Evaluating and Improving True and False Positive Annotations in Non-targeted (nano)ESI-DIMS and U(H)PLC-MS Based Clinical and Environmental Metabolomics

By

Judith Blessing Ngere

A thesis submitted to the University of Birmingham for the degree of DOCTOR OF PHILOSOPHY

School of Biosciences

University of Birmingham

Edgbaston

Birmingham

B15 2TT

August 2021

Abstract

Interest in the chemical exposome is increasing due to mounting evidence of the ubiquity of chemicals in the environment. Metabolomics informs on biological perturbations in response to stressors, including chemicals. Metabolomics and exposomics can therefore find utility in chemical risk assessment. However, since exposomics, the study of all non-genetic exposures an organism experiences from conception to death, is still emerging, the chemical coverage, detection reproducibility, and limitations of non-targeted analysis (NTA) methods applied are unknown. Moreover, all NTA methods face challenges in providing confident identification of analytes. The current strategy for confident identification is through fragmentation. However, good quality fragmentation requires sufficient ion intensities, yet it is known that ~70% of metabolites are too low intensity to give good quality fragmentation spectra, whilst chemicals found in biological samples are about ~1,000 times lower in concentration than endogenous metabolites. As such, they seldom yield good quality spectra for confident spectral matching. This means a large proportion of NTA rely on MS1 data annotation, yet there have been comparatively few investigations into how annotation parameters affect accuracy of annotations.

To characterise NTA chemical coverage, reproducibility, and limitations, direct infusion mass spectrometry (DIMS) and liquid chromatography mass spectrometry (LC-MS) were applied for analysis of chemical mixtures (standards in solvent, fortified serum, house-dust, and wristband extracts) as part of a global ring trial called the US Environmental Protection Agency's (EPA) Non-Targeted Analysis Collaborative Trial (ENTACT) Initially, sample compositions were unrevealed, and annotation was achieved by matching against a suspect-screening reference list with 4,462 chemicals (ToxCast library) based on *m/z* only for both techniques. The second time around, sample compositions were revealed, and this knowledge was used to create small databases containing only chemicals revealed to be in each sample for the DIMS methods, and retention time (RT) databases containing only chemicals detected using the LC-MS methods, against which to match. True and false positive rates (TPR and FPR) were calculated to evaluate method performance. To fill the knowledge-gap about how annotation parameters affect accuracy of MS1 annotations, a software called Birmingham mEtabolite Annotation for Mass Spectrometry (BEAMS) was used to annotate

four LC-MS datasets (serum), varying all parameters used in the annotation steps to ascertain which parameters impacted annotation the most through calculation of TPR and FPR.

For both techniques (DIMS and LC-MS, respectively), using the ToxCast library for annotation, lower sample complexity yielded higher TPRs of (48-74% and 0-94% for analysis of standard mixtures in a clean solvent matrix), which were reduced by increasing sample complexity. LC-MS methods yielded higher TPRs than DIMS methods. However, both techniques resulted in high FPRs (254-879% and 650-2031%), which increased with increasing sample complexity. The use of smaller tailored databases during annotation for DIMS and RT databases for LC-MS reduced FPRs (0-24% and 23-130%). However, for LC-MS methods, the TPR of annotation also decreased (0-74%) since RT databases created were incomplete, containing only chemicals that had been detected repeatedly in three MS1 injections. For DIMS methods, annotation against smaller databases increased the TPR (53-84%).

Optimisation of BEAMS parameters showed that using RT similarity and correlation analysis to group degenerate features, the maximum RT difference parameter had no big impact on the total annotations achieved. However, tighter correlation thresholds reduced the total number of annotations, including both TPRs and FPRs. Mass error tolerances also affected the number of annotations achieved, with tolerances between 0.5-3ppm reducing both true and false positive annotations, with the former highest between 3-10ppm. Finally, the reference lists used for annotation of degenerate features (adducts, isotopes, and neutral losses) TPRs and FPRs of annotation, with longer, more accurate lists created based on each dataset increasing true positive annotation for the positive ion mode datasets relative to shorter default lists. However, these results also demonstrated the pitfalls of using longer reference lists, as TPRs of annotation were reduced for some negative ion mode datasets.

Chemicals in environmental and biological samples can be screened for using both DIMS and LC-MS, yielding high TPRs. However, these techniques do not offer 100% TPRs, therefore new methods are required to increase chemical coverage. The use of smaller tailored reference lists and RT databases for annotation can reduce occurrence of false positive annotations but exemplifies the challenges in creating such small databases. BEAMS optimisations show which parameters affect MS1 data annotation, reduce FPR and maximise TPR. Although these results are specific to datasets and instruments applied herein, they are relevant to anyone using such approaches to group degenerate features, they can be used to guide selection of

appropriate annotation parameters. Continued efforts into maximising MS1 data annotation are required.

Acknowledgements

I would like to thank UKRI and Thermo Fisher Scientific for funding this work. A huge thank you also goes to Professor Warwick Dunn for his efforts in supervising this work, for providing a relaxed working environment, and for putting mental health above all else. I'd also like to thank Dr Catherine Winder for all the advice, support, laughs, coffee breaks, and cake. The four and a half years would not have been bearable without her support.

To Dr William Nash, you are invaluable! Thank you for putting a smile on my face every morning and helping me to stay positive to the end. I appreciate the endless scientific discussions we have had, and I am forever grateful for your bubbly and funny personality and strength of character!

Soon to be Dr Annie Woolven, thank you for your amazing mental health coaching, and emotional support. You are a friend like no other; loyal, kind, committed, and downright hilarious. Also, soon to be Dr Julia Malinowska, thank you for sharing your amazing cat's pictures with me, they filled me with joy and quite often, the giggles. Thank you for your exemplary scientific endeavours, strength, and courage, and thank you for the many DIMS-related conversations and debates. To all the metabolomics 4th floor ladies, you inspire me by being exceedingly strong women in science, and I thank you for that.

Special thanks go to those I hurled endless questions at, Dr Martin Jones, Dr Lukas Najdekr, Dr Andrew Chetwynd, Dr Elliott Palmer, Dr Thomas Lawson, and Dr Andris Jankevics. Thank you for your patience and willingness to help, every single one of you never turned me away.

Finally, I'd like to thank the rest of the metabolomics team on the 4th floor of the School of Biosciences. There was never a shortage of someone wanting an after-work "pint", and I thoroughly enjoyed whiskey and gin-club!

Last, but certainly not least, thank you to my wonderful sister, who had no idea what I was talking about regarding my work but listened anyway. A big thank you goes to my parents, who uprooted their lives from Zimbabwe to the UK to pay for my education. I promise to never go back to "school" again, and yes this is the end of my official education. I hope nobody points out to you that "learning never stops".

Table of Contents

List of abbreviations

(nano)ESI	nano electrospray ionisation
(nano)ESI-DIMS	nano electrospray ionisation direct infusion mass spectrometry
AA	acetic acid
AC	alternating current
ACMG	Analytical and Clinical Metabolomics Group
ACN	acetonitrile
APCI	atmospheric pressure chemical ionisation
API	atmospheric pressure ionisation
APPI	atmospheric pressure photo ionisation
aqRP	aqueous reversed phase
BEAMS	Birmingham mEtabolite Annotation for Mass Spectrometry
CA	correspondence analysis
CE	capillary electrophoresis
CI	chemical ionisation
CRM	charge residue model
C-trap	curved linear ion trap
DA	discriminant analysis
DAD	diode array detector
DART	desorption in real time
DC	direct current
DESI	desorption ionisation
DIMS	direct infusion mass spectrometry
DIOS	desorption on silicone

DMSO	dimethyl sulfoxide
EC	European Commission
ECHA	European Chemicals Agency
EI	electron impact/electron ionisation
EIBPC	extracted ion base peak chromatogram
EIC	extracted ion chromatogram
ENTACT	EPA's non-targeted analysis collaborative trial
EQS	environmental quality standards
ESI	electrospray ionisation
FA	formic acid
FDR	false discovery rate
FN	false negative
FN	false negative
FNR	false negative rate
FNR	false negative rate
FP	false positive
FPR	false positive rate
FT	Fourier Transform
FTICR	Fourier Transform ion cyclotron resonance
FWHM	full width at half maximum
GC	gas chromatography
GC-MS	gas chromatography-mass spectrometry
HILIC	hydrophilic interaction chromatography
HPLC	high pressure (performance) liquid chromatography

HTS	high throughput screening
IEM	ion evaporation model
IM	ion mobility
IPA	isopropanol
IT	ion trap
JRC	Joint Research Centre
LC	liquid chromatography
LLE	liquid-liquid extraction
LTQ	linear trap quadrupole
m/z	mass-to-charge ratio
MALDI	matrix assisted laser desorption ionisation
MI-pack	metabolite identification package
MP	mobile phase
MS	mass spectrometry/mass spectrometer
MS1	full-scan MS
MS2	MS/MS
N.B.	nota bene (Latin)
NCE	normalised collision energy
NEC	new emerging compounds
NMR	nuclear magnetic resonance
NP	normal phase
NRC	National Research Council
NTA	non-targeted analysis
PAHs	polyaromatic hydrocarbons

PC	principal components
PCA	principal components analysis
POCs	persistent organic contaminants
POPs	persistent organic pollutants
ppm	parts per million
QC	quality control
QIT	quadrupole ion trap
Qs	quadrupoles
REACH	registration, evaluation, authorisation, and restriction of chemicals
RF	radio frequency
RP	reversed phase
RSD	relative standard deviation
RT	retention time
S/N	signal to noise ratio
SFC	super critical fluid chromatography
SFE	supercritical extraction
SIM	selected ion monitoring
SLE	solid liquid extraction
SNR	signal to noise ratio
SP	stationary phase
SPE	solid phase extraction
SRMs	standard reference materials
SS	suspect screening
TA	targeted analysis

TOF	time of flight
TP	true positive
TPR	true positive rate
U(H)PLC	ultra-high-performance liquid chromatography
U(H)PLC-MS	ultra-high-performance liquid chromatography-mass spectrometry
UoB	University of Birmingham
UPLC	ultra performance liquid chromatography
US EPA	US Environmental Protection Agency
UV	ultraviolet
v/v	volume in volume
v/v/v	volume in volume in volume
WHO	world health organisation

List of Figures

List of Tables

1 CHAPTER 1: Introduction

1.1 General Introduction: A Rationale for the Project

Few things are more important for human survival than treatment of disease. There is no debating that penicillin is amongst the greatest scientific discoveries, saving millions of lives globally since its debut (1,2). However, it is increasingly evident that treating diseases costs time and money. One therapeutic drug can take 10-15 years to develop whilst costing an eye-watering £620 million (3), although artificial intelligence is accelerating this process and improving cost-effectiveness (4,5) . This has spawned different approaches to human health studies, with focus shifting from solely treating disease to probing causation, exploring prediction, and moving towards prevention.

