40 °C, whilst shaking at 200 rpm for 10 minutes. Into this tube, 160 μL of 50:50 v/v methanol: acetonitrile was added, and the samples were incubated at 40 °C, whilst shaking

at 200 rpm for 50 minutes. The samples were then centrifuged for 5 minutes at 9800 RCF and transferred to an LC vial for analysis.

- Sample Preparation 2

A 3mm hole punch was used to punch a hole in the blood spot (or an empty blood spot card for the blank) which was placed in the centrifuge tube. To this, 200 μL of 30:30:40 v/v/v methanol: acetonitrile: 0.01% formic acid in water was added, and the samples were incubated at 40 °C, whilst shaking at 200 rpm for 1 hour. The samples were then centrifuged for 5 minutes at 9800 RCF and transferred to an LC vial for analysis.

- Sample Preparation 3

A 3mm hole punch was used to punch a hole in the blood spot (or an empty blood spot card for the blank) which was placed in the centrifuge tube. To this tube, 200 μL of methanol was added, and the samples were incubated at 40 °C, whilst shaking at 200 rpm for 1 hour. The samples were then centrifuged for 5 minutes at 9800 RCF and transferred to an LC vial for analysis.

- Sample Preparation 4

A 3mm hole punch was used to punch a hole in the blood spot (or an empty blood spot card for the blank) which was placed in the centrifuge tube. 200 μL of LCMS grade acetonitrile was added, and the samples were incubated at 40 °C, whilst shaking at 200 rpm for 1 hour. The samples were then centrifuged for 5 minutes at 9800 RCF and transferred to an LC vial for analysis.

The samples were analysed using Agilent 1290 Infinity II UHPLC system, coupled to a 6550 Q-ToF with iFunnel source. The column used was an Agilent Eclipse Plus C18 (2.1 x 100mm, 1.8 μM particle size) as a starting point for reversed-phase analysis. The mobile phases were chosen to keep it a simple procedure, as they are the solvents that the sample is dissolved in. The formic acid is added to aid ionisation of molecules in the mass spectrometer.

Table 4.1 shows the LC-MS method used for the initial analysis given the shallow gradient, which is repeated in positive and negative ionisation mode.

Table 4.1. RP-LC-MS Method, with gradient for development

HPLC-MS					
Instrument	Agilent 1290 Infinity II UHPLC + 6550 Q-ToF with iFunnel				
Column	Agilent Eclipse Plus C18 RRHD. 2.1 x 100mm, 1.8 µm				
Oven (ºC)	40°C				
Pump	Mobile Phase A	0.1% Formic Acid in LCMS Water			
	Mobile Phase B	0.1% Formic Acid in 50/50 LCMS Acetonitrile/Methanol			
	Flow Rate	0.4 mL/min			
	Gradient	Time / (min)	%A	%B	
		1.0	95	5	
		11.0	5	95	
		11.1	2	98	
		12.0	2	98	
		12.1	95	5	
		15.0	95	5	
	Runtime (min)	15			
Injector	Volume (µL)	2			
MS	QTOF/QQQ/TOF	Mass Spec Type:	QTOF	Mode	+ve & -ve
	Source –Dual Jet Stream ESI	Gas temp	200°C	Gas flow	14 L/min
		Sheath Gas temp	350°C	Sheath gas flow	11 L/min
		Nebuliser pressure	241 kPa	Nozzle Voltage	100 V
		VCap	3500 V	Fragmentor	380 V

Table 4.2 and **Table 4.3** shows two methods for HILIC analysis which are to be compared and optimised for these DBS samples. The gradient method was optimised though the most advantageous column and mobile phase combination are to be determined. The first method uses an Agilent HILIC-Z column with ammonia used as an additive in the mobile phases.

Table 4.2 showing the HILIC method 2:

Table 4.2. HILIC_LC-MS Method 1.

The gradient is for development and column/mobile phases for comparison

HPLC-MS _ HILIC 1					
Instrument	Agilent 1290 Infinity II UHPLC + 6550 Q-ToF with iFunnel				
Column	Agilent InfinityLab Poroshell 120 HILIC-Z. 2.1 x 150mm, 2.7 μm				
Oven (ºC)	40°C				
Pump	Mobile Phase A	0.3% NH3 in LCMS Optima Water			
	Mobile Phase B	0.3% NH3 in LCMS Optima Acetonitrile			
	Flow Rate	0.4 mL/min			
	Gradient	Time / (min)	%A	%B	
		1.5	15	85	
		14.0	20	60	
		16.0	40	60	
		16.1	15	85	
		20.0	15	85	
	Runtime (min)	20			
Injector	Volume (μL)	2			
MS	QTOF/QQQ/TOF	Mass Spec Type:	QTOF	Mode	+ve (-ve)
	Source –Dual Jet Stream ESI	Gas temp	200°C	Gas flow	14 L/min
		Sheath Gas temp	350°C	Sheath gas flow	11 L/min
		Nebuliser pressure	241 kPa	Nozzle Voltage	100 V (1000 V)
		VCap	3500 V	Fragmentor	380 V

The second method for comparison with the HILIC method in **Table 4.2** is seen in **Table 4.3** where the column and mobile phases are changed. The column for comparison is Waters XBridge Amide column, used with ammonium formate as a buffer in the mobile phases.

Table 4.3 showing the HILIC method 2:

Table 4.3. HILIC_LC-MS Method 2.

The gradient is for development and column/mobile phases for comparison

HPLC-MS _ HILIC 2					
Instrument	Agilent 1290 Infinity II UHPLC + 6550 Q-ToF with iFunnel				
Column	Waters XBridge Amide 3.5µm, 2.1 x 150 mm				
Oven (ºC)	35°C				
Pump	Mobile Phase A	10mM Ammonium Formate at pH 3 in 90:10 Water:MeCN			
	Mobile Phase B	10mM Ammonium Formate at pH 3 in 90:10 MeCN:Water			
	Flow Rate	0.4 mL/min			
	Gradient	Time / (min)	%A	%B	
		1.5	0	100	
		10.0	40	60	
		12.0	40	60	
		12.1	0	100	
		18.0	0	100	
	Runtime (min)	20			
Injector	Volume (µL)	2			
MS	QTOF/QQQ/TOF	Mass Spec Type:	QTOF	Mode	+ve (-ve)
	Source –Dual Jet Stream ESI	Gas temp	225°C	Gas flow	11 L/min
		Sheath Gas temp	350°C	Sheath gas flow	10 L/min
		Nebuliser pressure	241 kPa	Nozzle Voltage	100 V (1000 V)
		VCap	3000 V	Fragmentor	380 V

4.4.2 Sample Repeatability

During this Experiment, the procedure followed is one that the group adhere to where Volunteer DBSs are used as QC samples. This is due to the low volume of actual DBS samples available; a pooled QC sample could not be made accurately and efficiently. Volunteer DBS (VDBS) cards are sample cards that have been made by a volunteer in the lab, this can be anyone involved the group at Cambridge and the identity of the volunteer is kept anonymous on the card. These are usually sampled, left to dry for 24 hours then stored at RT in a small bag with desiccators in, away from light and heat. For this short Experiment, VDBS cards were used to test the repeatability of the instrument and the sample preparation. This Experiment used 3 different dried blood spots for the extraction procedure. Alongside these samples, 4 standards were also used for compounds that are assumed to be in the blood; these are tryptophan, theobromine, theophylline and caffeine but these were only used as peak markers, therefore the concentration of the standard is unknown. The sample

preparation was kept the same for each of them, using the same sample preparation and LC methods chosen in **Section 4.4.1.**

Inter-sample Repeatability

Each dried blood spot card has space for 5 spots from the same one person, therefore inter-sample repeatability can be examined. Peak markers from the 4 chosen standards were run on the method to accurately extract the compound from the blood spot. This ensures that compounds are extracted from the spot, as well as determining how repeatable the chosen workflow is. Using the sample preparation chosen from **Section 4.4.1**, 3 samples were extracted and run on the same LC method stated in **Table 4.1**.

Intra-sample Repeatability

The same sample vial was injected 6 times to determine whether the instrument and method were working well. This intra-sample variance needs to be kept to a minimum given the larger inter-sample variance. If this was kept to the minimum, then this allows for more variation within the sample preparation.

4.4.3 Sample Stability

The final analysis would take multiple days to complete and therefore would be in the autosampler for multiple days at room temperature. Therefore, the stability of the samples was assessed. Randomly selected samples were run four days apart to see if there are any changes over time. These samples were run on the LC method in **Table 4.1** and were samples prepared as best determined in **Section 4.4.1**.

4.5 Results

4.5.1 Sample Preparation and Chromatographic Methods

Since the components of the blood were unknown, it was assumed that the "best" sample preparation method and LC-MS method was the one that found the greatest number of features when using Agilent's 'Find by Molecular Feature' in 'Agilent MassHunter Qualitative Analysis' software. For consistency, the blank card samples were also analysed using this feature extraction method for each sample preparation to ensure that the features were genuinely extracted from the blood spot, and not from the card matrix. The molecular feature extraction (MFE) results are given in the **Table 4.4** and **Table 4.5** for positive and negative ionisation mode, respectively. The overall number represents the average of A and B repeats, minus the number of features found in the blank card.

Table 4.4. Molecular Feature Extraction for Positive Ionisation Mode showing the number of features found in each sample and an average number of features found, minus the card blank

		No. of Features	Average No. of Features (minus Card Blank)
Prep Method 1	**A**	3578	385.5
	B	3609	
	Card Blank	3208	
Prep Method 2	**A**	3498	278.0
	B	3566	
	Card Blank	3254	
Prep Method 3	**A**	3188	305.5
	B	3133	
	Card Blank	2855	
Prep Method 4	**A**	3045	355.5
	B	3164	
	Card Blank	2749	

Table 4.5. Molecular Feature Extraction for Negative Ionisation Mode showing the number of features found in each sample and an average number of features found, minus the card blank

		No. of Features	Average No. of Features (minus Card Blank)
Prep Method 1	**A**	2393	844.0
	B	2533	
	Card Blank	1619	
Prep Method 2	**A**	1996	255.0
	B	1882	
	Card Blank	1684	
Prep Method 3	**A**	1946	651.0
	B	2112	
	Card Blank	1378	
Prep Method 4	**A**	2090	718.5
	B	2229	
	Card Blank	1441	

In the tables, both positive and negative ionisation mode showed that the method that gave the greatest number of features consistently is sample preparation method 1. This is thought to be due to the wetting of the DBS card before using the extraction solvents.

The HILIC method development compared the two different column phases and mobile phases. The aim was to determine the method that gave the best separation and the greatest number of features. In positive ionisation mode, method HILIC 1 in **Table 4.2** showed overall 1681 features compared to only 1134 features extracted using method HILIC 2 (**Table 4.3**). Therefore, the method 1 was determined to be most appropriate going forward.

Figure 4.2 shows the compounds extracted using HILIC method 1:

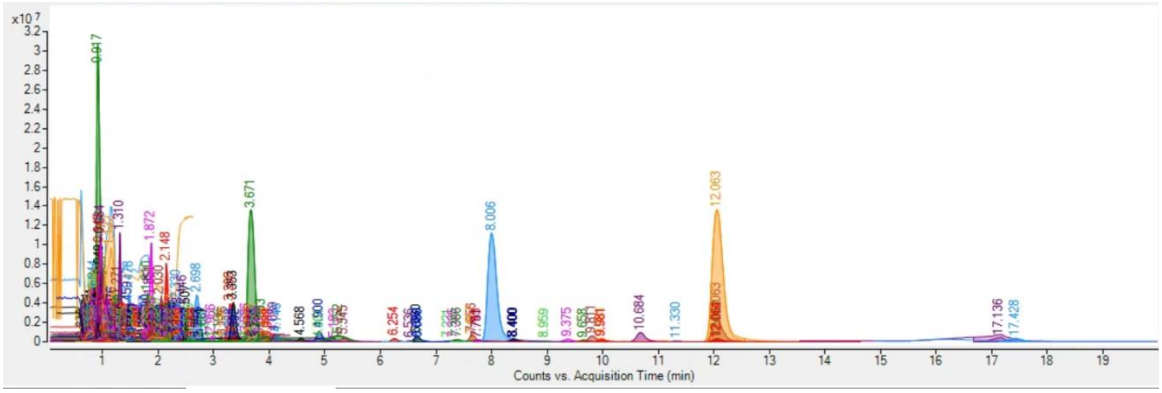

Figure 4-2. All 1681 compounds extracted using HILIC method 1

The compounds are spread across the whole run time, with the last compound eluting at 17 minutes where the wash period is. There are 1681 compounds extracted from sample 1 on this method.