Historically in epidemiology, which is the study of the distribution and determinants of disease in selected populations, the monocausal theory of disease, wherein onset of disease was attributed to one factor (e.g., a pathogen), was widely popular. Today epidemiologists debate whether other models should be considered, and multi-causal theories have been posited, wherein a disease develops due to multiple factors such as pathogenic, socio-economic and behavioural factors (6,7). Moreover, epidemiology itself has merged with other fields such as toxicology and environmental studies. Where historical perspectives focused mainly on causes of disease being pathogenic, and, to a lesser extent, diseases of old age, epidemiology now considers chemical and environmental exposures, amongst other factors such as any perturbation that acts as a stressor or toxicant. Theoretically, once a causal agent has been identified, it can be avoided or eliminated to aid prevention, or treatments can be more targeted where relevant.

Key to developing therapeutics is understanding cellular mechanisms occurring during disease onset and progression. Indeed, in the last 20 years a new field has emerged, focusing on metabolic changes occurring in living organisms in response to a wide variety of stressors, including disease and chemical pollution, and has been termed metabolomics. Metabolomics is the study of small molecules, usually <1,000 Da, found within a biological system (8). In 1998 this collection of small molecules was first referred to as the “metabolome”, and henceforth the field of metabolomics has gained popularity (9,10). N.B. that not all metabolites are <1,000 Da. Some components of the metabolome such as lipids are larger, and mass ranges of up to 2,000 Da and greater have been reported (11,12). The metabolome consists of both

endogenous (originating from within) and exogenous (originating from outside) metabolites. Not only does metabolomics facilitate understanding mechanisms in action during illness, thereby improving targeted drug design, other applications predict the risk of future onset of diseases before symptoms manifest by looking for metabolic biomarkers that can point to later onset of disease (e.g., prediction of Parkinson's disease, although this research is on-going and not currently applied routinely in clinical analyses) (13). Additionally, metabolomics-based precision medicine is rising. The metabolic response of a living organism to any type of exposure, whether disease, socio-economic, behavioural, environmental, or therapeutic, can be measured. Clinically, this could inform on response to selected drug therapy, overall health of an individual, and recovery progress (14). Metabolomics plays a critical role for human health investigations and disease treatment, but its applications are not limited to human health alone. Any living organism will have a metabolome that will fluctuate in response to a variety of exposures or stressors and can thus be measured.

Metabolomics studies have expanded to focus on other challenges such as chemical pollution, which can adversely impact both human and environmental health. Metabolomics of various plant species has been used to determine and mechanistically understand toxic effects of chemicals (15–18). In-fact, environmental metabolomics exists, aiming to characterise the interaction of biological organisms and their environment (19). Here, the metabolome is used to find biomarkers, or indicators, of chemical or drug toxicity (20–22). The metabolome has also been used for chemical risk assessment in environmental sciences. For example, biomarkers of exposure, i.e., biomarkers that indicate an organism has been exposed to a stressor such as chemical pollution, can be used to assess and monitor the level of chemical pollution in a selected environment (23). Moreover, the metabolome can inform on absorption, distribution, and metabolism of chemical pollutants, alongside biotransformation of these chemicals and their toxic effects. As such, the exposome, defined as 'the totality of non-genetic exposures a living organism is exposed to from conception to death', has gained popularity for both human and environmental health studies (24). The aim of exposomics is to measure both endogenous and exogenous metabolites whilst simultaneously characterising the mixture of exposure compounds effecting metabolic changes (25). Currently exposomics is challenged by a lack of well-defined analytical protocols for characterising the various exposure components a biological system can be exposed to (26).

Metabolomics and exposomics studies, defined above, appear solely biological studies, but in practice the central aspect is the analytical process involved. Technological advancements continuously improve measurement capabilities, which in turn yield more accurate and comprehensive biological interpretations. There are hundreds of thousands of compounds that constitute the metabolome, all with different physico-chemical properties and concentrations and therefore requiring specific preparation and analysis procedures. The analytical platform chosen must be suitable for analysis of a selected sample type, sensitive enough for small sample volumes and low concentration analytes, whilst the methods used must be selective towards different classes of compounds. Tens of thousands of compounds within a metabolome are also similar in shape and often these compounds have the same mass and elemental composition, making them isomers, or the same mass and different elemental composition, making them isobars (27). An appropriate analytical platform must be able to accurately distinguish between them where possible. This adds immense complexity to the analysis, and even more so to data processing. Due to physico-chemical similarities within the metabolome, identification of compounds is very challenging. Two or more isomers could easily be interchanged and given an incorrection annotation or identification (where an annotation is an unconfirmed assignment that required validation and an identification is a confirmed assignment), resulting in a false positive (FP). Conversely, a low concentration analyte may be present within a sample, but not detected due to low analytical sensitivity. This is a false negative. A true positive (TP) annotation occurs when a correct assignment is given. Collectively these issues are termed false discovery, and great efforts are required to minimise FP and negative annotation, whilst maximising TP annotation. The implications of false discovery are obvious. Where metabolites are used as a diagnostic tool, misdiagnosis could be fatal. If used to determine toxicity of a compound, then an incorrect structural annotation could put millions of humans at risk. There are multiple approaches to reducing the false discovery rate. Chemical bonds in isomers may be structurally broken by applying energy to generate fragments. These fragments inform on the three-dimensional structure of analytes and as such can aid in distinguishing isomers. Alternatively, a level of confidence can be assigned to each result. In this way, only high confidence results can be used to make important decisions. It is also paramount that results are reproducible as this increases confidence. The research presented in this thesis will explore all these approaches for reducing false discovery, from the analytical methods to the data processing strategies.

Specifically, method capability will be characterised, and appropriate data processing parameters will be extensively evaluated, and recommendations made for how best to achieve good quality results in the rapidly developing fields of metabolomics and exposomics.

1.2 Metabolomics

Metabolomics is an approach that aims to simultaneously detect and, where possible, quantify the metabolome, generally comprised of metabolites <1,000Da (8), although some components of the metabolome such as lipids (e.g., triacyclglycerols, phospholipids, and sterols) have molecular weights of greater than 2,000 Da (28). The 'metabolome', a term first used by Oliver in 1998, encompasses all low molecular weight metabolites present in a biological sample (29). It is the most downstream expression of biochemical pathways, and therefore integrates the biochemical processes in cells or tissues with the environment (30). It can be used to evaluate exposure effects, alongside a myriad of other applications, and yield important information about the biological status of the system under examination (31).

The importance of studying metabolites has been noted since ca. 1500 BC, when Chinese physicians tasted urine for sweetness to diagnose diabetes and noted that ants were attracted to this sweet-tasting "honey-urine" (32). This was still a far cry from observing the whole metabolome, but it set the foundation for metabolites being useful biomarkers for disease. An Arabic physician also touched on the concept by stating in his book that "the body is in a continuous state of change due to dissolution and nourishment" (33). In the 1940s and 50s, Roger Williams and co-workers proposed that individuals may have "metabolic patterns" whilst studying patients with schizophrenia, reviewed by Gates and Sweeley in 1978 (34).

In 1971, Horning and Horning first coined the term metabolic profile in their work describing gas phase methods for the analysis of metabolites. Here, they highlighted that "these procedures for obtaining metabolic profiles may be used in various ways, including studies of abnormal conditions, drug metabolism, and the effects of drugs on metabolic pathways, as well as for human developmental studies" (35). In 1999, Jeremy Nicholson and co-workers first proposed the concept and nomenclature of metabolomics, calling it metabonomics, whilst discussing metabolic changes in response to xenobiotic exposures (36). It must be noted that Nicholson's metabonomics has a slightly different definition to the metabolomics studies

of today. Metabonomics aims to capture the global, dynamic metabolic changes in response to biological or genetic stimuli thereby measuring systemic changes as a function of time (37,38), whilst metabolomics measures all metabolites present in a complex biological sample, reflecting on what was happening within that system at the point of sample collection. This makes these types of studies a snapshot in time, observing metabolism only at the selected point of sample collection (8,37,38). As the field has developed, these distinctions have faded. Indeed today, the term metabonomics is seldom used.

The development of the field of metabolomics was not restricted to academic research. Technological advancements in the 70s and late 90s propelled it even further (technologies used in metabolomics are discussed in section 1.5). Indeed, as will be discussed in chapter 6 of this thesis, technological advancements have played such a crucial role that more are currently required to further improve metabolomics.

Living organisms are the culmination of multiple, often simultaneous, biochemical processes, including metabolic reactions, operating within cells, tissues, organs, or whole organisms. The processes involve different characteristics, divided into genotype (genetic characteristics) and resultant phenotype (physical characteristics). Specific genotypic characteristics result in corresponding phenotypic expression. Simply put, genes control what *can* happen in a cell by switching on and off (39,40). Genetic information is inherited and fixed, i.e., one cannot eat more genes, and the study of all genetic material (genome) within an organism is called genomics. The transcripts resulting from switched on genes encode information that facilitates the synthesis of proteins. The study of all transcripts (transcriptome) in a living organism is called transcriptomics and indicates what *is* happening within a cell at a genetic level. Transcribed information directs the production of proteins, which bind to a variety of substrates and/or receptors, have a wide range of functions, and *facilitate* what is happening within a cell. The collective study of these proteins (proteome) is called proteomics, and indirectly reflects on the genetic composition of an organism. Since the transcriptome and proteome are governed by the genome, they too are fixed (alas, one can indeed eat more proteins, but consumed proteins cannot bind to substrates and/or receptors and effect biological processes). The genome, transcriptome and proteome's primary purposes are cellular regulation, growth, and reproduction. Metabolites have a unique position in a biological system. Not only can they be influenced by the transcriptome and proteome,

reflecting on what *has* happened within a cell, they also interact with the outside environment. The metabolome can therefore inform on both the biological organism's internal and external environments. It is the downstream response effected by upstream changes (genome, transcriptome, and proteome), whilst simultaneously interacting with the environment (41). The human metabolome is therefore comprised of endogenous metabolites such as glucose and glutamate, and exogenous metabolites derived from food, environmental exposures, and pharmaceutical components.

Molecular information does not always travel down a one-way system towards the metabolome. Indeed, the metabolome can convey molecular information up-stream as well. For example, methylation of DNA (epigenetics) is facilitated by the metabolome (S-adenyl methionine donates a methyl group with the help of methyltransferase enzymes) (42) and regulates gene expression (43). The functions of endogenous metabolites are therefore not just limited to metabolism, but also include cell signalling (44,45), regulation (46,47), control of gene expression (48), structure formation (e.g., cell walls) (49), stimulatory and inhibitory effects on enzymes (50), fuel (51), catalytic activity (52), defence (53), and interactions with other organisms (e.g., pheromones) (54). The changes in metabolite concentrations, influenced by anabolism and catabolism (55), for all the various roles listed and more are often referred to as the metabolic phenotype, with changes in the genome, transcriptome, and proteome effecting physical (phenotypic) responses that are the metabolome and can be measured. Moreover, the interaction of these various parts of a biochemical system with the environment also results in a measurable metabolic phenotype (Figure 1).

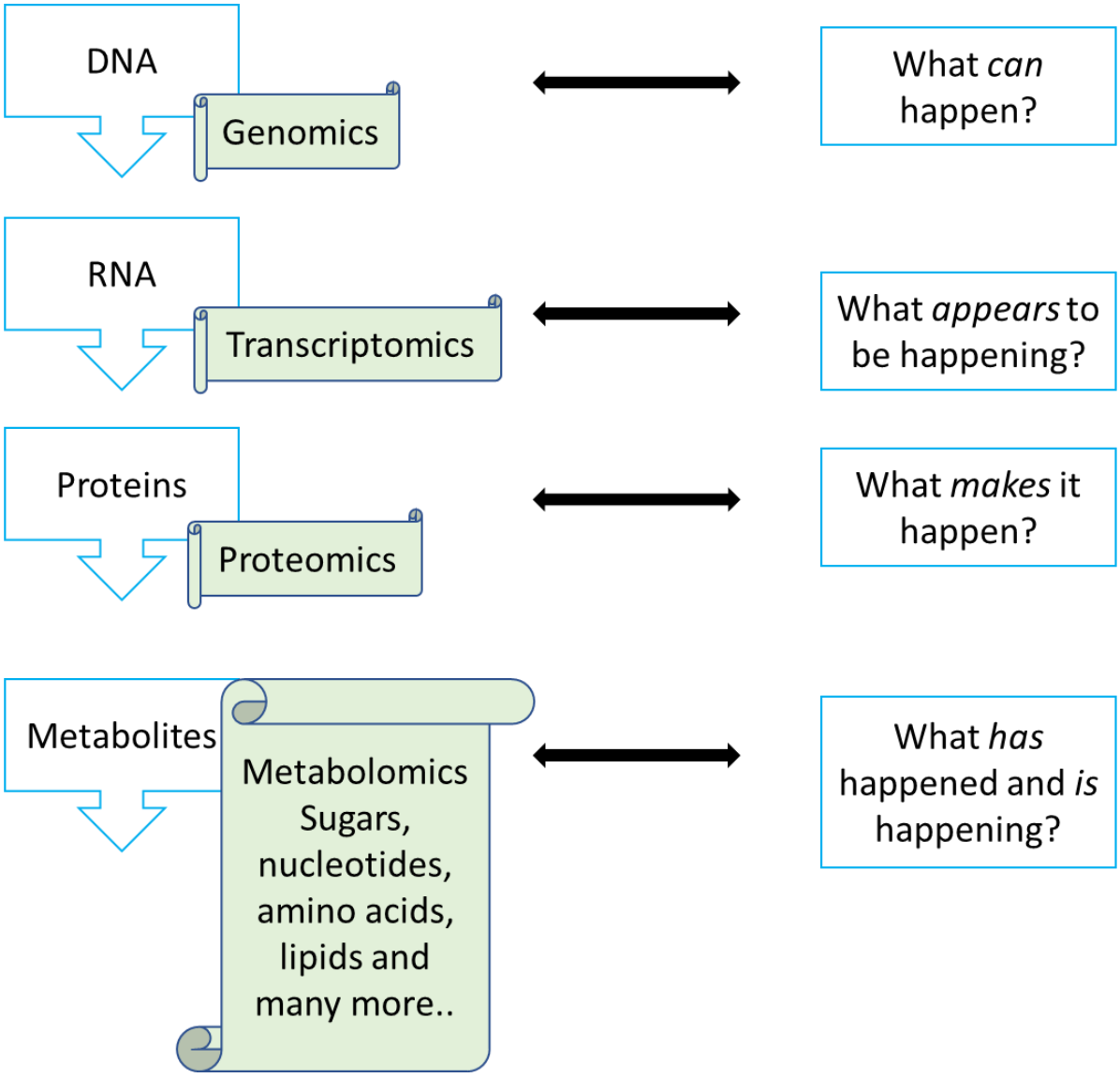


Figure 1: Adaptation of a figure (41) showing the relationship between the genome, transcriptome, proteome, and metabolome. The genome governs what can happen in a cell, the transcriptome results in RNA, which carries encoded information about what proteins to produce. The proteome effects genetic information, and the metabolome is influenced by these up-stream processes whilst also interacting with the outside environment. This is known as the "omics cascade".

1.2.1 Types of Metabolomics Methods

Metabolites have a wide variety of chemical classes and concentrations that span across several orders of magnitudes of concentrations (56). To fully characterise the metabolome, multiple analytical strategies for measuring different chemical classes are often employed, whilst metabolomics studies themselves are often split into two types of assays: targeted and non-targeted assays.

1.2.1.1 Targeted Assays

Targeted analyses (TA) are analytical techniques used to measure a small number of analytes, usually 1-100 analytes, use authentic chemical standards against which to reference and calculate absolute analyte concentrations, and offer high levels of confidence in metabolite identification. Targeted metabolomics analyses often look at specific classes of metabolites with *a priori* knowledge of the system response*,* and by using internal standards and calibration curves offer quantitative (exact concentrations) or semi-quantitative (estimated concentrations) measurements. Targeted metabolomics takes advantage of knowledge of biochemical pathways to which metabolites belong, along with an understanding about enzymatic processes, their end products and enzyme kinetics (56). These types of analyses have many advantages. The use of authentic chemical standards offers the highest level of confidence in identification possible (see section 1.6.4). Additionally, the small number of analytes means sample preparation and analysis methods can be optimised to reduce high abundance analytes dominating results, whilst improving sensitivity for the desired analytes. Data processing is significantly less complicated as targeted metabolomics datasets contain less information (chromatographic and spectral where U(H)PLC-MS analyses are carried out, (see 1.5.3.1) and are therefore smaller (56). However, there are also some disadvantages to targeted metabolomics. Firstly, it requires some prior knowledge of which metabolites or metabolite classes or pathways to target. This is not always possible, particularly for species or biological questions that have not been studied before. It also requires authentic chemical standards, which are both costly and not available for many endogenous metabolites. Method development/optimisation for TA requires expertise and can be time consuming; but several guidelines define how (62,63), (57–60). Finally, TA misses out on vast amounts of information about other metabolites not measured. This is particularly important where a metabolic pathway perturbation has been hypothesised. Targeted metabolomics studies still play a vital role in the field. They are often used to validate hypotheses generated from non-targeted metabolomics studies and targeted biochemical assays offer a route to translation of concepts in to practice such as biomarker panels for disease diagnosis (61,62). There is a myriad of successful targeted metabolomics assays, including targeted metabolomics to discern correlation between the metabolome and nutrition (63,64), identification of phenolics in fruits and beverages (65,66), and looking for brain and blood metabolite signatures of pathology and progression in Alzheimer's disease (67,68).

1.2.1.2 Non-targeted Assays (NTA)

Non-targeted metabolomics assays, or non-targeted analyses (NTA) potentially detect all metabolites in a sample above the assay's limit of detection, do not routinely or extensively use authentic chemical standards or internal standards and so only offer relative quantification of metabolites. Relative quantification here means the area of the observed signals of each metabolite is used together with statistical analyses to inform on concentration variations between metabolites and different study samples applying the assumption that concentration is correlated to the observed signal (peak height or peak area). These studies are used for hypothesis generation. Method development is less time consuming, and once established less expertise is required to apply these methods. They also cost less as authentic chemical standards are not required, and they offer a large volume of information about each analysed sample. However, these methods yield less confidence in metabolite annotation as authentic chemical standards are not used to reference against, whilst annotation is itself challenged by the vast number of similar chemical signals that can be observed (see 1.6.4). Indeed, data handling is difficult as the files are often extremely large, require conversion prior to data analysis and can take a few months for data interpretation. Results from a non-targeted metabolomics method must be validated using a targeted metabolomics method for greater confidence. A review of validated biomarkers discovered through metabolomics was carried out by López-López and co-workers (69).

A summary comparison of targeted and non-targeted metabolomics studies is shown in Table 1.

Targeted Metabolomics	Non-targeted Metabolomics
Hypothesis-testing approach	Hypothesis-generation approach
Identity of metabolites known *a priori*	Identity of metabolites unknown *a priori*
Absolute quantification	Relative quantification
Low number of metabolites studied (1-100)	Large number of metabolites studied (100-1, 000s)
Validation not required, or used as validation approach	Validation required
PROS	**PROS**
Absolute quantification	Comprehensive analysis
Higher accuracy, sensitivity, and selectivity	Useful for biomarker discovery
Easier data interpretation	Higher throughput
CONS	**CONS**
Not comprehensive	Semi or relative quantification
Lower throughput	Difficult data interpretation and large unknowns

Table 1: Summary of targeted vs non-targeted metabolomics studies, adapted from (70).

1.2.2 Advantages and Applications of Metabolomics

1.2.2.1 Advantages of Metabolomics

Metabolomics is the newest of the -omics approaches but carries significant advantages over the others. As described in section 1.2, the measurement of the metabolome informs directly on function, being itself the observed changes, where changes in the genome and proteome do not necessarily reflect cellular changes. Moreover, genes and proteins can vary greatly across multiple organisms and species. This makes method development and transfer from species to species much more challenging (71). Metabolites are more global across various species, albeit spanning many chemical classes. This means class specific methods can be developed and transferred for the metabolomics analysis of different species with ease.

At present, it appears that there are fewer endogenous metabolites than genes and proteins, although the full size of the metabolome is yet to be comprehensively measured. This makes data complexity smaller and increases the likelihood of detecting meaningful changes. This

by no means suggests that metabolomic datasets are not complex but rather draws a comparison with proteomics and genomics databases based on currently available knowledge. It is important to note that to date, many metabolites have not been detected and identified, and this comparison is therefore limited (Table 2). Lastly, metabolomics analyses are at the current time cheaper and quicker than genomics, transcriptomics and proteomics analyses (72).

DATABASE	Genecode	Ensembl	RefSeq	CHESS
Protein-coding genes	19,901	20,376	20,345	21,306
IncRNA genes	15,779	14,720	17,712	18,484
Antisense RNA	5,501		28	2,694
Miscellaneous RNA	2,213	2,222	13,899	4,347
Pseudogenes	14,723	1,740	15,952	
Total transcripts	203,835	203,903	154,484	323,827

Table 2: Table taken from literature (73), showing the number of human genes within various proteomics/transcriptomics databases. The total number of genes and transcripts currently known is much larger than the largest repository of human metabolites for metabolomics experiments. This currently contains ~114 000 metabolites, a value regularly updated but still smaller than the number of genes and transcripts (74,75).

1.2.2.2 Applications of Metabolomics

Metabolomics applications span many fields. Clinical metabolomics is focused on human health studies, specifically human disease. The most notable application in clinical metabolomics is biomarker discovery. A look at publication trends in PubMed over the last 20 years (Figure 2) shows great increases in metabolomics studies investigating biomarkers (76).