Figure 4.3 shows the compounds extracted using HILIC method 2:

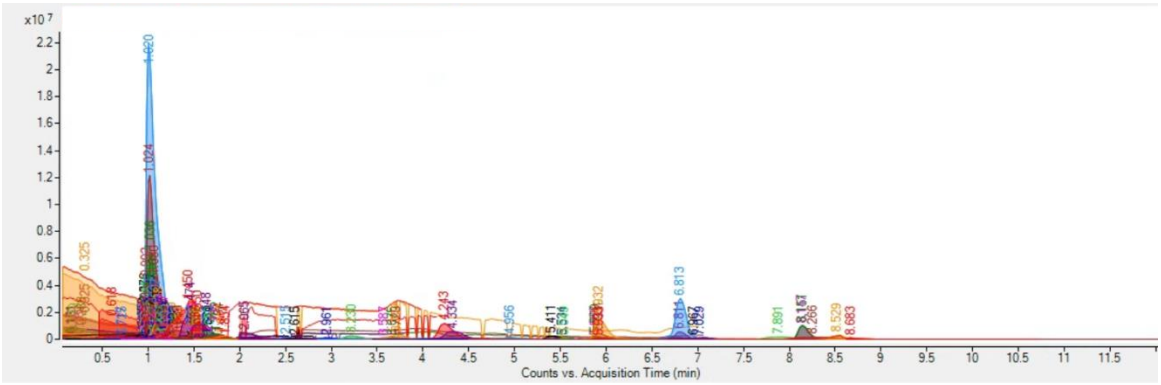

Figure 4-3. All 1134 compounds extracted using HILIC method 2

The compounds are not spread across the whole run time, they are only within the first 9 minutes of a 20 minute run analysis time. There are no significant peaks except those at 1 minute and therefore shows this method is not ideal.

The comparison of the extracted compounds was repeated with other samples, which all showed that the HILIC method 1 extracted a greater number of features with a better spread across the whole analysis time. Therefore, it was decided that HILIC method 1 is the best method.

4.5.1.1 Summary of Method Decisions

- The sample preparation method conducted was as follows:
  A 3mm hole punch was used to punch a hole in the blood spot (or an empty blood spot card for the blank) which was placed in the centrifuge tube. To this, 40 μL of LCMS grade water was added to the sample and placed on the incubator at 40 °C, whilst shaking at 200 rpm for 10 minutes. 160 μL of 50:50 v/v methanol: acetonitrile was added, and the samples were incubated at 40 °C, whilst shaking at 200 rpm for 50 minutes. The samples were then centrifuged for 5 minutes at 9800 RCF and transferred to an LC vial for analysis.
- The reversed phase liquid chromatography (RP-LC) method used is shown in **Table 4.6**. There is a slight development from the previous method to create an even shallower gradient than previously shown.

Table 4.6. RP-LC Method to use

HPLC-MS					
Instrument	Agilent 1290 Infinity II UHPLC + 6550 Q-ToF with iFunnel				
Column	Agilent Eclipse Plus C18 RRHD. 2.1 x 100mm, 1.8 μm				
Oven (ºC)	40°C				
Pump	Mobile Phase A	0.1% Formic Acid in LCMS Water			
	Mobile Phase B	0.1% Formic Acid in 50/50 LCMS Acetonitrile/Methanol			
	Flow Rate	0.4 mL/min			
	Gradient	Time / (min)	%A	%B	
		1.0	98	2	
		11.1	2	98	
		12.0	2	98	
		12.1	98	2	
		15.0	98	2	
	Runtime (min)	15			
Injector	Volume (μl)	2			
MS	QTOF/QQQ/TOF	Mass Spec Type:	QTOF	Mode	+ve (-ve)
	Source –Dual Jet Stream ESI	Gas temp	200°C	Gas flow	14 l/min
		Sheath Gas temp	350°C	Sheath gas flow	11 l/min
		Nebuliser pressure	241 kPa	Nozzle Voltage	100 V (1000 V)
		VCap	3500 V	Fragmentor	380 V

- The HILIC method to be used, which should show the compounds that are not eluting in the main analysis region on the R*P*-LC method, is shown in **Table 4.7**.

Table 4.7. HILIC Method to use

HPLC-MS					
Instrument	Agilent 1290 Infinity II UHPLC + 6550 Q-ToF with iFunnel				
Column	Agilent InfinityLab Poroshell 120 HILIC-Z. 2.1 x 150mm, 2.7 µm				
Oven (ºC)	40°C				
Pump	Mobile Phase A	0.3% NH3 in LCMS Optima Water			
	Mobile Phase B	0.3% NH3 in LCMS Optima Acetonitrile			
	Flow Rate	0.4 mL/min			
	Gradient	Time / (min)	%A	%B	
		1.5	15	85	
		14.0	20	60	
		16.0	40	60	
		16.1	15	85	
		20.0	15	85	
	Runtime (min)	20			
Injector	Volume (µL)	2			
MS	QTOF/QQQ/TOF	Mass Spec Type:	QTOF	Mode	+ve (-ve)
	Source –Dual Jet Stream ESI	Gas temp	200°C	Gas flow	14 L/min
		Sheath Gas temp	350°C	Sheath gas flow	11 L/min
		Nebuliser pressure	241 kPa	Nozzle Voltage	100 V (1000 V)

4.5.2 Sample Repeatability

Inter-sample Repeatability

Figure 4.4 shows the overlaid TIC of three different samples from one person.

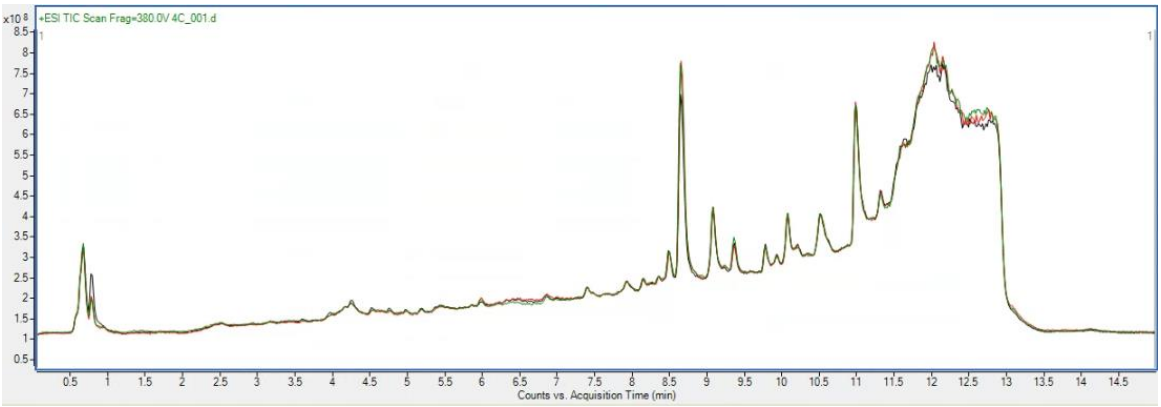

Figure 4-4. Overlay TIC of three different samples from one person.

TIC traces overlay perfectly with no additional or absent peaks therefore repeatable preparation.

The trace looks repeatable and therefore each of the masses were extracted in the samples to determine the area of each peak. **Table 4.8** shows the peak areas, with the mean, standard deviation (SD) and the percentage relative standard deviation (%RSD).

Table 4.8. Table showing peak areas throughout three repeat samples.

The table includes the mean, standard deviation and %RSD of peak areas of theobromine, theophylline, caffeine and tryptophan.

	Theobromine Area	Theophylline Area	Caffeine Area	Tryptophan Area
A	1738884	889906	2848514	340245
B	1437570	788448	2486777	281758
C	1512551	851487	2780422	295411
Mean	1563002	843280	2705238	305805
SD	156864	51224	192232	30597
RSD%	10.0	6.10	7.10	10.0

At a parts per billion (ppb) level, an acceptable %RSD is approximately 10% or below, though FDA guidance shows accepted RSD up to 30% depending on sample matrix and analyte chemistry, in metabolomic analysis. The FDA also states that ±15% RSD is good for targeted analysis, though the target for repeatable analysis is below 5. It is assumed that these compounds are present in ppb levels and therefore each %RSD is accepted, although they are on the upper limits. This variance can be expected from 3 separate preparations from 3 spots on one card for compounds at such low levels on a difficult sample matrix.

Intra-sample Repeatability

To test the repeatability of the injections, one sample was injected multiple times. **Figure 4.5** shows a TIC overlay of 6 repeated injections from one sample.

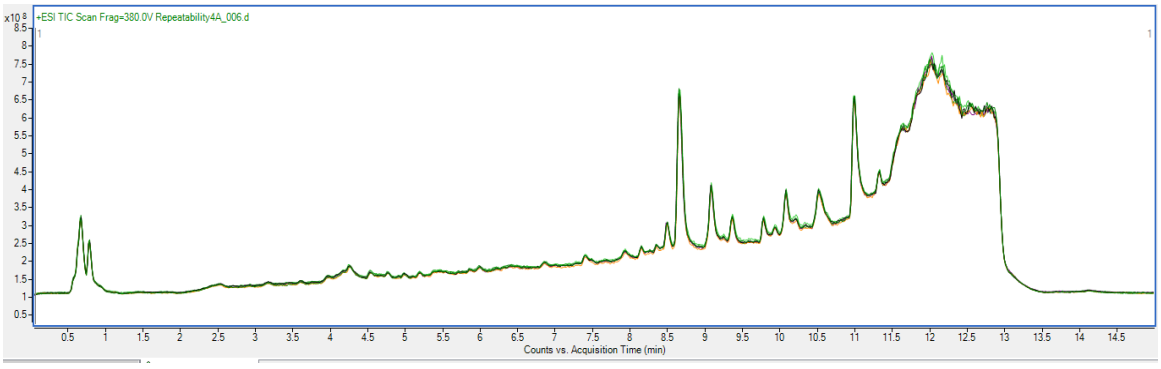

Figure 4-5 TIC overlay of 6 repeated injections.

TIC overlay perfectly, with no additional or absent peaks therefore repeatable injections.

The traces were repeatable, and no differences can be seen suggesting the method was working well. The same 4 masses from the inter-sample repeatability Experiment were extracted to determine if these compounds are repeatable across the 6 injections. The abundance of each

compound across the six repeated injections in tabulated in **Table 4.9**. Again, the mean, SD and %RSD can also be found in this table for each compound. The %RSD should be below 2% for a highly reproducible method.

Table 4.9. 6 Repeat injections showing the mean, SD and %RSD also of peak areas of theobromine, theophylline, caffeine and tryptophan.

	Theobromine Area	Theophylline Area	Caffeine Area	Tryptophan Area
1	1764945	893474	3297069	374251
2	1728562	889798	3177846	403920
3	1728664	900283	3144723	363836
4	1718662	913916	3217220	359664
5	1761691	905124	3204252	376069
6	1733707	870933	3244004	364843
Mean	1739372	895588	3214186	373764
SD	19207	14797	52938	16074
RSD%	1.10	1.65	1.65	4.30

Tryptophan has the largest RSD however it is still below 5% which is the 'good' acceptance for ppb level compounds. Three of the four compounds have a %RSD of less than 2 which shows the instrument and LC method are highly reproducible and reliable.

4.5.3 Sample Stability

To check the sample stability over four days, randomly selected samples were run and the TIC was overlaid to check for any changes over time. **Figure 4.6** shows the TIC overlay of the samples, the red trace being the original and the black trace showing the same sample four days later.

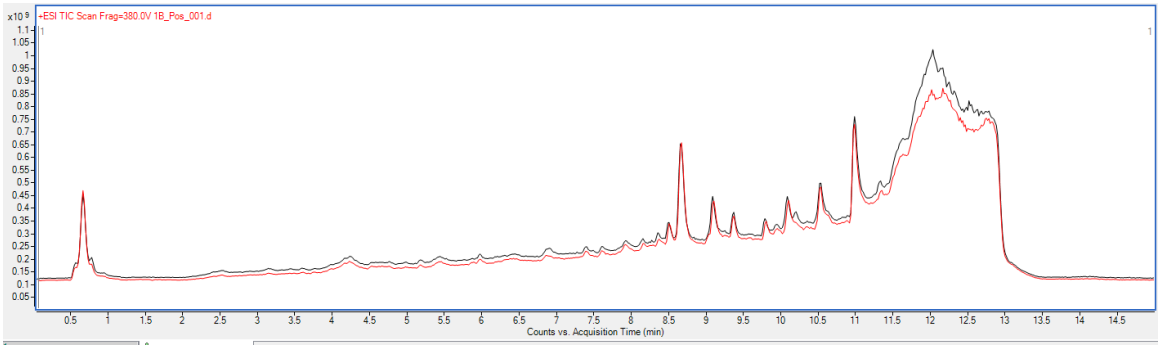

Figure 4-6. TIC Overlay on sample stability.