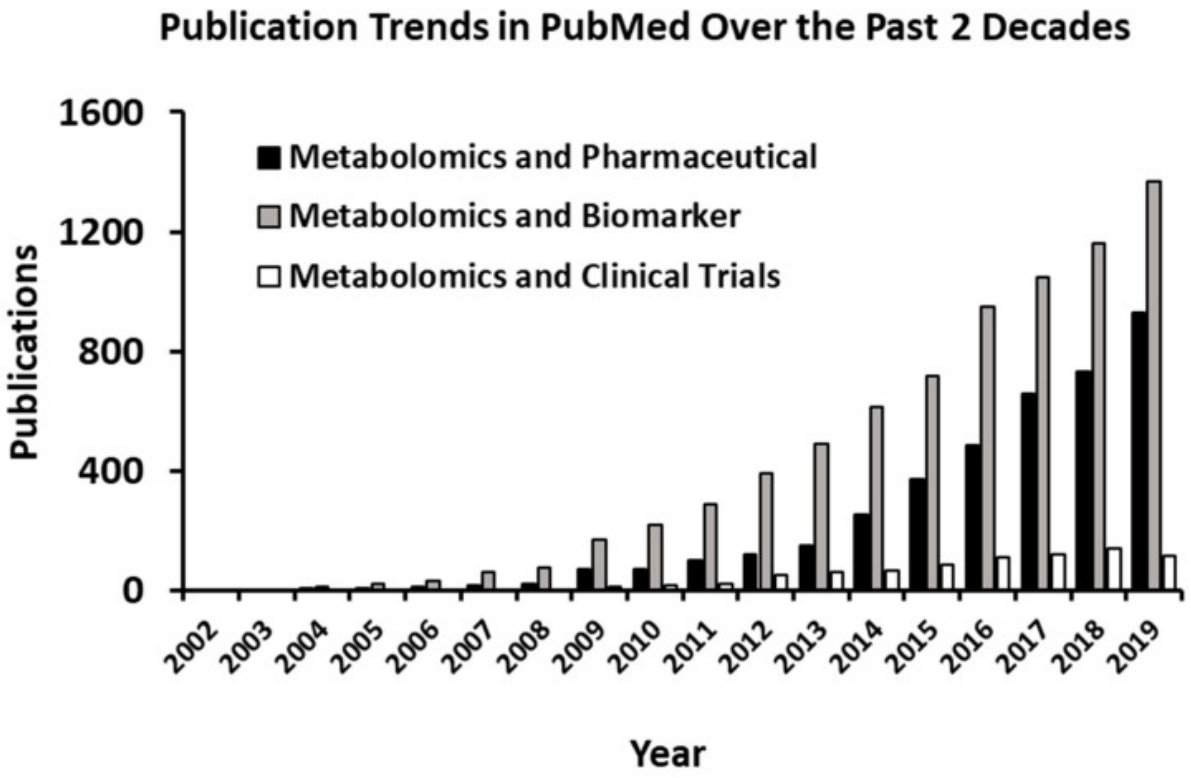


Figure 2: Bar plot presented in literature (76) showing an increase in research papers with keywords 'metabolomics' and 'biomarkers'.

Metabolites possess all the right characteristics to be ideal biomarkers. For example, they are readily available in blood or urine, and have high sensitivity facilitating early detection (77) thus biomarkers for inborn errors of metabolism (78), insulin resistance (79), various cancers (80,81), and cardiovascular diseases (82,83) were found using metabolomics approaches mining for biomarkers. Not only can biomarkers aid diagnosis, evaluating metabolites can improve understanding of cellular mechanisms of action (pathophysiological processes) and allow monitoring of disease progression. Knowing what metabolic changes to expect at disease onset and during progression potentially introduces new targets for disease intervention, paving way for precision medicine. Moreover, biomarkers for drug response can be sought, allowing drug response monitoring (pharmaco-metabolomics). Metabolomics studies can even allow classification of an individual's metabolic profile (metabotype) based on the individual's genotype, merging the fields of metabolomics and genomics (84). For example, precision medicine using metabolomics has been applied to characterise molecular subtypes of various cancers (85–87), whilst cancer related pharmaco-metabolomics has also

been demonstrated (88,89). Other studies have shown the successful application of metabolomics to drug-response testing and monitoring (90).

Metabolomics studies have been used in conjunction with genomics to further understand and identify genes responsible for phenotypic changes (91,92). Several other studies have also investigated the relationship between the genome and the metabolome (93–99). This type of analysis has even allowed testing of genetic modifications through monitoring the metabolome phenotype (100). Nutrition has increasingly become a topical research area, and the merger between it and metabolomics has begun. For example, nutrition related exogenous and resulting endogenous metabolites have been studied by various researchers (101–104). Metabolomics is a versatile field, continuously adapted for various applications. These applications are not solely focused on human health and a clinical setting. Indeed, they cross over into environmental health.

Metabolomics can be applied in toxicology. Classic toxicology tests have been successful at correctly determining the target organs of a toxin through easy observation of clearly defined endpoints such as mortality, clinical symptoms, or examination of tissue damage (histopathology) (105). These approaches are well established, with toxicological studies carried out for selenium (106,107), aluminium (108,109), gallates (110,111), styrene (112), and aflatoxins (113) to name but a few. Although these approaches correctly point to a toxin's target organs, a multitude of off-target effects are not observed. The use of metabolomics in toxicology gives a more comprehensive understanding of chemical toxicity, not only suggesting target organs through perturbed metabolic pathways without the need to analyse each (organ) individually, but also other impacted biochemical pathways. Biomarkers of toxicity can and have been discovered in this way (114–117), although data interpretation is often complicated by common, non-specific pathway perturbations (118–120).

The implementation of metabolomics in toxicological studies has helped to develop another sub-branch of metabolomics known as environmental metabolomics. Environmental metabolomics "applies metabolomics to areas such as ecophysiology-with a focus on understanding the underpinning biochemical responses of organisms to abiotic and biotic stressors in their environment; and to ecotoxicology and ecological risk assessment, which typically involves chemical toxicity testing within a controlled laboratory environment to inform upon potential risks of chemicals to organisms in the natural environment" (121).

Simply put, environmental metabolomics studies interrogate interactions between organisms and the environment with an end-goal of evaluating organism functionality at the molecular level. This information ultimately gives insight into environmental factors such as levels of chemical pollution (122). Environmental metabolomics studies have successfully been demonstrated with various applications (123–128).

1.3 Chemical Pollution: A Global Challenge

The importance of metabolomics studies in fields such as toxicology and environmental health studies can be demonstrated through discussing the wider challenge of chemical pollution in the environment. The scale of chemical pollution and the resultant effect on the environment is sobering and requires immediate circumvention. A study by the World Health Organisation (WHO) highlighted 4.9 million deaths attributable to environmental exposures during the year of 2004 (129), rising to ~8 million deaths in 2021 (130,131). There were an estimated 140,000 chemicals on the EU market (132) potentially being released into the environment, with grossly insufficient information on the hazards they pose on environmental and human health. However, new initiatives are in place to register chemicals, and currently ~22000 chemicals are registered on the EU market (133). Environmental ecosystems are threatened by long term exposures to low or sub-lethal concentrations of single chemicals, or more realistically, mixtures of chemicals. For example, approximately 90% of water and fish samples from aquatic environments contain pesticide contamination (134,135), even though the adverse effects of pesticides on multiple organisms have been demonstrated in *Daphnia magna (D. magna)* (136), macro-fauna populations in field ditches (137), and *Clarias batrachus* (138). Other common chemical contaminants include persistent organic chemicals (POCs), which, due to their stability and inertness have found widespread uses in many industries (139). These have been shown to have adverse effects on zebra fish and other aquatic organisms (140,141). Nanoparticle pollution has also emerged recently, increasingly used in consumer products and medicines, and is posing a real threat to environmental health, as demonstrated on a study using *D. magna* (142).

This global problem poses a significant human and environmental threat and requires immediate action, focussing on the characterisation of as much environmental contamination as possible, and discerning toxic effects. Chemicals released into the environment from

various sources such as chemical waste from the pharmaceutical and chemical manufacturing industries must be carefully controlled through the introduction of and adherence to various government regulations and legislations. These should require chemical manufacturers, packagers, and distributors to provide evidence of non-toxicity at given chemical levels. Provision of this evidence undoubtedly requires toxicity studies to be carried out; for which there are already well-established methods (143,144). Indeed, these methods are so well established that they have not changed much in the last 50 to 80 years (145). Despite their popularity for many years, these toxicity testing approaches do come with caveats. They often involve the usage of animal testing and require high chemical dosages, presenting ethical challenges. They are costly and time-consuming, whilst the extrapolation of toxicity from high to low, environmentally relevant doses is applied. More efficient approaches are required to attempt to eliminate or minimise these challenges. It is important to know which chemicals have drained into the environment and at what concentrations. To date most methods investigating environmental chemicals have been targeted, looking for known analytes or groups of analytes such as aromatic compounds within environmental samples (146). Non-targeted methods are an attractive alternative as they provide a more comprehensive picture of environmental contamination by allowing for the identification of as many contaminants as possible. However, most non-targeted methods have been developed for "omics" fields looking at biological samples rather than environmental. Consequently, the chemical coverage of these methods for environmental samples has gone unexplored. Knowing this information could aid in improving methods to allow for greater selectivity, sensitivity, and coverage; and understanding where gaps in environmental contaminant knowledge may arise.

1.3.1 Governing Organisations and Corrective Legislations

Many organisations have spawned globally to control chemical pollution, with many legislations being introduced over the years. As far back as 1973, Japan introduced a national protection law called the 'Chemical Substances Control Law' which is concerned with the evaluation and regulation of chemical substances that pose a risk to human health and the environment (147). The Great Lakes Water Quality Agreement was introduced five years later in Canada and the USA. Its focus is the quality of surface waters, aiming for restoration and

preservation of the Great Lakes waters (148). In 1998, the UNECE persistent organic pollutants (POPs) Protocol was launched. This focuses on 16 substances singled out due to risk criteria. 11 of these 16 compounds are pesticides, and the protocol bans the use of some of these chemicals completely (149). Additionally, the Stockholm convention was launched with objectives to protect human and environmental health from "pollutants that remain intact in the environment for long periods of time" (150). Also in 1998, the US Environmental Protection Agency (US-EPA) introduced the Toxic Substances Control Act which regulates the introduction of new or already existing chemicals (151). In 2006, REACH was founded under the European Commission enforcing registration, evaluation, authorisation, and restriction of chemicals. This legislation places responsibility on industry to provide safety information relating to chemicals and to manage associated risks, with the information required to be entered into a central database in the European Chemicals Agency (ECHA) in Helsinki (133). Preceding this, increased public demand for cleaner rivers and lakes yielded the EU Water Framework Directive issued by the European Commission, which aims to evaluate water bodies on chemical pollution and ecological preservation (152). These are an almost insignificant few examples of the range of legislations implemented to monitor and improve environmental health globally.