No additional or absent peaks therefore passes sample stability on TIC eye level.

There are very little differences in the TIC traces, except for the baseline noise level and the wash section at the end. This shows some difference and therefore reiterates the need for a QC sample to be used throughout the run to check for any sample instability, as well as instrumental drift.

4.5.4 Overview of Samples from the Cambridge Group Results

A set of 106 dried blood spot samples were received anonymously from the clinic run by Professor Sabine Bahn. These include 50 samples from patients diagnosed with Major Depressive Disorder (MDD), 50 age, sex and BMI matched control samples and 6 volunteer DBS (VDBS) samples as QC controls. The samples were prepared as tested and described in **Section 4.5.1** and were run of the Q-ToF in positive and negative ionisation mode on RP-LC method in **Table 4.6** and the HILIC method described in **Table 4.7**.

4.5.4.1 Traditional Analysis and QC Check

One QC sample (VDBS) was injected multiple times throughout the analysis to monitor the method and instrument, enabling to see if there's instrumental drift and check that it was working consistently throughout the analysis.

Figure 4.7 shows a QC Check on RP-LC Positive Mode:

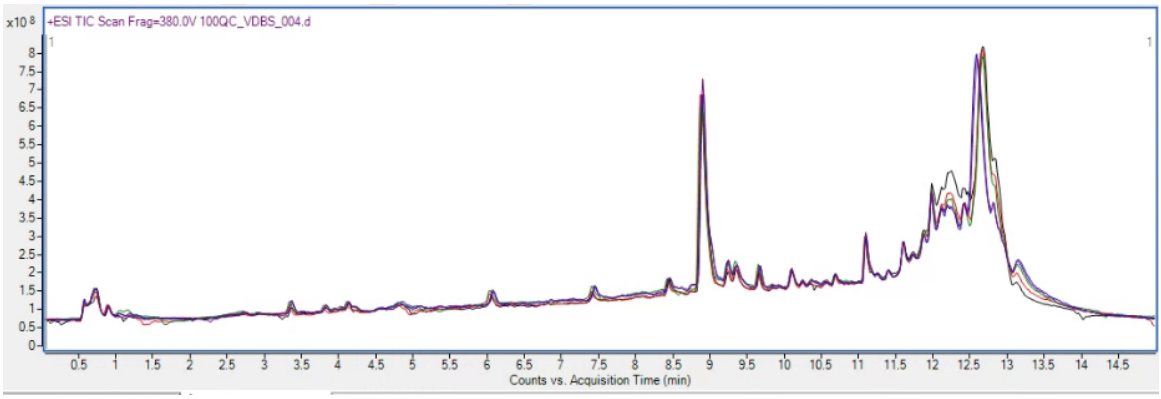

Figure 4-7. TIC overlay of RP-LC Positive Mode QC samples.

No additional or absent peaks therefore QC passes tests by traditional analysis on TIC.

The slightly elevated background is the first QC sample run pre analysis. There are no additional or missing peaks and the samples chromatograms are overlaid, showing that the instrument and method is working sufficiently.

Figure 4.8 shows a QC Check on RP-LC Negative Mode:

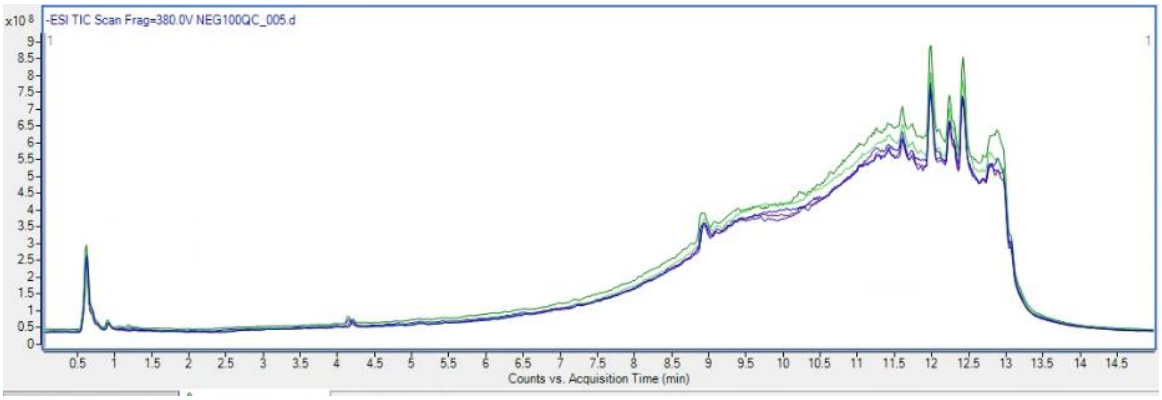

Figure 4-8. TIC overlay of RP-LC Negative Mode QC samples.

No additional or absent peaks therefore OK.

The trace showing an elevated background throughout, particularly in the wash region, is the first QC sample run in negative mode. There are no additional or missing peaks and the samples chromatograms are overlaid, showing that the instrument and method is working sufficiently.

Figure 4.9 shows a QC Check on HILIC Positive Mode:

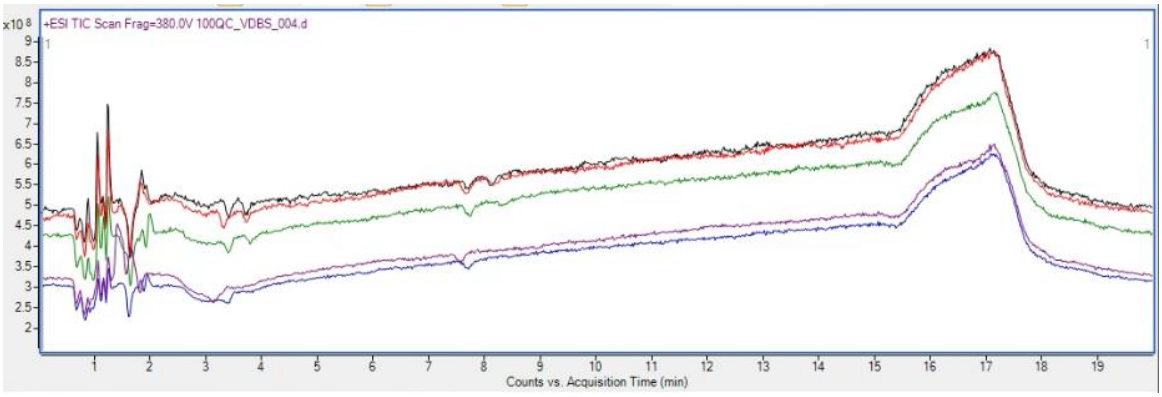

Figure 4-9. TIC overlay of HILIC Positive Mode QC samples.

No additional or absent peaks therefore OK.

The trace shows that the background throughout the run decreases. This could be due to multiple reasons, including a long equilibrium of the column in HILIC mode over time. There are over 100 samples run between the first and last QC sample from **Figure 4.9** meaning that there is over 30 hours in between these samples Despite this uncertainty, there are no additional peaks in any samples, and the large peak in the wash region at the end, and the retention time of the solvent front peaks at the beginning are overlaid, despite the difference in background. There are no peaks in the analysis to check any retention time drift other than these. The lack of peaks shows the need for software that can extract any possible compounds in the data.

QC Check on HILIC Negative Mode:

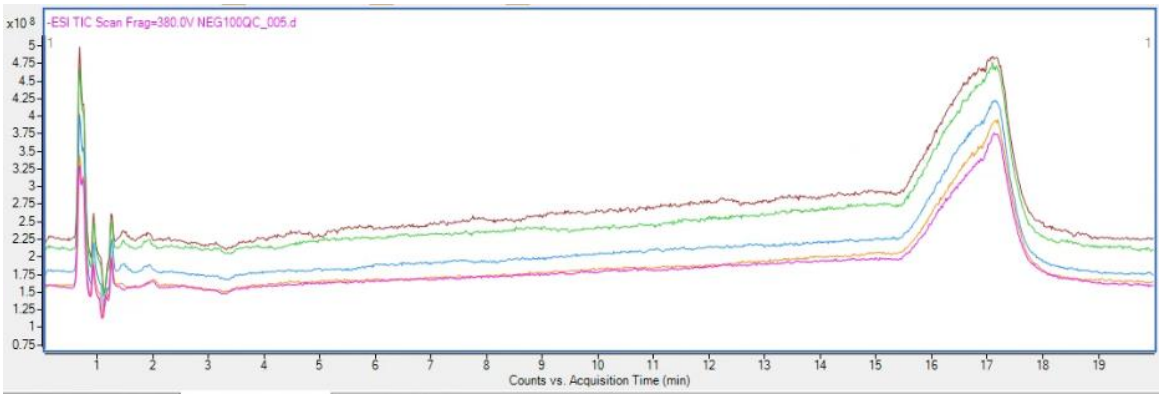

Figure 4-10. TIC overlay of HILIC Negative Mode QC samples.

No additional or absent peaks therefore OK.

The trace in **Figure 4.10** shows that the background throughout the run decreases again. However, there are no additional peaks in any samples, and the large peak in the wash region at the end, and the solvent front peaks at the beginning are overlaid, despite the difference in background. The lack of peaks shows the need for software that can extract any possible compounds in the data.

4.5.4.2 Mass Profinder

Whether the difference between control and MDD samples are significant was determined on the fold change between the two. If the mass is significant, the fold change will be greater than 30. Any fold change above the value of 10 was investigated.

RP-LC Positive Mode

Using Mass Profinder, differences between the two sample groups were investigated. As the intention was to discover a biomarker that is present in MDD samples and absent from control group samples, the results will look at the average abundance across the groups and discover is any of them have an abundance close to 0. **Table 4.10** shows the average abundance of the masses that are less than 10,000.

Table 4.10. Table of average abundance in Profinder RP-LC positive mode. Only one shows a fold change above 2 with a mass 662.4479 Da. The crossed-out features show those with a fold change less than 2 as these are insignificant for the data analysis.

Mass (Da)	RT (min)	Average Abundance		Fold Change
		Control	MDD	
~~188.0393~~	~~2.660~~	~~3075~~	~~4889~~	1.59
~~127.0985~~	~~0.890~~	~~1600~~	~~2277~~	1.42
~~127.0993~~	~~1.250~~	~~648~~	~~191~~	0.29
~~375.2525~~	~~9.010~~	~~329~~	~~1~~	0.00
662.4479	12.240	0	3608690	3608690.00
~~420.3099~~	~~9.000~~	~~0~~	~~1~~	1.00

It can be seen in the table that most of the masses are not significantly different between the two sample groups, hence they have a strike through. There is one mass of 662.4479 at 12.240 min that seems to have a significant difference, however when this mass is reviewed in Profinder, it does not appear to be a real compound and only a shoulder of the peak at 12.6 min with a neutral mass of 662.4476 Da. These peaks can be seen in **Figure 4.11** where the left-hand side shows the peak at 12.24 min, though only two of these have an extracted peak and there is the trace of a larger peak that is not integrated. On the right-hand side, there is a peak extracted that is consistent throughout and seems to be the peak that should've been extracted originally.

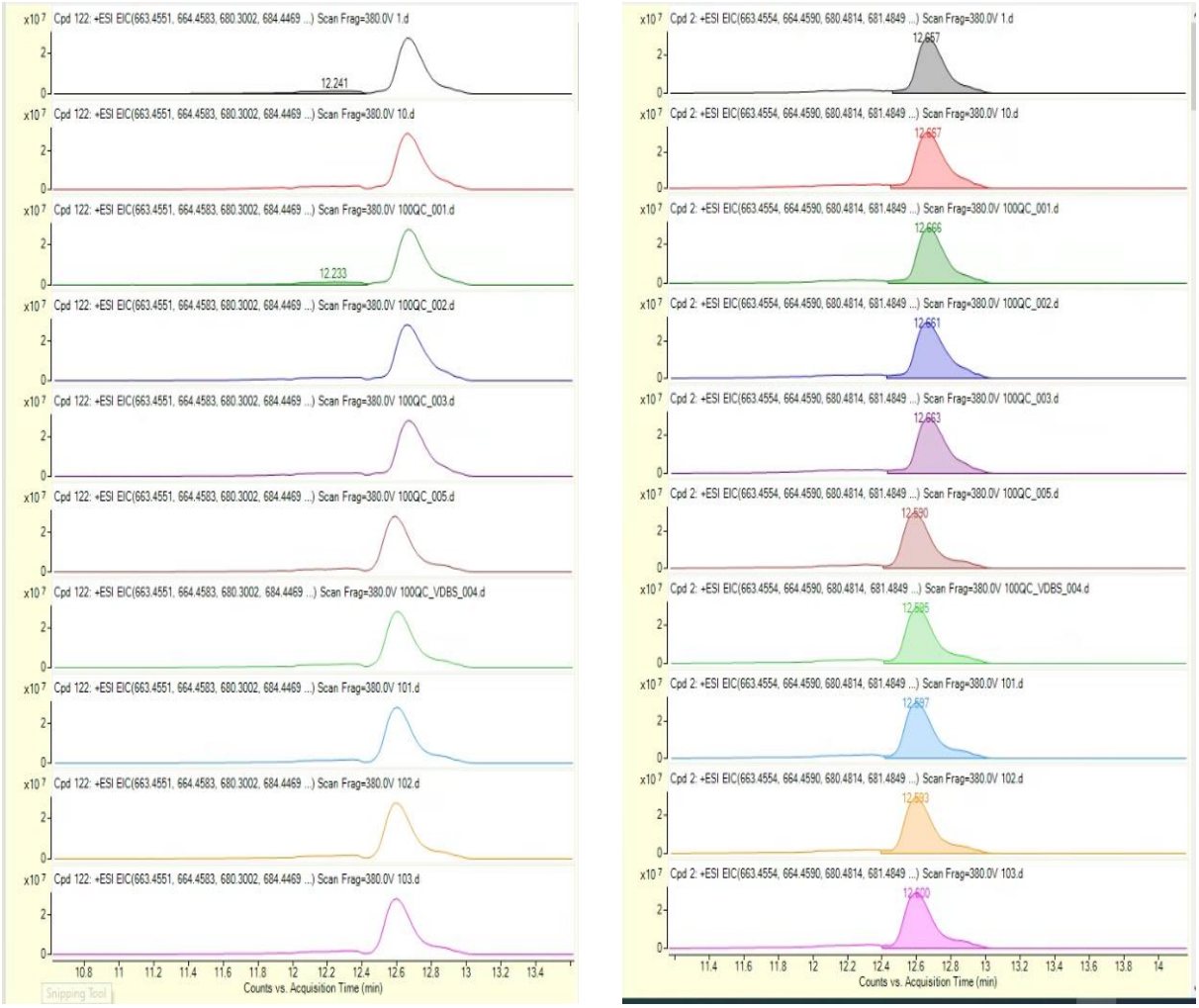

Figure 4-11. Peak with neutral mass of 662.447 Da in Profinder positive mode.

Shows the peak on the left wrongly extracted at 12.3 minutes, where-as it should be the peak extracted on the right at 12.6 minutes. Therefore, this is a false positive and is the same abundance throughout all samples.

RP-LC Negative Mode

As previous, compounds that are potentially biomarkers have been searched in Profinder and **Table 4.11** shows the lowest average abundance of the masses that are less than 10,000.

Table 4.11. Table of average abundance in Profinder RP-LC negative mode. Only one mass at 327 Da has a fold change above 2. The crossed-out features are those that have a fold change less than 2.

Mass (Da)	RT (min)	Average Abundance		Fold Change
		Control	MDD	
327.0958	2.900	3650	47874	13.12
~~375.2507~~	~~9.040~~	~~568~~	~~310~~	0.55
~~347.3022~~	~~9.350~~	~~1~~	~~1~~	1.00

The compound average masses in **Table 4.11** are not significantly different, except for 327.0958 at 2.9 minutes. This mass shown in Profinder averages, however, does not cooperate what the raw data

shows. **Figure 4.12** shows some selected samples in Profinder where there is a peak in 2 MDD samples, but there is also a peak in a control sample which shouldn't be present at such a high abundance according to the Profinder data. There is a peak missing from 5 MDD samples showing it is not an ideal biomarker.

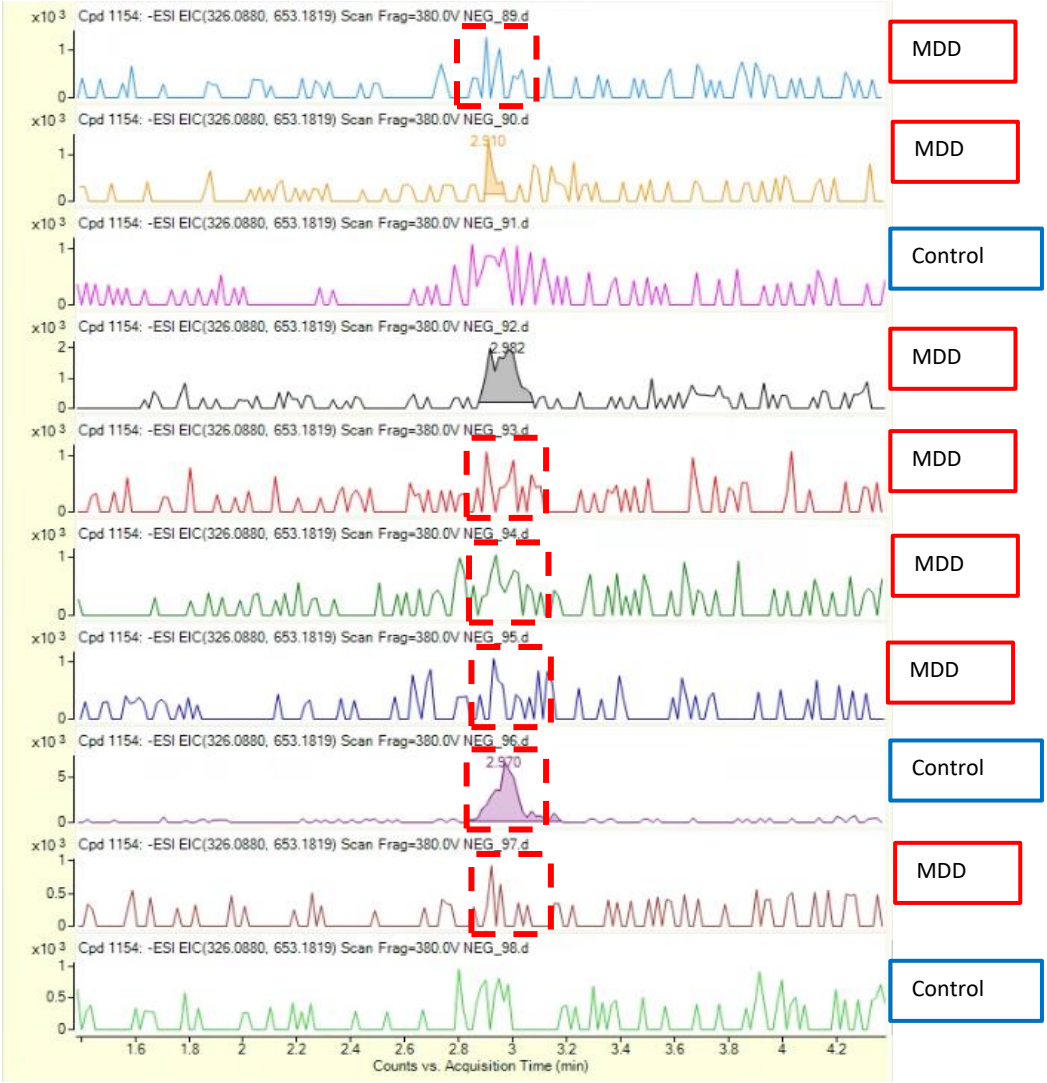

Figure 4-12. Profinder of peak with neutral mass 327.0958.

The red dashed lines show the wrongly absent/present peaks and therefore showing that it is a random mass that does not show a trend of present in only MDD group. This is a false positive.

There are no other masses that show a significant difference between the two groups that also check out in the raw data when checked. The compound with neutral mass 823.7592 Da at 12.92 minutes does not appear in this data.

The analysis conducted in HILIC mode also showed no differences between control and disease data.

4.5.4.3 MPP

Any differences between the control and MDD group should be visible on a PCA plot. They would cluster into distinct groups or show some trend within the data.

RP-LC Positive Mode

PCA Plots

Figure 4.13 shows the PCA plot for the RP-LC positive ionisation mode data.

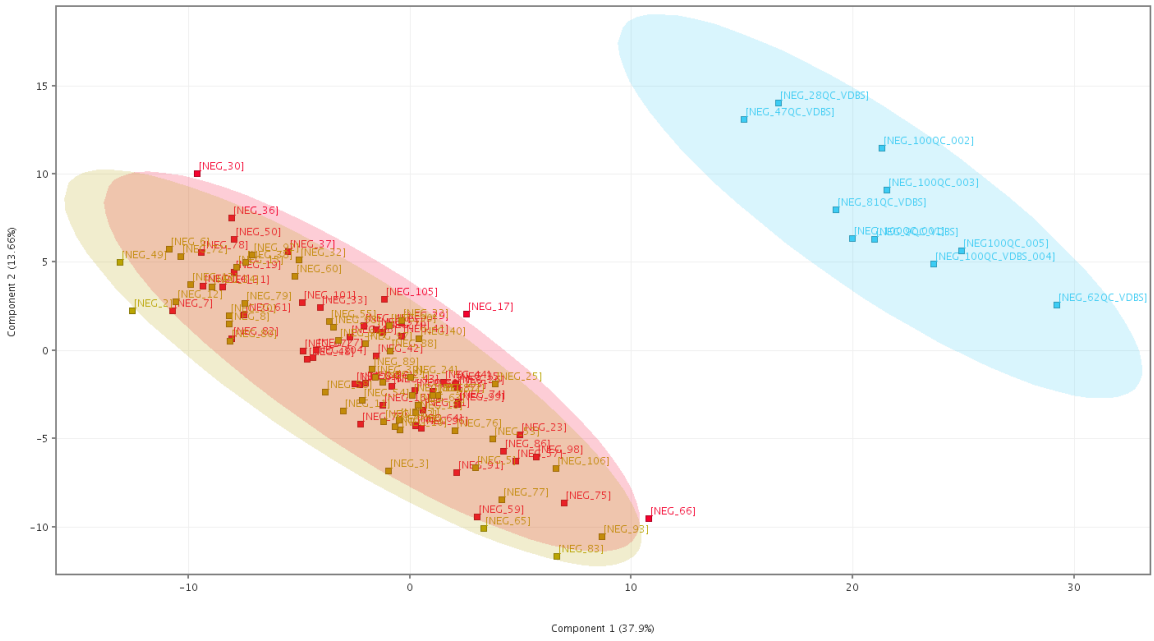

Figure 4-13. 2D PCA plot on RP-LC Positive Mode.

The PCA plot shows the lack of clustering in MDD and control samples. The QC VDBS samples are clustered into one group (blue), then MDD and control samples are clustered into one group (red & yellow). The largest variance is shown on the x-axis with 37.90% and the next largest variance on the second component is 13.66% (y-axis).

The PCA plot in **Figures 4.13** shows that the biggest variance in the data is the difference between the QC samples (VDBS) and the rest of the samples. There is no difference between the MDD and control sample groups in this PCA plot, they are all mixed with some slight spread along the 3 axes. They are very tightly clustered, but sit far away from the QC samples, spread along the x-axis with a variance of only 38%. The sample '100QC' was injected 5 times and these are clustered together, with some slight spread along the x- and y-axes seen in the PCA plot. The other QC samples spread along the y-axis mainly, with only 14% variance from the extremes, showing that the instrument is working OK. These QC samples act lot like quality assurance (QA) samples. The purpose of these is to

check whether there are any plate effects, for example if the samples all go in one line with the QC's in numerical order of injection, then this shows that there is potentially a problem with sample stability or a problem with the sample injector, causing plate effects within the autosampler positions.

Unique Compounds

Looking at the Venn diagrams of unique compounds to each group, there are multiple masses that are incorrectly labelled as unique. These correlate to the Profinder results, therefore showing that as the Profinder results are not entirely accurate, the MPP results are also not entirely accurate. There are some masses that show some differences but none of these are absent in only one sample group, and none are showing a significant difference. No compounds with a *p*-value of less than, or equal to, 0.01 and a fold change of greater than 30, between the MDD and control groups.