1.4 Exposomics

Despite chemical pollution control efforts through regulation and legislation (section 1.3.1), the advanced field of toxicology, the application of metabolomics to toxicological studies and indeed environmental metabolomics studies, the wildfire that is chemical pollution is far from under control. Most chemical risk assessments and toxicity testing have considered single chemicals one at a time (153). The European Union's Joint Research Centre (JRC) stated that environmental quality standards (EQS) are not satisfactorily protective against chemical mixtures (154). Moreover, "challenges at the European scale include the identification, assessment and management of the riskiest chemicals and mixtures and the prediction of trends in chemicals expected due to societal-change processes, such as increased population size and ageing" (155). The European Food Safety Authority (EFSA) also acknowledge the more realistic scenario in chemical exposure and potential toxicity. Humans and other species in the global environment are not exposed to single chemicals at a time but rather a mixture

of chemicals. They recommend identifying priority mixtures by considering not only chemical toxicity but also exposure or anticipated exposure risk to these chemicals. It is also recommended that where toxicity data is not available, chemicals are pooled into classes and evidence compiled to show that similar chemicals cause similar adverse effects on organisms. This creates "so-called assessment groups" which can be utilised for combined chemical toxic effect predictions, where chemicals can act through similar modes of toxicity allowing doses to be added together to predict effect (dose addition), interact with one another to cause a more toxic combined effect (synergism), or interact with one another to give a less toxic effect (antagonism). EFSA also states the need for a better understanding of how chemicals interact within an organism and the implications of these interactions. Finally, EFSA calls for continued and better use of mathematical and biological models for predicting chemical elimination from organisms and toxicity modes of action (156).

Metabolomics methods can be used to measure simultaneous effects of chemical mixtures on biological systems. Indeed, temporal metabolomics studies can inform on dynamic metabolic responses to longer term exposures to mixtures of chemicals. As such, the 'exposome', first defined by Christopher Wild in 2005 as the "totality of an organism's life-course environmental exposures (including life-style factors) from conception to death" was coined (24). To fully characterise it, both the exposures and the changes they induce in an organism over time, if any, must be measured. Under this definition, Wild further divided the exposome into three broad categories. The internal exposome, including processes such as metabolism, specific external exposome including various agents such as ionisation radiation, and general external exposome including wider psychological influences (157). Most exposome-related research has thus far centred around chemical pollution (external exposome) and human health (internal exposome, including the metabolome) (158–164), although not without challenges. It requires the collation of large amounts of data from wildly variable populations, whilst lifetime human exposure impacts are unfeasible to measure. Proposed solutions have included use of already available samples that have been collected and stored appropriately, but this raises questions about sample quality, quantity, availability, and identification of unknown compounds (165).

Exposomics has the potential for great utility in environmental health assessment. It facilitates the measurement of chemicals in a variety of ecological environments. This is in

line with traditional environmental monitoring techniques which employed targeted analytical analyses to measure external exposures such as chemical pollution (166). However, exposomics studies not only measure chemicals in the environment (external exposome) but also attempt to measure both chemical and non-chemical exposures plus their effect on biological systems. A case study can be used to demonstrate application and advantages. The honeybee (*Apis mellifera*) lends an irreplaceable role to the environment as a pollinator worth about 200 billion US dollars globally (167). Unexplained losses in bee colonies have been observed (168,169). The "in-hive pesticide exposome" was measured to evaluate its potential role in colony losses. In a top-down approach, pesticide exposure was measured through-out the bee-keeping season, bee health monitored, and potential relationships mined for between exposure and colony well-being. In this instance, pesticide burden was summarised using three mathematical models which illustrated a simplified risk to pesticide exposure (169). By contrast, a bottom-up approach would focus on individual pre-defined categories of external exposures to give a quantitative measure of contaminants that can be summed up to give individual external exposomes. Disadvantages with this method lie in the thousands of unknown analytes to be annotated. This approach also neglects components of the internal exposome. A top-down approach therefore offers the greatest value in attempting to measure the whole exposome, encompassing both endogenous and exogenous exposures (170). Rappaport discusses the advantages of top-down and bottom-up approaches; stating that a top-down approach would find value in "non-targeted omics methods to measure features of exposures in biological fluids" and offers benefits in the form of hypothesis generation to identify specific exposures (170). To better adapt the concept of the exposome to environmental studies, the National Research Council (NRC) of the US National Academy of Sciences defined an "eco-exposome" as "the extension of exposure science from the point of contact between a stressor and receptor inward into the organism and outward to the general environment, including the ecosphere" (171). The eco-exposome is characterised by both internal and external exposures, although the way in which the exposome is implemented would undoubtedly vary from organism to organism (172).

1.5 Analytical Technologies Applied in Metabolomics and Exposomics

Comprehensive chemical risk/exposure assessment, metabolomics, and exposomics studies require a means by which to measure and, where possible, quantify environmental chemicals, the metabolome, and the exposome respectively, in a rugged manner. The measurement of environmental chemicals has been achieved through the application of TA (described in 1.2.1.1). Suspect screening (SS) and non-targeted analyses (NTA) allow interrogation of samples containing suspected or unknown analytes of interest, respectively. Advantageously more information can be accrued through these strategies, showing a more comprehensive picture of chemical contamination as described in 1.2.1.2.

SS and NTA are applied through a wide range of analytical instruments. To measure the full spectrum of chemical classes requires powerful instrumentation that offers sensitivity to detect low intensity chemicals, selectivity to target desired analytes without detrimental interference from the matrix, high resolution to differentiate analytes of interest whose signals are close in value, signal accuracy, low signal background noise, cost effectiveness and user friendliness.

The most applied analytical techniques in metabolomics are nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry (MS). The former is not described here as this technique is not directly relevant to work presented in this thesis. Indeed, mass spectrometry is by far the most widely applied technique in metabolomics studies, with more than 80% of metabolomics publications to date using this technique (173).

1.5.1 Mass Spectrometry (MS)

Mass spectrometry (MS) is an analytical technique that measures samples with an instrument called a mass spectrometer (also abbreviated to MS). It circumvents the challenges faced in NMR spectroscopy offering much higher resolution and sensitivity which provides detection of a larger number of analytes. Using MS analyses, structural elucidation can be performed, although this requires the use of authentic chemical standards which are not always available and can be expensive. There are various types of MS instruments, each offering advantages and disadvantages. These will be discussed in the following sections, with a focus on Orbitrap instruments described in 1.5.1.3.1 which have been used in the research presented in this thesis.

MS separates and detects compounds based on their molecular mass to charge ratio (*m/z*). Analytes must be gaseous and carry an ionic positive or negative charge. Every MS has five main components to it. At the front-end is a sample inlet, facilitating sample introduction to the MS. An ionisation source then converts samples to the gaseous phase if they are not already, and form ions of each analyte within the sample. Generally, the sample inlet and ionisation source are located outside the MS vacuum, although this is dependent on the specific MS setup. Once ionised, samples are accelerated or transmitted through the MS towards a mass analyser, which operates under vacuum to avoid sample interaction with air molecules like oxygen and nitrogen. This is where separation by mass, in the form of a mass-to-charge (*m/z*) ratio, occurs. Generally, smaller, lighter molecules can travel faster than bigger, heavier molecules, arriving first at the last component of the MS, the detector. This is a general premise for how separation based on *m/z* ratio is achieved, with each different type of MS using these general rules but having slightly different designs. Finally, all MS must be attached to a data recorder, often a computer, which stores the data collected by the MS in the form of a mass spectrum and normally controls the instrument operation. A diagrammatic summary of the main components of a MS is shown in Figure 3.

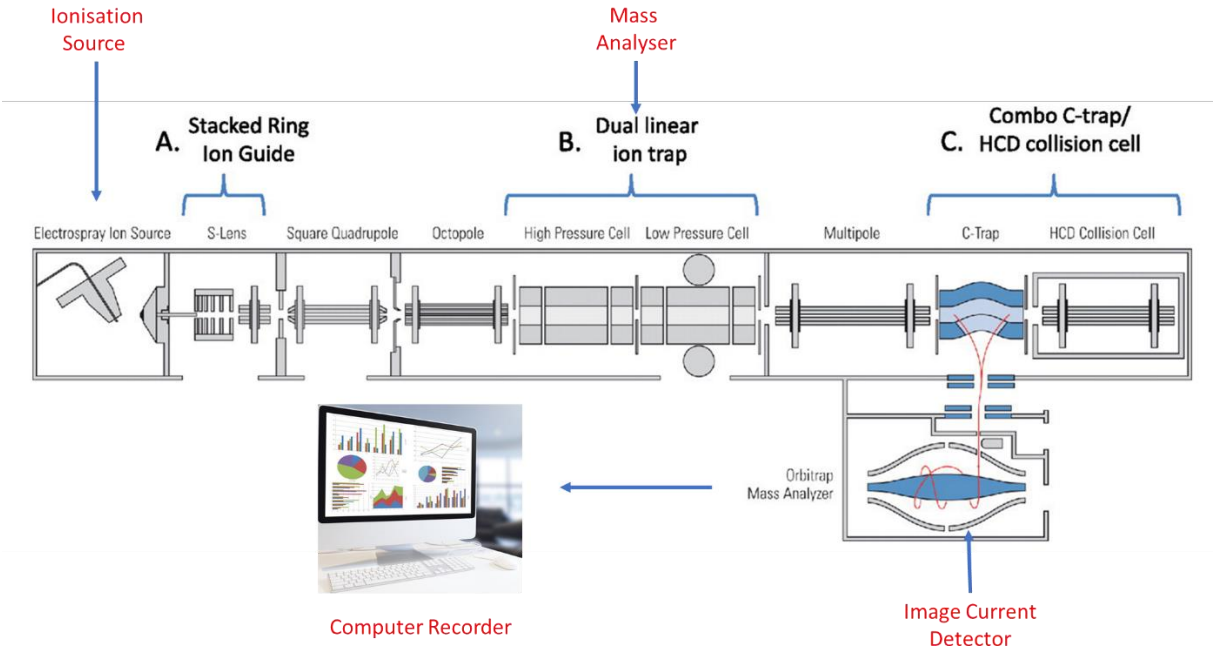


Figure 3: Figure summarising the main components of a mass spectrometer, taken from (174). Sample arrives at the ionisation source from a sample inlet. Ions are transferred into the MS and filtered at the mass analyser. Finally, they are detected at the detector, and a data recorder such as a computer records the data in the form of a mass spectrum.

1.5.1.1 Sample Inlets

There are many ways samples can be introduced into mass spectrometers. Most MS have a syringe pump and liquid samples can simply be pumped into the MS from a syringe at a user defined flow rate. This is usually a single syringe requiring manual sample loading and is not a feasible approach for high-throughput metabolomics and exposomics studies, which often have hundreds of biological and/or xenobiotic samples. For such large studies, an automated sample introduction system is more appropriate. Liquid chromatograph autosamplers or indeed autosamplers of any MS compatible separation technique (see sections 1.5.3, 1.5.3.1, and 1.5.3.1.1, 1.5.3.1.1.2) can be used as sample inlets. In such a scenario, they bypass their separation capabilities (175) and directly introduce many samples into the MS in an automated approach. This is called direct infusion mass spectrometry (DIMS). Samples are automatically introduced for analysis adhering to user defined analysis sequences. N.B. all these separation techniques can also be used for sample pre-separation prior to detection by MS.

The user defined flow rate is an important sample inlet parameter. To vaporise analytes in solution, the sample flowing from the sample inlet into the ionisation source forms droplets (see 1.5.1.2.1.1). The size of these droplets varies in size depending on the width of the spray needle orifice through which they travel to arrive at the ionisation source, an applied electrical voltage, and the flow rate. The higher the flow rate, the bigger the size of the droplets of sample formed. Typical metabolomics and exposomics studies use flow rates in the µL/min range, termed microflow analysis. One study showed that nL/min flow rates can offer increased sensitivity and reduce ionisation competition (ion suppression) between analytes, whilst also using minimal sample volumes, recommending that these benefits were more pronounced at flow rates <50nL/min (176). Studies involving flow rates as low as this are termed nanoflow analyses. These nanoflow analyses do have their caveats. The improved ionisation efficiency awarded by the low flow rates can result in mass spectrometric saturation, longer analysis times, and require higher pump pressures (177). Both micro and nano-flow analysis types will be discussed in detail (section 1.5.1.2.1.1).

1.5.1.2 Ionisation Sources

Sample inlets flow samples into the ionisation source at micro or nano flow rates. There are different types of ionisation methods. Conventional techniques include electron impact (EI) ionisation, also called electron ionisation or electron bombardment ionisation, which is the oldest ionisation technique in MS. In the EI source, electrons are produced by heating a wire filament whilst passing an electric current through it. The resultant electrons are then accelerated towards a sample, usually entering the EI source at a perpendicular angle to the fired electrons. The electrons are typically accelerated to 70eV as a standard for use in all gas chromatography-mass spectrometry (GC-MS) systems globally (described in 1.5.3). The energetic electrons fired at the sample and the low pressure in the EI source destabilise the electric field around analytes of interest. If the analyte's ionisation energy is surpassed, an electron is lost from the analyte, resulting in ionisation, whilst the analyte's covalent bonds are simultaneously broken, resulting in fragmentation (178). The fragmentation of analytes is beneficial as MS fragments can be used to aid structural elucidation. However, particularly labile analytes could be obliterated beyond useful information using EI, whilst the detection mass range offered by this ionisation technique is also limited to small molecules <1,000 Da since higher molecular weight analytes often have higher boiling points, and thus higher vapour pressures. Chemical ionisation (CI) was popularised as a circumvention to the "hard" ionisation of EI, offering a "soft" ionisation approach with minimal fragmentation (179). This technique works in an electron impact ionisation source through inducing chemical reactions between the analytes and some ionising reactants such as methane, thereby preserving the structural integrity of the analyte (180). Both EI and CI operate inside the MS, under vacuum. This makes it challenging when analysing liquid samples to combine other analytical techniques, such as separation, prior to MS analysis, as these other techniques often work at atmospheric pressure. Indeed, these challenges are so vast that conventional application excluded biomolecules and many other organic compounds, either due to the mass detection limitation or because analysis of these compounds was more optimal with separation of analytes prior to MS detection.

1.5.1.2.1 Atmospheric Pressure Ionisation Sources (APIs)

MS instruments operate under high vacuum pressures to eliminate interaction of the analytes with potential contaminants such as oxygen and nitrogen. Traditional ionisation methods such as electron impact (EI) and chemical ionisation (CI) also operate under high vacuum pressures. This has made it extremely challenging to couple MS analyses with any liquid introduction systems, described in 1.5.1.2. The development of atmospheric pressure ionisation (APIs) techniques has made it possible to apply MS analysis together with other analytical methods such as liquid chromatography, introduced in 1.5.3.1. This expands MS application to life sciences, not only facilitating coupling of MS to separation techniques but also significantly increasing the mass range of mass spectrometers to allow detection of high molecular weight molecules including proteins.

API sources include atmospheric pressure photo-ionisation (APPI), atmospheric pressure chemical ionisation (APCI), desorption electrospray ionisation (DESI), matrix-assisted laser desorption ionisation (MALDI), direct analysis in real time (DART), liquid extraction surface analysis LESA), secondary ion mass spectrometry (SIMS) and desorption/ionisation on silicone (DIOS). Each of these techniques has advantages and disadvantages. A summary comparison of these ionisation techniques is given in Table 3, and ESI, which has been applied in the work presented in this thesis, will be described in section 1.5.1.2.1.1.

Ionization Technique	Maximum Mass (Da)	Degradation	Complex mixtures	Sensitivity	Analyte type	Disadvantage
(Nano)ESI	70,000	None	OK but low flow rates can present problem	High zeptomole to low femtomole	Polar Analytes	Low salt tolerance
ESI	70,000	None	Excellent	High femtomole to picomole	Polar analytes	Low salt tolerance
MALDI	300,000	Photo degradation and matrix reactions	Possible	Low to high femtomole	Biomolecules, fragile molecules	Thermal degradation
DESI	2,000	None/minimal	Not possible	Attomole	Whole cells, tissues, or organs	Matrix background can be problematic for low molecular weight analytes
DART	1,200	Thermal degradation	Excellent	Attomole	Whole cells, tissues, or organs. Polar and non-polar analytes	Lower spatial resolution
EI	500	Thermal degradation	Very limited	Picomole	Low molecular weight	Hard ionisation technique, excessive
CI	500	Thermal degradation	Very limited	Picomole	Low molecular weight	Limited to volatile analytes and yields less fragments
SIMS	1,000	None/minimal	Possible	>Attomole	Whole cells, tissues, or organs. Polar and non-	Lack of quantitative data

Table 3: Table summarising different ionisation techniques, detailing advantages, and disadvantages of each. This table details both atmospheric pressure and vacuum ionisation techniques. Taken and adapted from literature (179).

1.5.1.2.1.1 Electrospray Ionisation (ESI) and Nano Electrospray Ionisation (nano)ESI

Electrospray ionisation (ESI) (nano or micro-flow) is the gold standard for metabolomics and exposomics analyses, adapted for MS by Nobel Prize winners Masamichi Yamashita and John B. Fenn in 1984 (181). The ESI source functions as an electrochemical cell. It can be divided into four processes, formation of ions in solution, droplet formation, droplet shrinkage, and gaseous ion formation (182). Dissolving solid compounds into an aqueous liquid, such as water, often leads to the disruption of covalent or ionic bonds by the water molecules. Consequently, compounds break (dissociate) into their smaller components such as atoms, ions, or radicals. Dissociation is often reversible. The reversible dissociation of a compound reaches an equilibrium (K_a) under the right conditions, with a balance reached between the concentration ratio of dissociated components to the undissociated compound (183). K_a is called the dissociation constant (Equation 1). Samples analysed using ESI are always prepared in solution, thus sample components undergo dissociation, whilst each individual analyte has its own K_a.

$$K\,a\ = \frac{[A\ -][H\ 3O\ +]}{[HA][H\ 2O]}$$

Equation 1: The dissociation constant (K_a) governs the ratio of proton donation to an undissociated acid.

The strength of acidity is called pH (Equation 2). The larger K_a is, the greater the acid dissociation, thus the stronger the acidity of the solution (lower pH). As such, pH can be used to manipulate the dissociation of compounds to form ions. Increasing the protons available in solution, for example by adding an acid such as formic acid to the water solution, slows down the production the H3O+ ions of water molecules, whilst lowering the concentration of hydrogen ions will have the opposite effect. Therefore, pH can be used to encourage formation of ions in solution (183). For reproducibility of results, the pH must be maintained as constant as possible throughout the analysis. To achieve this, buffer solutions, which resist changes to pH due to the production or consumption of protons, are created (184).

$$pH = -\log [H+]$$

Equation 2: The pH is the negative logarithm (to base 10) of the hydrogen ion (proton) concentration.

Droplet formation is achieved through the dispersion of liquid (containing ions in solution) to form an aerosol. Aerosols are fine liquid or solid droplets suspended in a gas, and are formed when the liquid or solid material, whilst enclosed in a container under high pressure, is released by a propellant as a fine spray. ESI aerosol formation is achieved by applying a high electrical field to a sample in solution as it arrives at the tip of the capillary (a thin, hollow, metal tube used to introduce samples to the ionisation source). The ions in the solution are attracted to the electrical field applied at the tip of the capillary if the charges are the same. For example, positively charged ions in solution will be attracted to a high positive electrical field at the tip of the capillary. This interface between the liquid sample solution and the metal capillary behaves as an electrochemical cell. Here, more ions are formed through a series of charge-balancing reactions, such as $2H_2O \rightarrow 4H^+ + 4e^- + O_2$ (185). Other ions are formed due to acid-base reactions, occurring since most sample solutions are acidic. These ions and charged molecules simulate a current, moving in the gas phase, whilst electrons flow through the wires connecting the capillary to the MS (185). The capillary acts as the anode when carrying out analyses in positive ion-mode (analysing positively charged analytes), whilst the MS behaves as the cathode (185).

Accumulation of charged ions on the surface of the solution as it exits the capillary into the ionisation source causes a Taylor cone to form because of surface tension. The electrical field exerts pressure on the liquid, causing it to shrink and take up the minimum possible surface area (maximum surface tension) and the edges of the liquid to adopt a convex shape and a rounded tip. If the applied electrical field is high enough, the Taylor cone forms droplets of solution as the surface tension of the solution is exceeded. This occurs when the Taylor cone is elongated at the rounded tip. The droplets form through a process called budding, breaking off from the elongated end of the Taylor cone. The size of these droplets is influenced by the strength of the electrical field applied, the flow rate at which the solution is introduced, and the properties of the solution itself (182). The next phase is droplet shrinkage. The newly formed liquid droplets gradually decrease in diameter due to multiple processes. Heat can be

applied to help evaporate the solvent present in the droplets, whilst dry nitrogen is also often employed to aid this drying step. As the droplets decrease in diameter and solvent evaporates, they are prevented from losing excessive energy and therefore momentum through collisions with other droplets. The liquid droplets become so small that the combined charge carried by the ions in the droplet is greater than the surface tension holding the droplet together, and droplet fission (Coulombic explosion or fission) occurs. The maximum amount of charge required to overcome the surface tension of the droplet is called the Rayleigh limit (Equation 3)(182,186). This process repeats, with solvent evaporation and fission iteratively occurring until nanometre droplets are formed. These are the immediate precursors to gaseous ions, and these nano sized droplets form in both micro and nano ESI gaseous ion formation processes (187).

$$zR = \frac{8\pi}{e}\sqrt{\varepsilon 0}\gamma \mathrm{R3}$$

Equation 3: Rayleigh limit equation representing the relationship between the Rayleigh limit z_R, the radius of the droplet, vacuum permittivity, and surface tension. $_zR$ is the droplet radius, ε_0 is the vacuum permittivity, and γ is surface tension.