RP-LC Negative Mode

Figures 4.14 shows the PCA plots for the RP-LC negative ionisation mode data.

PCA Plot

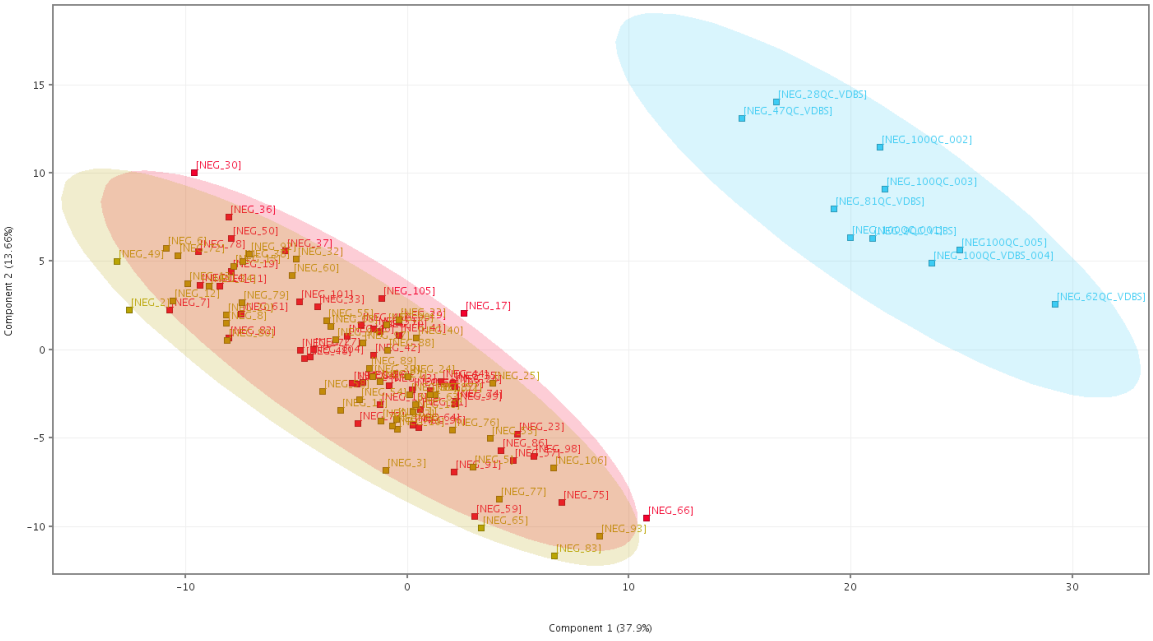

Figure 4-14. PCA plot on RP-LC Negative Mode.

The PCA plot shows the lack of clustering in MDD and control samples. The QC VDBS samples are clustered into one group (blue), then MDD and control samples are clustered into one group (red & yellow).

The PCA plot in **Figures 4.14** shows the same that the positive ionisation mode shows, which is that the biggest variance shown in the data is the difference between the QC samples (VDBS) and the rest of the samples. There is no difference between the MDD and control sample groups in this PCA plot, they are all mixed with some slight spread along the 3 axes. They are very similarly clustered as they were in the positive mode since the samples sit far away from the QC samples by the spread along the x-axis showing a 38% variance. The QC samples are grouped together, with some spread along the three dimensions, however the 5 repeat injections are grouped closer together than the rest of the QC samples.

Unique Compounds

Along with the positive ionisation mode, there are multiple masses that are incorrectly assigned as unique to one group in the Venn diagrams within MPP. The masses have a difference up to 30% between the two groups overall on average however this is not significant. There are no compounds with a *p*-value of less than, or equal to, 0.01 and a fold change of greater than 30, between the MDD and control groups.

4.5.4.4 Mass Profiler

RP-LC Positive Mode

Using Mass Profiler, the method was to look for the compounds with a fold change larger than 30 between the MDD and control groups. However, there were no compounds with a FC above 10. When the method was repeated with FC above 5, one compound showed up, which is shown in **Table 4.12**.

Table 4.12. Table of fold change masses in Profiler RP-LC positive mode

RT (min)	Mass (Da)	Peak Abundance		Log2(A1/A2)	FC
		MDD	Control		
12.300	564.6546	115652	15858	2.87	7.31

However, when this mass is extracted in the raw data the difference is due to the poor peak shape causing the integration not to be consistent. Some samples it is shown to be small, other times it is a

wide 'peak'. This is shown in **Figure 4.15** below on the right-hand side. The left-hand side of **Figure 4.15** shows the different samples with very similar abundances and areas, despite the MDD and control sample groups.

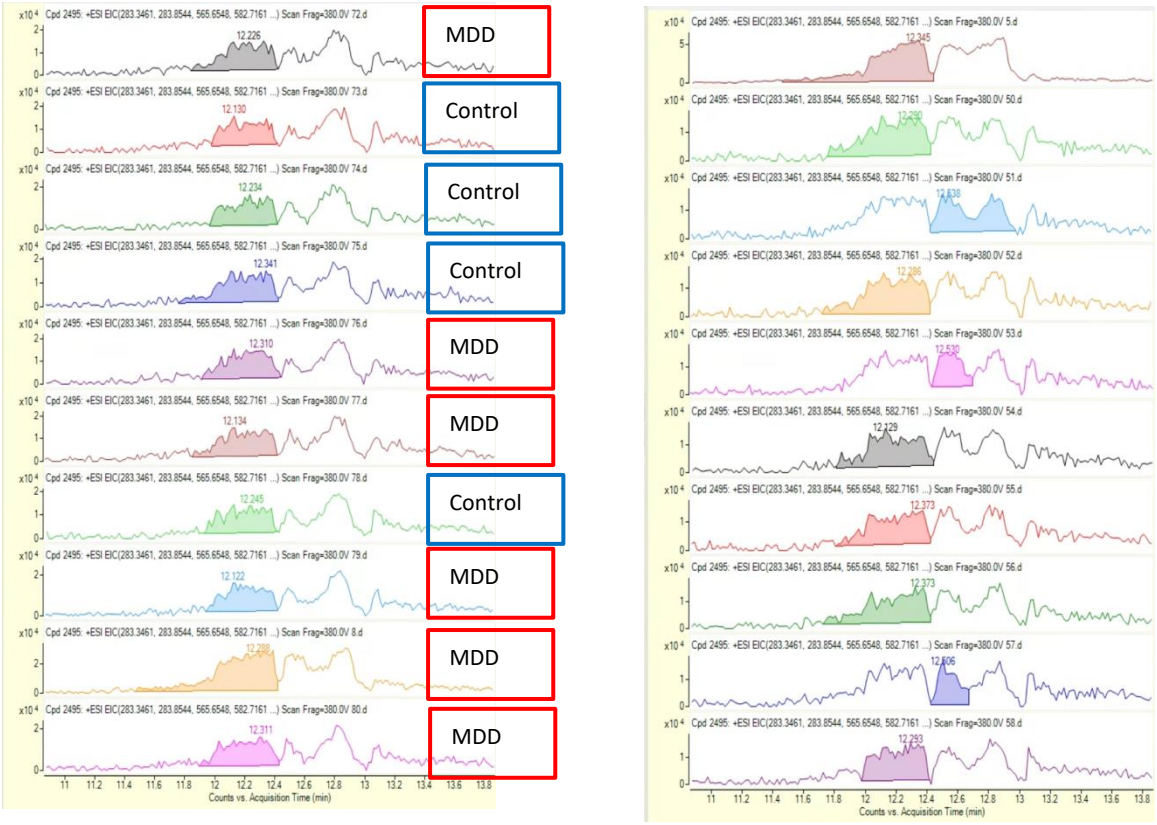

Figure 4-15. Raw data showing presence of neutral mass 564.6546 compound throughout selected samples. It shows that there is no

RP-LC Negative Mode

The method is Mass Profiler brought up no compounds with a fold change larger than 5 for this mode and HILIC positive and negative mode. The method was adapted to a fold change larger than 2 and some features were extracted.

However, for this RP-LC negative mode, the results were still only bringing compounds that were u*p*-regulated in control mode compared to the MDD sample groups. This can be seen in **Table 4.13**.

Table 4.13. Table of fold change masses in Profiler RP-LC negative mode

RT (min)	Mass (Da)	Peak Abundance		Log2(A1/A2)	FC
		MDD	Control		
11.990	434.3268	54262	270712	-2.32	-4.99
12.585	434.3340	92914	275327	-1.57	-2.97

These are therefore not ideal markers as the fold change is less than 30. These compounds are also eluting close to the wash in the run and therefore the mass may be slightly in-accurate due to the busy spectra with possible ion suppression. They are also of very similar neutral mass and suggest that they are the same compound eluting over the wash region.

4.5.4.5 MetaboAnalyst

RP-LC Positive Mode

Fold Change Results

When looking at the fold change results between the two sample groups, the following two features have a fold change between 4 – 7, **Figure 4.16**. However, looking at the samples showing this difference in the MDD groups, there are only a few samples that possess the increased concentration. Therefore, these are not potential biomarkers. There are no other features with a significant fold change showing that the sample groups are not significantly different to each other.

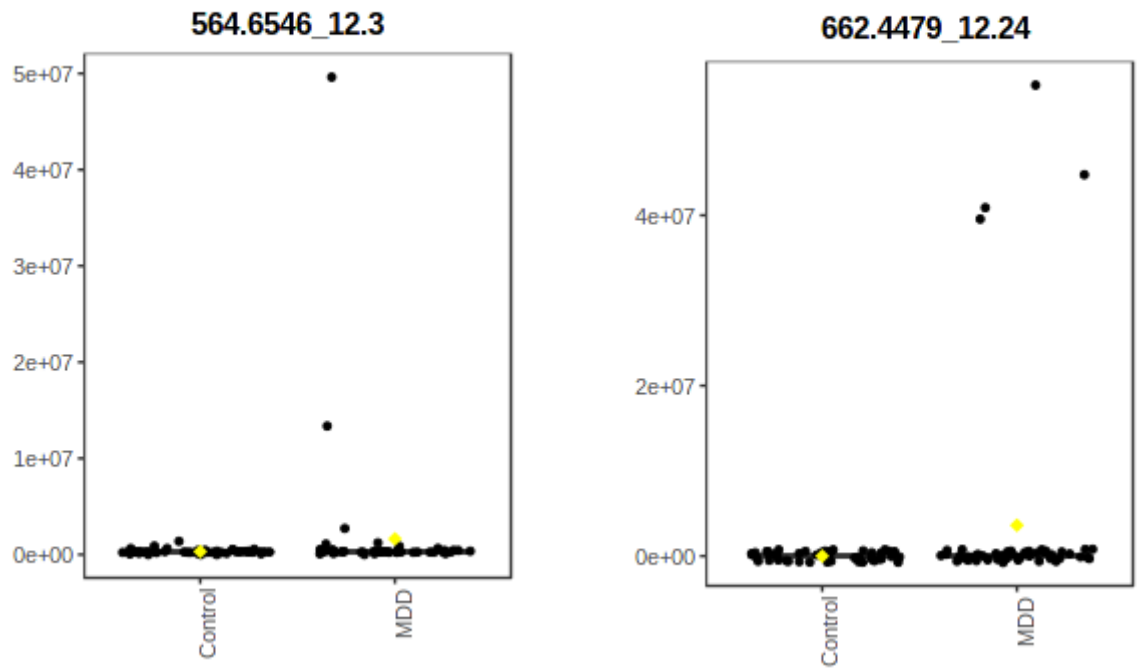

Figure 4-16. MetaboAnalyst Fold Change results for RP-LC Positive mode. FC 4-7.

The figure shows that the compounds with neutral mass of 564.6546 and 662.4479 Da both have less than 5 samples with a positive abundance in MDD sample group.

T-Test Results:

There are no significant features with a *p*-value of 0.01, this was increased to 0.05 and there were still no significant features. Therefore, it can be deduced that there is no significant difference between the two sample groups according to MetaboAnalyst.

There are other tests that can be performed in MetaboAnalyst, however since it is consistent with the other analysis software of bringing up no significant differences, these tests were not performed.

RP-LC Negative Mode

Fold Change Results:

There is one feature with a fold change of 5.3 in **Figure 4.17**. This means there are some differences between the mean averages of MDD and Control groups. However, this was not consistent and can be seen to be at the same increased level in some of the control patients, as well as in the MDD. There are also a lot of samples where the compound concentration is at/near 0. This is again, not a

potential biomarker for MDD due to the lack of consistency in all samples. A biomarker needs to be present at a significant concentration in all samples within one group, and at a significantly lower concentration in all samples within the opposite sample group.