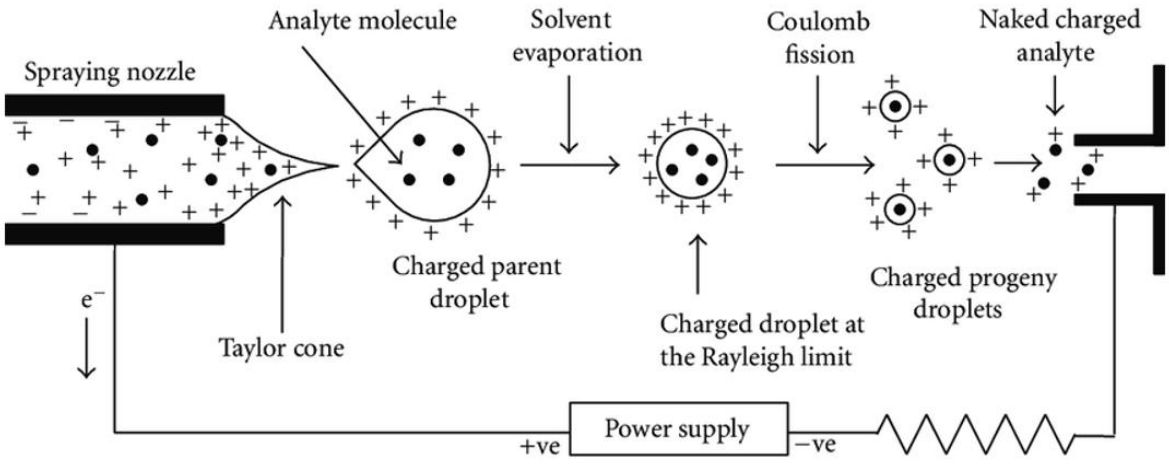


Figure 4: Schematic taken from (188) demonstrating the process of repeated solvent evaporation and droplet fission until the Rayleigh limit is reached, at which point small nano sized droplets are formed as immediate precursors to gaseous ion formation.

The final step is gaseous ion formation, and this step is not fully characterised due to difficulty in experimentally observing these small, gaseous ions as they form. Some mathematical

models attempt to describe the mechanism by which gaseous ions are formed, described below.

1.5.1.2.1.1.1 The Ion Evaporation-model (IEM)

The ion evaporation-model (IEM) of the last step of ESI, developed by Iribarne and Thomson in 1976 (189,190), occurs when the electric field of the charged droplets, usually <10nm in diameter, causes ejection of solvated ions from the surface of the droplet. This reaction process is limited by a principle called the method of images (191) which induces a pull on the ion as it attempts to escape, and on the repulsion of the ion by the surface-located charge that pushes it away (185). The rate at which the ion is ejected is mathematically represented (Equation 4).

$$k = \frac{kBT}{h} e - \frac{\triangle H}{RT} e \frac{\triangle S}{R}$$

Equation 4: The theory of absolute reaction rates equation, describing parameters governing the rate at which a gaseous ion is ejected from the solvent droplet during the last phase of ESI. k is the reaction rate constant, K_B is the Boltzman constant, h is Planck's constant, T is temperature, ΔH is the enthalpy of activation, and ΔS is the entropy of activation.

Ion ejection has also been simulated using molecular dynamics. As the ion leaves the droplet, it initially remains attached through "a sticky string of solvent molecules" (185,192,193) before fully detaching. This is known as bridging and is more prevalent in organic solvents which have less surface tension than aqueous ones. The ion evaporation process produces a single analyte ion with a cluster of solvent molecules around it. This "solvent shell" is lost as the ion travels towards the MS, facilitated by heat and dry nitrogen (185,192,193). The IEM is the preferred model for how gaseous ions of small molecules are formed.

ESI assumes that solvent droplets formed only contain one analyte ion, so there are advantages to the IEM. Firstly, the rate and efficiency of ionisation are not influenced by the analyte in question, but rather by the ability to successfully form small droplets, efficiently evaporate the liquid phase, and detach from solvent nano-droplets to form gaseous ions. This means that the smaller the nano droplets formed, the easier the selected solvent is to

evaporate, and the easier it is to detach a single ion and its solvent shell from the droplet. This increases the analytical sensitivity (194), and is the major advantage of (nano)ESI.

1.5.1.2.1.1.2 The Charged Residue Model (CRM)

Small molecules enter the gaseous phase through the IEM. Large molecules such as proteins on the other hand, enter the gaseous phase using the charged residue model (CRM). Here, droplets near the Rayleigh limit and containing single analytes evaporate until completely dry. Importantly, as the solvent shell fully evaporates, the charge is transferred to the large molecule analyte (187). Droplets are thought to remain close to the Rayleigh limit throughout the drying process, suggesting that charge reduction must occur as the droplet size decreases (185). Charge reduction can occur via the IEM, with the loss of small ions and protons (195).

The nomenclature ESI can be misleading. The steps involved in ESI do not directly form charged ions of molecules, but rather play a crucial role in forming singular, charged gaseous molecules for analysis. Formation of ions is primarily achieved through chemical manipulation whilst the analytes are still in liquid form, and ESI ensures that these charged analytes can be observed individually in their gaseous form.

1.5.1.2.1.2 Advantages of ESI and (nano)ESI

ESI, the most frequently applied MS ionisation technique, offers many advantages. It is an atmospheric pressure ionisation technique (API), allowing location of the ionisation source to be outside the high vacuum of the MS, and therefore allowing coupling to other analytical techniques. It permits the analysis of very large molecules (up to 70,000 Da) (179) as multiply charged species can be produced. Mass analysis is measured as a mass to charge ratio (*m/z*), therefore formation of multiply charged species reduces the required detection mass range limit, allowing even low mass range instruments to detect large molecules. Sensitivity can go as low as femto and pico-molar concentrations, and the soft ionisation approach allows measurement of intact analytes (179).

1.5.1.3 Mass Analysers

1.5.1.3.1 Fourier Transform (FT) Mass Analysers: The Orbitrap

A key distinction between quadrupole (196)and ion trap mass analysers (197), and Fourier transform (FT) mass analysers is that for the former, mass-to-charge ratio (*m/z*) is determined based on ion stability whilst in the latter it is based on the ion's frequency of motion. This key distinction is important, as Orbitrap mass analysers offer much higher mass resolution than IT and quadrupole mass analysers. The Orbitrap MS is the only new MS introduced on to the market within the last 30 years. The general mechanism by which the Orbitrap works involves a quadro-logarithmic field (196) inducing ion motion inside the Orbitrap for subsequent detection.

The Orbitrap is made of two curved outer electrodes sandwiching a central electrode (Figure 5). Upon entering the Orbitrap, ions are subjected to a quadro-logarithmic field (Equation 5) and static field and fall into motion. A radially induced electric field pushes ions towards the central electrode whilst the velocity they carry upon entry creates antagonistic centrifugal forces. This causes the ion to fall into circular motion around the central electrode. An axially induced electric field is also applied to the Orbitrap and causes the ion motion to utilise the widest parts of the Orbitrap as the force pushes out on the ions. This results in harmonic axial oscillations, which can be calculated using Equation 6. Equation 7 demonstrates the frequency of axial oscillation in relation to the Quadro-logarithmic field. Depending on size, each ion exhibits a specific axial ion frequency which can be detected by outer electrodes on the Orbitrap, which detect the image current induced by the frequency of each ion of different *m/z* harmonic oscillation frequency. For example, a positively charged ion in harmonic axial motion will attract increasingly negative charge onto the outer electrode as it gets closer and closer to it. This charge is specific to the ion's frequency, which in turn depends on the *m/z* of the ion. This is called image current detection (198).

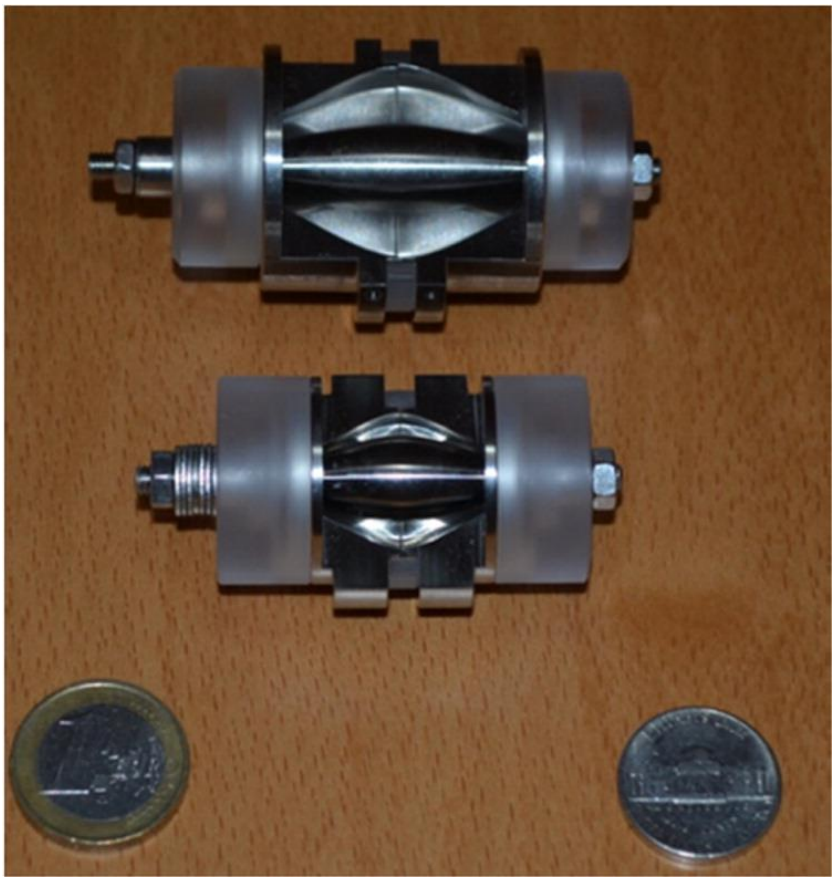


Figure 5: Image showing a standard Orbitrap (top) and a high field Orbitrap, comprised of two curved outer electrodes enveloping a central electrode. Image taken from (198).

$$U(r,z) = \frac{k}{2\left(z^2 - \frac{r^2}{2}\right)} + \frac{k}{2}.(Rm)^2.\ln\left[\frac{r}{Rm}\right] + C$$

Equation 5: Equation calculating the Quadro-logarithmic field distribution exerted by the electrical voltages applied to the Orbitrap electrodes. This Quadro-logarithmic field aids in ion trapping and facilitating harmonic ion oscillations.

$$\omega = \sqrt{\frac{e}{m \div z} \times k}$$

Equation 6: Equation demonstrating frequency of harmonic axial oscillation, where ω = frequency of axial oscillation, m = mass, z = charge, k = field curvature, e = elementary charge (1.602 x 10-19 C).

$$\omega = \sqrt{\frac{e}{m \div z} \times \frac{2 \times Ur}{R_m^2 \ln\left(\frac{R_2}{R_1}\right) - \frac{1}{2}[R_2^2 - R_1^2]}}$$

Equation 7: Frequency of harmonic axial oscillation calculated in relation to the Quadro-logarithmic field.