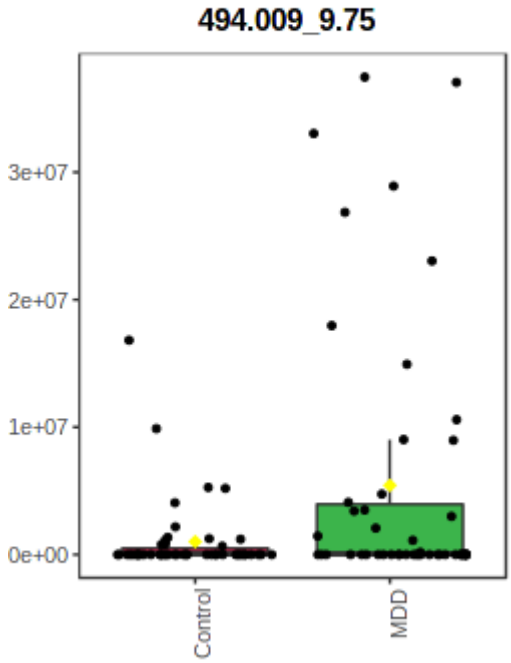

Figure 4-17. MetaboAnalyst Fold Change results for RP-LC Negative mode. FC 5.3.

The figure shows that the compound with mass of 494.0090 Da has varying abundance in both sample groups which makes it a non-ideal biomarker due to the inconsistencies across the control and disease group.

T-Test Results:

There are no significant features with a *p*-value of 0.01, this was increased incrementally to 0.1 and there were still no significant features. Therefore, it can be deduced that there is no significant difference between the two sample groups.

Due to the lack of consistent differences between the datasets, no other tests were performed in MetaboAnalyst.

4.5.4.6 R

Using R, Mann Whitney T tests and Chi Squared tests were performed to determine the *p*-value, q-value (corrected *p*-value) and fold change results for each feature. R was then used to determine

how many of these features have a *q*-value of significance; less than, or equal to 0.05. To continue with the analysis conducted in other software, features with *p*-values ≤ 0.01 are of significance. The chi squared tests determine how many missing values are present in each group, with the test stating how many features are only present in one group.

RP-LC Positive Mode

The Mann Whitney test declared there are 0 features of significance. The Chi-Squared test also declared 0 features that are missing in only one sample group.

The top 5 features with the lowest *p*-values, alongside the fold change results are shown in **Table 4.14**.

Table 4.14. Top features of significance with lowest p-value*, fold change also listed.*

Ordered by significance of p-value*.*

Mass_RT	Fold Change	*p*-value
380.3245_12.01	0.77	0.021
641.5642_12.25	0.79	0.056
166.0058_2.39	23.23	0.058
300.0563_12.87	1.07	0.086
343.3015_9.65	1.09	0.086

The **Table 4.14** shows that no features have a *p*-value of significance, and the fold change results are also all less than 30, with only 1 feature with a value higher than 1.1. This shows that these features are very similar abundance in both sample groups and therefore are not biomarkers.

RP-LC Negative Mode

The Mann Whitney test declared there are 0 features of significance. The Chi-Squared test also declared 0 features that are missing in only one sample group.

The top 5 features with the lowest *p*-values, alongside the fold change results are shown in **Table 4.15**.

*Table 4.15. Top features of significance with lowest p-*value*, fold change also listed.*

*Ordered by significance of p-*value*.*

Mass_RT	Fold Change	*p*-value
345.8688_12.94	1.05	0.017
752.5888_12.02	0.71	0.018
769.0609_8.93	0.52	0.019
303.8325_13.17	0.62	0.020
527.1302_11.98	0.76	0.035

The **Table 4.15** shows that no features have a *p*-value of significance, and the fold change results are also all less than 30, with the highest fold change of 1.05. This shows that these features are very similar abundance in both sample groups and therefore are not potential biomarkers of MDD.

Overall, R shows that there are no features of significance in any of the phases or polarities.

4.6 Discussion

4.6.1 Results Discussion

Overall, there are no potential biomarkers that are consistent in more than one software analysis package. When one software package claims a potential compound that can be used as a marker, the raw data in Qualitative analysis software shows it is a false positive.

The QC samples are VDBS, which means that they are freshly spotted within the months leading up to the analysis. The MDD and control samples are older, spotted as early as 2017. A stability Experiment was conducted at the start of the Delta Study by University of Cambridge and some minor differences were seen, though it was concluded that majority of the sample remained unchanged due to the strict storage of the samples [28]. Within the literature, the stability of DBS samples hasn't been examined for much longer than 100 days, with some exceptions. However, these samples have been stored for 5 years and therefore there is most likely some decay occurring

within the samples. There is one paper which shows the decrease in abundance of a randomly selected small molecule from an average of 1000 to 250 over the space of 5 years, indicating that even when the DBS cards are stored at low temperature, the compounds are unstable from one year to the next [29]. This paper does not mention the use of a desiccator within the DBS packet, which has shown to be essential for keeping moisture from hydrolysing the compounds within the samples [29]. Therefore, this is possibly why the compounds were not stable over the years due to hydrolysis occurring. The PCA plots throughout Chapter 4 show that the QC samples sat away from the rest of the samples. A possible reasoning for this would be the age of the samples, showing that there is some decay within the samples overtime that is leading to this difference seen by the first principal component throughout all the modes of analysis. However, all the samples used have been taken within the same time-period, there are no samples later than 2017 and therefore it may be assumed that the potential decay of some compounds within the samples have all occurred at a similar rate, so they can be compared against each other directly. The QC samples can still be used to test for plate effects and instrument stability.

The same project but using more fresh samples would be ideal, giving more of an overview of the compounds within the patient's blood at the time of sampling, rather than allowing the concentrations to potentially fall over time. Any potential differences within the sample groups could be missed by allowing this change to occur before analysing the samples. It would also be ideal if the same volume of blood could be collected each time as some spots of the same diameter will have a different volume of blood due to the people's different haematocrit levels. Volumetric absorptive microsampling (VAMS) is an advanced technique that overcomes this issue as it uses capillary action to take a fixed volume of blood, regardless of haematocrit level, in less than 4 seconds [31]. The VAMS are then dried at room temperature and can be sent through mail [32]. This simple dried blood collection technique is user friendly and is more reliable than DBS cards for use of blood biomarker studies due to the fixed volume of blood taken each time. This technique was used during COVID-19 clinical trials due to easy sampling, delivery and storage [32].

It is also possible that blood is not the most ideal sample for this analysis. It is possible that another matrix like cerebrospinal fluid would be more suited, however this is more invasive technique and would require professionals to take the samples. The fact that people can sample their own blood in their own time and send the samples through the post makes the DBS analysis ideal, cheap and easy to get more willing volunteers, than if they were to need a hospital appointment to take a portion of their cerebrospinal fluid.

4.6.2 Software Discussion

Each piece of software has given the same results; that there are no biomarkers present. There have been some false positives throughout in every piece of software, particularly in MPP. XCMS online was not used in this project due to the false positives and false negatives it has given in the previous Chapters and due to the large number of files that would need to be uploaded, there was a lack of data space to have all 112 sample files. R was used, however this required the input of a .csv file and so the deconvolution was conducted in Mass Profinder first. Therefore, it is just the statistical analysis that is compared to MPP and MetaboAnalyst as they will all have the same deconvolution parameters. R is a free programming language software that is widely used; however, it is the most complicated software used throughout this project due to its language. R scripts are needed to be written, which is in a different language and lacks visual techniques. It is possible to create simple graphics, which are customisable, however these do not compare to those provided in MPP as they are so simple. Once a script for some analysis is written, the analysis is simpler as the same script can be used and adapted slightly for each different .csv data file. Though, the fold change values were limited after 10,000 as it describes this value as infinite, therefore these cannot be ordered more than this. It also didn't give an option to export the average abundances and so the whole list could only be ordered by the limited FC. The *p*-value or *q*-values were determined by median of group abundances and rounded to a set value, giving the same values for multiple features. It is a free software platform and given that the results can be exported as simple lists, it is useful to use for

metabolomics. There are support platforms online where other users share scripts and offer help, making it a nice helpful community for quick analysis time.

Due to the results of this analysis, it is difficult to see which software was the most accurate since there are no lists results to look at, though the fact that they all consistently gave no results is a positive result, showing consistency throughout. However, there were false positives that appeared throughout that needed to be confirmed in the raw data.

The reasoning behind Mass Profinder showing so many false positives could be because there were significantly less QC files than group sample files, the parameters had to be set to a compound showing consistently in 'at least 5 samples within at least one group' since there were 5 repeat QC samples. Without the QC samples, it could be set to a higher value, possibly showing in a minimum of '40 out of 50 samples within at least one group'. Though, when this Experiment was performed, only compounds that were in both groups showed and therefore this was disregarded. The reason that only 80% of the files needed to have a possible marker in it is to allow for possible miss-diagnosis of patients. MDD is hard to diagnose, given that it is based upon a GP asking the patient open ended questions with very little time to listen to the answers. This limited time is what often leads to bipolar being diagnosed with MDD, causing problems with any potential biomarkers present only in MDD patients [30].

4.6.3 Overall Discussion

There is no difference between the control and MDD sample groups. This could be due to a lot of different factors, but the main one assumed was the sample preparation stage or the age of the samples. During the project, it came to light that it is not ideal to look at 'control vs disease' datasets as it is then unknown whether any potential markers found are due to the MDD diseased state, or general stress markers due to MDD. Two diseased states directly compared would be more ideal as both would produce stress on the body and therefore the markers determined would be due to the individual state only. Once these are determined, a control set can then be used to see whether these markers are also present, leading to whether they may be up or down regulated in the diseased state compared to the control state.

The sample preparation step could be a reason behind the lack of differences between the sample groups as any lipid type molecules could be missed with the current preparation. The most simple, effective procedure was previously used to try capture the whole picture of the blood sample, though further research shows that the current method is not ideal for lipids due to the presence of water, where-as a lot of lipids associated with psychiatric disorders are hydrophobic molecules. If the major differences between MDD and control groups happen to be a hydrophobic, lipid molecule, then this would have been missed in this analysis as it would have not been extracted from the DBS sample card.

4.7 Conclusions

The data throughout this study has shown to be repeatable confirmed by using the QC samples. The RP-LC method showed to be most reproducible and the QC samples had minimal variability. The HILIC method showed a drop in the baseline which needs to be investigated further, but still no new peaks appeared showing some success.

The statistical approach used were not able to find significant differences in this data, with no markers discovered, though each piece of software used agreed that there were no statistically significant differences, which shows that they are all working well with each other. The multivariate statistics employed for this study were successful in proving this lack of differences. The PCA plots did not cluster into groups, except showing that the fresh VDBS QC samples are different to the older MDD and control samples.

Additional investigations need to be conducted to confirm no changes within the sample groups. This would entail changing the sample preparation and LC methods to cover the molecules that were potentially not extracted from the DBS card, or not visible on the TIC due to incompatibility with the LC columns chosen. For example, the methods (both sample preparation and LC method) are not optimised for lipid profiles and therefore potential important data may be missed.

4.7.1 Future Work

Following on from this, the next step was to go back with the analysis results to the University of Cambridge and tell them that there seemed to be no difference within the groups for the data collected, using this sample preparation and the two LC methods.

A study was conducted by another research group, though this time the study focused on bipolar vs MDD. This solved the issue of looking at control vs disease, as both bipolar and MDD are seen as 'diseased', ensuring that this study cannot lead to discovery of a "stress" biomarker rather than an MDD or bipolar biomarker. The study focused on the lipidomics with more concentrated samples and a potential difference between two sample groups was found. This data analysis was conducted using a QQQ which means that the data was not of high accuracy as only masses to 1 decimal place were collected, whereas the Q-ToF collects data up to 5 decimal places. Their study showed that there were differences in some lipids between the two sample groups, with one thought to be a hex-ceramide. The external company sent over the samples in 96 well plates, dried and stored under

nitrogen. These were re-constituted in 50:50 methanol:chloroform and a method more suited to lipids was developed using a C8 column. Results showed a difference in the levels of glucosylceramide, with elevated levels in patients with MDD. The standards of galactosylceramide and glucosylceramide were ordered and separated using a polymer column, hence the positive identification of glucosylceramide using MSMS and retention time confirmation. It is possible that if this one ceramide compound is showing differences in the dried blood spots, then other lipids or ceramide compounds may also show differences that were missed in the original analysis conducted on RP-LC or HILIC columns. Therefore, in future work the project would be adapted to look at the new sample sets on the untargeted screening method using the Q-ToF, with the sample preparation and LC methods optimised for lipid type compounds using chloroform that dissolves these ceramides. Using the workflow with the deconvolution in Mass Profinder then using the comparison software packages seen previously (MPP, R, Profiler and MetaboAnalyst), these differences should be observed.

4.8 References

1 Li K, Naviaux J, Monk J, Wang L, Naviaux R. Improved Dried Blood Spot-Based Metabolomics: A Targeted, Broad-Spectrum, Single-Injection Method. Metabolites. 2020;10(3):82.

2 Zakaria R, Allen K, Koplin J, Roche P, Greaves R. Advantages and Challenges of Dried Blood Spot Analysis by Mass Spectrometry Across the Total Testing Process. EJIFCC. 2016;27(4):288-317.

3 Guthrie R, Susi A. A simple phenylalanine method for detecting phenylketonuria in large populations of newborn infants. Pediatrics. 1963;32(3):338-343.

4 Martial L, Aarnoutse R, Schreuder M, Henriet S, Brüggemann R, Joore M. Cost Evaluation of Dried Blood Spot Home Sampling as Compared to Conventional Sampling for Therapeutic Drug Monitoring in Children. PLOS ONE. 2016;11(12):e0167433.

5 E Balashova E, P Trifonova O, L Maslov D, G Lokhov P. Application of dried blood spot for analysis of low molecular weight fraction (metabolome) of blood. Health and Primary Care. 2018;2(5).

6 Han M, Jun S, Song S, Park H, Park K, Song J. Ultra-performance liquid chromatography/tandem mass spectrometry for determination of sulfatides in dried blood spots from patients with metachromatic leukodystrophy. Rapid Communications in Mass Spectrometry. 2014;28(6):587-594.

7 Manicke N, Abu-Rabie P, Spooner N, Ouyang Z, Cooks R. Quantitative Analysis of Therapeutic Drugs in Dried Blood Spot Samples by Paper Spray Mass Spectrometry: An Avenue to Therapeutic Drug Monitoring. Journal of the American Society for Mass Spectrometry. 2011;22(9):1501-1507.

8 Li Q, Cao D, Huang Y, Xu H, Yu C, Li Z. Development and validation of a sensitive LC-MS/MS method for determination of tacrolimus on dried blood spots. Biomedical Chromatography. 2012;27:327-334.

9 Denniff P, Spooner N. Effect of storage conditions on the weight and appearance of dried blood spot samples on various cellulose-based substrates. Bioanalysis. 2010;2(11):1817-1822.

10 Buckton A, Prabhu D, Cane P, Pillay D. No evidence for cross-contamination of dried blood spots excised using an office hole-punch for HIV-1 drug resistance genotyping. Journal of Antimicrobial Chemotherapy. 2009;63(3):615-616.

11 Johnson S, Lange B. Open-Access Metabolomics Databases for Natural Product Research: Present Capabilities and Future Potential. Frontiers in Bioengineering and Biotechnology. 2015;3.

12 Mirnaghi F, Caudy A. Challenges of analyzing different classes of metabolites by a single analytical method. Bioanalysis. 2014;6(24):3393-3416.

13 Organization World Health. Depression and other common mental disorders: global health estimates [Internet]. Apps.who.int. 2017 [cited 1 June 2022]. Available from: https://apps.who.int/iris/handle/10665/254610

14 Alonso J, Angermeyer M, Lepine J. The European Study of the Epidemiology of Mental Disorders (ESEMeD) project: an epidemiological basis for informing mental health policies in Europe. Acta Psychiatrica Scandinavica. 2004;109(s420):5-7.

15 Michalak E, Yatham L, Lam R. Quality of life in bipolar disorder: A review of the literature. Health and Quality of Life Outcomes. 2005;3(1):72.

16 Chesney E, Goodwin G, Fazel S. Risks of all-cause and suicide mortality in mental disorders: a meta-review. World Psychiatry. 2014;13(2):153-160.

17 van Wijngaarden B, Schene A, Koeter M. Family caregiving in depression: impact on caregivers' daily life, distress, and help seeking. Journal of Affective Disorders. 2004;81(3):211-222.

18 Fifth Edition. Diagnostic and statistical manual of mental disorders. Am Psychiatric Assoc. 2013;5(21):591-643.

19 Han S, Tomasik J, Rustogi N, Lago S, Barton-Owen G, Eljasz P et al. Diagnostic prediction model development using data from dried blood spot proteomics and a digital mental health assessment to identify major depressive disorder among individuals presenting with low mood. Brain, Behavior, and Immunity. 2020;90:184-195.

20 Mitchell A, Vaze A, Rao S. Clinical diagnosis of depression in primary care: a meta-analysis. The Lancet. 2009;374(9690):609-619.

21 Olesen J, Gustavsson A, Svensson M, Wittchen H, Jönsson B. The economic cost of brain disorders in Europe. European Journal of Neurology. 2011;19(1):155-162.

22 Hirschfeld R, Lewis L, Vornik L. Perceptions and Impact of Bipolar Disorder. The Journal of Clinical Psychiatry. 2003;64(2):161-174.

23 Patel R, Shetty H, Jackson R, Broadbent M, Stewart R, Boydell J et al. Delays before Diagnosis and Initiation of Treatment in Patients Presenting to Mental Health Services with Bipolar Disorder. PLOS ONE. 2015;10(5).

24 Kakuma R, Minas H, van Ginneken N, Dal Poz M, Desiraju K, Morris J et al. Human resources for mental health care: current situation and strategies for action. The Lancet. 2011;378(9803):1654-1663.

25 Olmert T, Cooper J, Han S, Barton-Owen G, Farrag L, Bell E et al. A Combined Digital and Biomarker Diagnostic Aid for Mood Disorders (the Delta Trial): Protocol for an Observational Study. JMIR Research Protocols. 2020;9(8)e18453.

26 Martial L, Aarnoutse R, Mulder M, Schellekens A, Brüggemann R, Burger D et al. Dried Blood Spot sampling in psychiatry: Perspectives for improving therapeutic drug monitoring. European Neuropsychopharmacology. 2017;27(3):205-216.

27 Ozcan S, Cooper J, Lago S, Kenny D, Rustogi N, Stocki P et al. Towards reproducible MRM based biomarker discovery using dried blood spots. Scientific Reports. 2017;7(1)45178.

28 Chambers A, Percy A, Yang J, Borchers C. Multiple Reaction Monitoring Enables Precise Quantification of 97 Proteins in Dried Blood Spots. Molecular & Cellular Proteomics. 2015;14(11):3094-3104.

29 Rus C, Di Bucchianico S, Cozma C, Zimmerman R, Bauer P. DBS Methodology Study for Biomarker Discovery in LSD. Metabolites. 2021;11(6):382.

30 Martin-Key N, Olmert T, Barton-Owen G, Han S, Cooper J, Eljasz P et al. The Delta Study – Prevalence and characteristics of mood disorders in 924 individuals with low mood: Results of

the of the World Health Organization Composite International Diagnostic Interview (CIDI). Brain and Behavior. 2021;11(6).

31 Denniff P, Spooner N. Volumetric absorptive microsampling: a dried sample collection technique for quantitative bioanalysis. Anal Chem. 2014;86(16):8489–8495.

32 Harahap Y, Diptasaadya R, Purwanto DJ. Volumetric absorptive microsampling as a sampling alternative in clinical trials and therapeutic drug monitoring during the COVID-19 pandemic: a review. Drug Des Devel Ther. 2020;14:5757.

5 – Overall Concluding Remarks

5.1 Summary

This research has resulted in the development and evaluation of different comparison software platforms for deconvolution and statistical analysis across a range of sample types, including plant and different human samples. Each study showed the same overall results of the software, giving confidence that the conclusions are accurate. The studies show how this workflow can be applied to any samples and by using Mass Profinder in combination with Mass Profiler Professional, accurate and reliable results can be produced. Results provided by Mass Profinder combined with MetaboAnalyst are just as reliable and provides a great alternate platform for those who cannot justify the cost of MPP, however the multivariate statistics are not as high in quality as those provided in MPP. During this work, XCMS online did not provide any reliable or accurate results which could be due to a lack of optimisation of the parameters since they were tried to be kept similar to the deconvolution parameters used in the other deconvolution software. R is focused more on the statistical analysis with complicated coding skills required that makes it more difficult to use than MPP and MetaboAnalyst. The same results can be obtained in R and therefore could be used as a backup in order to provide confirmation of results after MPP or MetaboAnalyst.

It's a common belief that the different results that are obtained throughout the different pieces of software are a part of a large issue that needs further examination across metabolomics. There may be companies trusting the software is giving the best results and potential research science changing biomarkers may be missed, should the wrong software be used. Analysis should be repeatable and reliable, yet the same data gives different results using the same parameters with the same vendor specific software programs, Mass Profinder and Mass Profiler, from Agilent. A wider range of software needs to be tested in order to get the most reliable answers, but from those tested throughout this thesis, Mass Profinder followed by MPP is a personal favourite that will be continued

to be used in the future for comparison work. This is due to its simplicity and ease of looking back at the raw data, as well as providing the most reliable results that have been found so far.

One step which has increased difficulty is identification, and/or annotations, of metabolites. These need to be evidence based, but there is an increased concern in the validity of proposed identities in metabolomics that do not seem to exist in other areas such as genomics and proteomic data analysis due to the potential easier identification procedure, with some exceptions i.e., protein post-translational modifications [8]. Improving the identification reporting procedure is crucial for metabolomics to maintain its value. At present, there are many poor, biologically plausible identifications in the literature with proposed molecules that do not comply with chromatographic data, including the physicochemical properties (hydrophobicity/hydrophilicity) [8]. To identify, or annotate, compounds reliably, database searches are often conducted which use retention times, accurate masses and in some cases, fragmentation patterns. The rest of the identification validity needs to be done by the user, in which the likeliness of an identification being plausible needs to be considered before claiming confirmed identity of a compound. The reporting procedure should stay compliant to the minimum reporting standard rules from Metabolomics Standards Initiative (MSI) to keep the community informed of how reliable an annotation or identification may be [10].

5.2 Conclusions

The work aimed to address the issues that can arise by using different software analysis platforms for the same analysis. The software a user chooses for chemometric analysis is another added variable for each experiment and therefore, it is of interest to determine whether by using the same parameters, the same biomarkers can be determined with each platform. These studies aimed to develop a data processing and statistical analysis workflow that's suitable for untargeted metabolomic studies. A comparison of results to see which software platform gives the most accurate, reliable and comprehensive results are important to enhance the understanding of the untargeted metabolomic workflow of choice. Although the software needs to be easy to use, it

should never be the case to totally remove the need for an expert to use the software as this can lead to flawed studies due to the non-expert solely basing parameters on the default settings [4]. Most scientists are not trained in programming or chemometrics, nor is it easy to find assistance in training to a substantial level unless you know who to ask, therefore this problem may also influence the output of data analysis. This problem is well known in the literature, however it continues to hinder application of chemometric approaches in deconvolution and statistical analysis [5-7].

The first study was used to develop and evaluate an untargeted LCMS workflow, whilst comparing the different software types to see what the benefits are of using each one using tea plant samples. Each piece of software used showed that differences between the tea samples could be determined, with MPP giving the best multivariate statistics to determine which samples were most similar and most different to each other. The statistical approach in MPP also gave accurate masses that causes the differences between each sample, with some putatively annotations provided for those compounds with matches database. The untargeted metabolomic workflow using Agilent's Profinder for deconvolution, then MPP for statistical analysis enabled each objective to be accomplished, something that could not be done without the comparison software. XCMS online and Mass Profiler were also used to deconvolute the data however the results were not as good as those provided by Profinder as a lot of false positives and negatives were found.

Following on from this, a different type of sample matrix was analysed in the next study. Human hair was now used to determine what additional information could be obtained by using the same sample preparation and LC method procedures currently followed by a company. The data that have been gathered in this work has shown overall to be reproducible and reliable with multiple masses that have repeatedly appeared in the results for potential biomarkers, despite the lack of a QC sample used. Using a statistical approach in each piece of software, a wide selection of masses was determined as potential biomarkers for bleached hair, with some of these appearing in multiple comparison software results. XCMS online showed multiple random masses that appear to be false

positives and in positive ionisation mode and many of the final chosen biomarkers did not appear in the results, despite the raw data confirming the presence in only the bleached samples, showing false negatives. The best combination of software used was the same as the first study, with Mass Profinder and MPP giving the most accurate results in the easiest format. MetaboAnalyst and R also gave accurate and reliable results however these do not conduct deconvolution and therefore Mass Profinder should also be used. The experiment needs to be repeated on a larger sample selection, with a more varied range of hair colours. Tests also need to be conducted to determine whether the potential biomarker is produced consistently, and the levels along the hair shaft should be examined.

Moving onto a different sample type in the next study, a liquid human blood sample through the use of dried blood spots was used with the untargeted LCMS metabolomic workflow proved to be repeatable. The VDBS QC samples showed minimal variability however the statistical analysis showed that there were no compounds of significant difference between the two sample groups, MDD and control (age, gender and BMI matched). This does not mean that no differences occur, only that the samples provided, with the simple sample preparation and the RP-LC / HILIC methods used do not show differences. The multivariate statistics also showed no difference between the sample groups which adds to the reliability of the statistical analysis. Further work is to be conducted looking at disease vs disease sample sets, with the focus moved onto lipid compounds with alternate sample preparation and potentially a more focused LC method.

Despite the determination of biomarkers in studies, there are clear differences in the results obtained from each software package. Each study showed that Mass Profinder is a key part of the comparison analysis. It comes with the 'Agilent MassHunter Suite' and therefore, if the instrument used is already an Agilent one, it is possible that this software is free. Without Mass Profinder, analysis in MPP, MetaboAnalyst and R is not possible as they do not have deconvolution ability in their software. They need the raw Agilent data files (.d) to be converted into another format for reading, such as .cef or .csv files. However, statistical analysis and multivariate analysis cannot be

conducted in Mass Profinder. Due to random noise present in untargeted LCMS data, it is important to visually inspect the data in Mass Profinder to ensure the correct peak is selected prior to exporting the data to ensure that false positives or negatives are kept to a minimum. I appreciate that this may take time to go through each feature and sample, however it will greatly improve the quality of data and in turn, improve the quality of statistical analysis in the next steps. This will then least to more reliable and robust findings.

MPP was overall the best secondary analysis, which can perform both in depth statistical analysis, and high-resolution graphics for multivariate statistics. The biomarkers declared were reliable and accurate throughout the studies. However, **Table 5.1** shows the cost of MPP to be starting at £20,500 for one licence code, meaning that it can be used on one device only. Each time you request an upgrade to a 'newer/better' version, there is an upgrade cost of £11,600 and therefore there is a high cost that comes with the great statistical analysis.

Table 5.1. Cost of each software used in the studies.

This shows that MPP is the most expensive, with MetaboAnalyst and R cheapest due to no cost to download.

	XCMS Online	**Mass Profiler**	**Mass Profinder**	**MPP**	**Metabo-Analyst**	**R**
Cost of Software	Free / Purchasable for £7300 a year	Free with Agilent MassHunter suite	Free with Agilent MassHunter suite	Starts at £20500	Free – online	Free – download app

Depending on how often this type of analysis is conducted, the price may not be worth the large cost since other software platforms give results just as accurate, but for free. The same type of statistical and multivariate analysis can be conducted in MetaboAnalyst, which also provides a high level of reliable biomarkers. The multivariate statistics graphics are not as high quality as those shown in MPP and are not customisable, however since MetaboAnalyst is free to use and run, it comes at a great advantage and severely reduces the need for purchasing MPP if it is used only sparingly. It would also depend on what the aim of the experiment is, for those only wanting to determine potential biomarker masses with retention times, the high-quality graphics are not needed and

therefore reduces the need for MPP since MetaboAnalyst can be used and provides accurate results, but does not include data validation. R can also be used, providing highly accurate and comprehensive results that are customisable, providing that you are able to understand and adapt the coding language. Using R, a large amount of statistical analysis can be conducted also, however this requires a large amount of training to determine how to conduct the analysis in the best, and easiest way. The graphics for the multivariate statistics, such as PCA scores plot, are not great however they are easy to read and can be customised, such as labels and colours. MetaboAnalyst provides results as accurate and reliable as R, but MetaboAnalyst is more user friendly.

XCMS online provided many false positives and the website online is more difficult to navigate than MetaboAnalyst. A large amount of the features it detected had poor peak shape or a high amount of noise, giving little confidence that they are true features. True features refer to the features that have a matching *m/z* value across multiple samples, and the RT is within the defined window. These values tend to be within 5 ppm mass error and 0.1 minute RT window. There seems to be 'bugs' on the website where various buttons/functions do not work, and when 'help' is contacted there is no reply. The multivariate statistics are basic and non-customisable, and the statistical analysis is not as accurate as MPP when the raw data are examined. However, only the free version of XCMS online was used and perhaps the upgraded version that is available may be more developed and easier to navigate. In contrast to this, there are multiple papers in the literature that use XCMS online and have succeeded in producing credible results. The reasoning behind the difference in experiences may be down to the optimisation of parameters as slightly different parameters can lead to different outcomes within data analysis. With significant time spent on manually optimising each parameter and selection criteria, deeper knowledge on programming languages and parameter interpretation is needed, but in turn provides better fine-tuning of the data and thus more robust deconvolution [3].

Mass Profiler can only be used for 1 vs 1 analysis, therefore it is not ideal when analysing the QC samples as well as the sample sets, as well as when there are more than 2 sample groups to be

compared. With that said, it is ideal for finding major differences between two datasets and can perform basic statistical analysis. Multivariate statistics are not able to be performed in this software which is a limitation. Despite this, the software can determine reliable and accurate biomarkers for diseased datasets.

To summarise, this research has successfully developed and evaluated a data processing and statistical workflow for the detection of markers in a sample group by using an untargeted metabolomic workflow with the Q-ToF, with software programs Mass Profinder and MPP giving the best results. The use of QC samples in some studies ensured that the methods were robust prior to statistical analysis. The studies show that different results can be obtained by using different comparison software platforms, hence the need for more than one software to be used at a time to ensure that no important information is lost. Using Mass Profinder and MPP for a project, only the practical sample preparation and LC methods need to be developed and optimised for each sample set; the statistical and multivariate analysis can stay consistent with few parameters needing optimising for each data set.

The results that have been obtained show that good results firstly depend on the quality of data obtained. Good chromatographic separation on a column phase that is suited to the needs of the analysis is crucial, i.e., lipid compounds that showed differences in DBS samples in chapter 4 were overlooked due to the methods used not being suited for lipids and therefore they were not a part of the initial data analysis. However, for full metabolomic profiling, no single method will be sufficient as in complex matrices, there are more than likely going to be a lot of metabolites belonging to various chemical groups that require multiple analytical techniques. Not only does this mean different column phases are required, but it also means that different chromatographic techniques should be considered such as liquid chromatography and gas chromatography. One technique is not the best for all sample sets.

There needs to be a significant understanding of statistical analysis in order to turn the large amounts of data collected during metabolomics experiments, into something that is easily understood and consistent throughout. Improper parameters can largely affect the analysis that software provides. For example, PCA plots can be manipulated to make it look like there are differences in the datasets if the parameters are not optimal. There are ways in MPP that PCA plots can be made on data with only a fold change above a selected value, and therefore will skew the data to make it look like there is a difference between datasets, when if all data was used, there would be no significant difference seen in the clusters [3]. This is the same with improper use of partial least squares discriminant analysis (PLS-DA) plots. This is a multi-variate statistical tool that is widely available in most software packages and using the default settings, is an easy tool to implement into data analysis. It is particularly useful for noisy data and can provide a lot of useful information such as scores and loading plots [9]. However, when it is used by those not fully trained in programming and chemometric analysis, there are many potential limitations when used alongside metabolomic data. Gromski *et al.,* discuss the alternatives to PLS-DA in which other functions can be used in conjunction with PLS, or as an alternative supervised learning method, such as support vector machines (SVM) and random forests (RF) which are seen to sometimes out-perform PLS-DA [9].

Overall, small variations in processing parameters can lead to different conclusions from the data analysis and potentially jeopardize the integrity, and reliability of the results. The thesis highlights the need for proper optimisation and a deeper understanding of statistical methods for robust results to be accomplished.

5.3 Future Work

Of the large number of comparison software available, only a small selection was studied and compared and an even smaller group of deconvolution software programs. Other platforms are available which should also be considered, i.e., MS-Dial and MZMine2. There is also a platform that combines MetaboAnalyst and R, to be used in conjunction with one another, which would be of interest to discover if it gives any additional information to what they give when the analysis is conducted separately. Each one would require extensive training from someone who has experience. It would also be of interest to determine if there are any differences between the results obtained on the XCMS online (free) and the XCMS that requires payment.

During these studies, any different unsupervised multivariate statistical approaches were taken, i.e., PCA plots and clustering. However, no supervised analysis approached were conducted, such as Partial Least Squares (PLS), despite its capability to build training set models with the ability of predicting sample groups of unknown samples [1-2]. These are some tests that could be conducted in further work to determine whether they give the same results throughout the different software programs, or if they differ largely.

Even though these studies determined some potential biomarker masses for sample groups, it has not been fully determined how reliable and accurate these software packages are. It would be interesting to determine how many markers each software could determine when used in a targeted study. Two solutions with some unique mixed standards with known concentrations could be run on the Q-ToF to determine whether the differences could be obtained in each piece of software.

The analytical techniques used, coupled with the data processing and statistical workflow, proved successful for differentiating the difference in sample groups. It helped discover compounds that have potential for use as biomarkers in hair bleaching and identification of tea clones/types and showed that there were no differences within the control and MDD groups in dried blood spots

supplied. The most useful pieces of software used for the untargeted metabolomic analysis overall were Mass Profinder for deconvolution, then MPP and MetaboAnalyst for a wide range of statistical analysis available, in the most comprehensive, yet simple ways. The application of the same techniques in other areas could be used, showing the further usefulness of these methods such as using gas chromatography, or in other disciplines such as environmental sciences.

5.4 References

1 Ren S, Hinzman A, Kang E, Szczesniak R, Lu L. Computational and statistical analysis of metabolomics data. Metabolomics. 2015;11(6):1492-1513.

2 Miller J, Miller J, Miller R. Statistics and chemometrics for analytical chemistry. 6th ed. Harlow: Pearson Education Limited; 2010.

3 Albóniga OE, González O, Alonso RM, Xu Y, Goodacre R. Optimization of XCMS parameters for LC–Ms Metabolomics: An assessment of automated versus manual tuning and its effect on the final results. Metabolomics. 2020;16(1). doi:10.1007/s11306-020-1636-9

4 Brereton RG, Lloyd GR. Partial least squares discriminant analysis: Taking the magic away. Journal of Chemometrics. 2014;28(4):213–25. doi:10.1002/cem.2609

5 Broadhurst DI, Kell DB. Statistical strategies for avoiding false discoveries in metabolomics and related experiments. Metabolomics. 2006;2(4):171–96. doi:10.1007/s11306-006-0037-2

6 Goodacre R, Broadhurst D, Smilde AK, Kristal BS, Baker JD, Beger R, et al. Proposed minimum reporting standards for data analysis in Metabolomics. Metabolomics. 2007;3(3):231–41. doi:10.1007/s11306-007-0081-3

7 Sumner LW, Amberg A, Barrett D, Beale MH, Beger R, Daykin CA, et al. Proposed minimum reporting standards for chemical analysis. Metabolomics. 2007;3(3):211–21. doi:10.1007/s11306-007-0082-2

8 Theodoridis G, Gika H, Raftery D, Goodacre R, Plumb RS, Wilson ID. Ensuring fact-based metabolite identification in liquid chromatography–mass spectrometry-based metabolomics. Analytical Chemistry. 2023;95(8):3909–16. doi:10.1021/acs.analchem.2c05192

9 Gromski PS, Muhamadali H, Ellis DI, Xu Y, Correa E, Turner ML, et al. A tutorial review: Metabolomics and partial least squares-discriminant analysis – a marriage of convenience or a shotgun wedding. Analytica Chimica Acta. 2015;879:10–23. doi:10.1016/j.aca.2015.02.012

10 L. Sumner, A. Amberg, D. Barrett, M. Beale, R. Beger, C. Daykin, T. Fan, O.Fiehn, R. Goodacre, J. Grin, T. Hankemeier, N. Hardy, and J. Harnly. Proposed minimum reporting standards for

chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). Metabolomics. 2007;3(3):211-221.