How the ions enter the Orbitrap mass analyser is important. Without special circumvention, ions entering the Orbitrap would oscillate out of phase (incoherently), and slowly start to fall into phase. This phenomenon would decrease the mass resolution of the instrument as ion harmonic axial oscillation frequencies could not be accurately recorded if they are changing. To ensure ions fall into coherent axial oscillation immediately, they must be injected in a specialised way. An efficient strategy is to use a curved linear ion trap (C-trap). This type of IT works exactly as described here (197) but has a curved shape, whose main advantage is that ions can be focused or packed into small volumes with a narrow diameter and then ejected in small, tight ion packets. Prior to injection into the Orbitrap, ions are stored in the C-trap. Ions are then injected into the Orbitrap from an off-centre entry orifice and in a very short time. This is called fast injection and enables ions to fall into axial oscillations immediately without the need for additional energy. Most Orbitrap instrument designs use a C-trap, but a linear ion trap can be used instead. Here, ions cannot be packed as tightly prior to injection, neither can they be injected as rapidly into to Orbitrap. Ion oscillation using this slow injection requires additional ion excitation (199). During ion injection, the gas in the linear or C-trap can be carried over into the Orbitrap. This is commonly minimised using transfer lenses located between the ion trap and Orbitrap, deflecting the ion beam slightly from its path towards the entry orifice and leaving gas molecules behind (199).

Additional to using a C-trap for ion injection, the electric field at the entry electrode is lowered as the ions enter the Orbitrap. Excessive ion motion results in collision with the electrodes and loss of the ion population. A lowered electric field ensures ions do not carry too much energy, and therefore have smaller radial and axial ion motion. Once the ions have entered the mass analyser, the electric field is gradually ramped up. As it increases, ions experience force that pushes them towards the central electrode, further keeping them safe from collision with the outer electrodes, whilst simultaneously decreasing the size of the ion

packet. This is known as electrodynamic squeezing, and ensures ions fall into orbit. The voltage ramping determines the *m/z* of trapped ions, so electrodynamic squeezing also serves to increase the *m/z* range of the ions that can be trapped. This squeezing continues until there is no longer a chance of ion loss due to collisions with electrodes, and once it stops voltages are held constant to minimise variations in frequency detection in the recorded image currents (Figure 6).

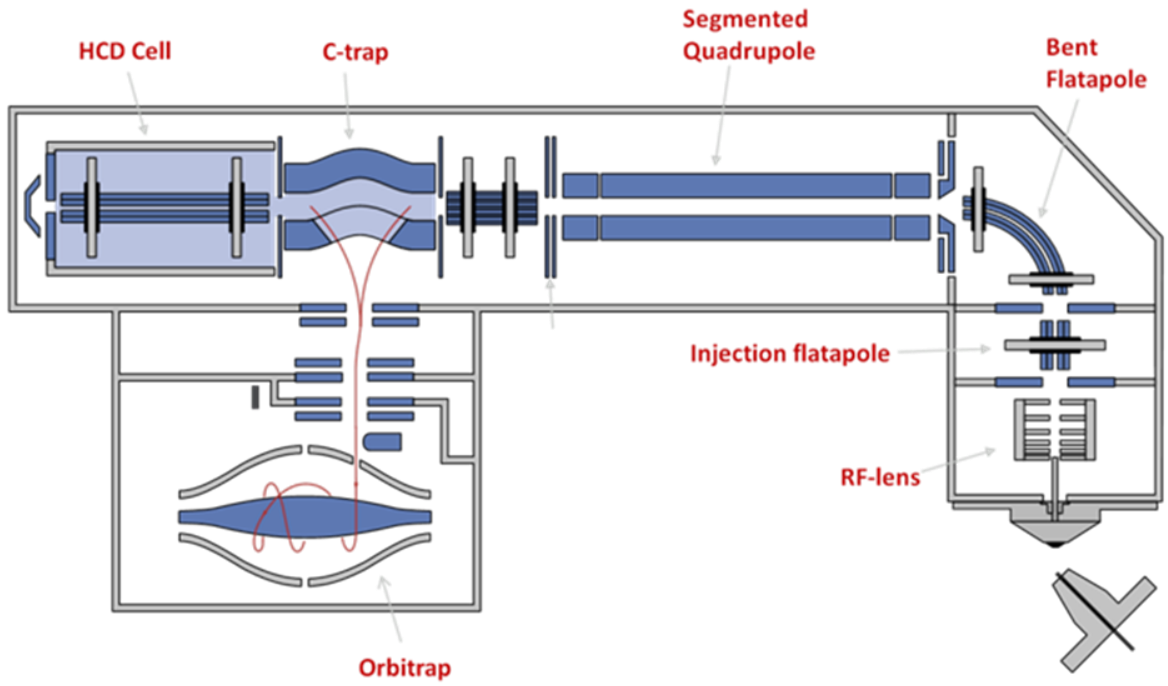


Figure 6: A schematic of an Orbitrap mass spectrometer. An API source introduces ions into the mass spectrometer (bottom right), which are then guided using RF lens towards a quadrupole. This can be used as a mass filter, allowing only ions of certain masses to pass through, or for CID fragmentation. Full-scan or fragmented ions are then passed on to the C-trap, where they are stored ready for fast injection into the Orbitrap. The HCD cell facilitates a higher energy fragmentation strategy than CID. The Orbitrap has outer plates which detect image current, and Fourier transform algorithms are used to convert image current to *m/z* (200).

Fourier transform is a mathematical operation that allows conversion of one function in one domain, such as time for FT mass analysers, into another, such as the frequency domain. This mathematical algorithm is used to convert image currents detected in the Orbitrap into a usable *m/z* reading (201).

Thermo Fisher Scientific has a series of continually improved Orbitrap instruments. Each iteration offers higher mass resolution and acquisition speed. Work presented in this thesis used the LTQ-Orbitrap Elite and the Q-Exactive Plus, so resolution and electrode voltage of these is summarised below (Table 4).

B

Instrument	Orbitrap central electrode voltage [kV]	Orbitrap analyzer size [mm]	eFT	Specified resolution @ m/z 200	Transient length [ms]	†Resolution per millisecond @ m/z 200
LTQ Orbitrap	3.5	30		140,000	1536	91
LTQ Orbitrap XL	3.5	30		140,000	1536	91
Orbitrap Velos	3.5	30		140,000	1536	91
Orbitrap Velos Pro	3.5	30		140,000	1536	91
Orbitrap Elite	3.5	20	+	336,000	768	438
Exactive	5.0	30		100,000	700	143
Exactive Plus	5.0	30	+	140,000	512	273
Exactive Plus EMR	5.0	30	+	140,000	512	273
Q Exactive	5.0	30	+	140,000	512	273
Q Exactive Plus	5.0	30	+	140,000	512	273
Q Exactive HF	5.0	20	+	240,000	512	469
Orbitrap Fusion	5.0	20	+	480,000	1024	469

† The resolution per millisecond is calculated from rounded values and may be inaccurate

Table 4: Summary information for the Orbitrap family of instruments. Taken from (202).

Other FT mass analysers include Fourier Transform Ion Cyclotron Resonance (FTICR) mass analysers, which are also Penning traps in which ions are constrained by a magnetic field move in circular motions of a specified frequency dependent on their *m/z*. Electromagnetic fields are applied to excite trapped ions into their cyclotron frequency. The orbiting ions create an image current, with each *m/z* value generating its own characteristic frequency. Fourier transformation converts frequencies into *m/z*; and since frequencies are measured so accurately this technique offers very high mass resolutions of up to greater than 1,000,000. It is, however, important to note that resolving power decreases with increasing *m/z*. Additional to high resolving power, these instruments offer high mass accuracy and compatibility with newer pulsed ionisation techniques such as MALDI. They are, however, expensive, space consuming and relatively difficult to operate in comparison to TOF and Orbitrap mass analysers.

1.5.2 (nano)ESI SIM-Stitch DIMS

Miniaturisation of analytical instruments has gained popularity over the last two decades (203–207). This has been facilitated by progress in micromachining, a field dedicated to creating small scale instruments (208) to save space and where possible integrate multiple sample preparation and separation techniques into a single step. These are so called lab-on-a-chip technologies. According to literature (209), microfluidic structures on microchannels are conventionally made in planar silicone or glass substrates. Photolithography (210,211), a method of patterning materials, is then used to pattern and then etch the surface of the silicone or glass (209). Finally, the microcapillary construction is covered (209).

Nano-electrospray can be achieved using this chip-based technology (212). By creating a large voltage differential between the mass spectrometer and the chip, liquid sprayed from an emitter etched into the chip forms a Taylor cone (see section 1.5.1.2.1.1) and ESI is achieved. One drawback with this approach is the potential for liquid escaping the emitter to distribute onto the other areas of the chip. This can cause cross contamination and can be resolved by creating a sufficiently large hydrophobicity gradient between the liquid and the chip (209). The concept of creating a hydrophobicity gradient between a solid surface and a liquid was discussed (213). Alternatively, conventional capillary emitters can be attached to the chip, although this can increase void volumes (209).

Nano-electrospray offers lower sample consumption and increased analytical sensitivity. It also gives more uniform ionisation efficiency (214) due to the smaller droplets of sample formed, the higher electric field produced at the nozzle, and excess protons formed because of this high electric field. This reduces ion suppression and enhancement effects, whilst reducing the analyte intensity variability often observed in conventional ESI (215). This makes (nano)ESI an ideal strategy for DIMS metabolomics and exposomics studies, introduced in 1.5.1.1, which suffer more adversely from ion suppression and enhancement, and have a wide range of analytes with different physico-chemical properties and thereby ionisation efficiencies. Furthermore, DIMS analyses have no chromatographic drift (see section 1.5.3.1), allowing more efficient alignment of mass spectra (see section 1.6.3.1.4), and no carry-over (216). The data produced by (nano)ESI-DIMS methods is complex, and strategies for improving this are discussed in section 1.6.3.1.

To further increase analytical sensitivity and dynamic range and improving on earlier methods developed on an FTICR instrument, (216,217) developed a spectral stitching (nano)ESI-DIMS method to be applied on high resolution, high mass accuracy Orbitrap instruments. Here, data are collected in SIM mode, which increases accumulation of ions in the C-trap, thereby increasing sensitivity by five-fold (218). Fifty *m/z* wide SIM windows that overlap by 10 *m/z* are created based on the optimisation carried out in the original work in a method that is only ~3 minutes long (217,218). These are then "stitched" together at the end to create a full mass spectrum. The *m/z* windows were overlapped as original method development had revealed a phenomenon called "edge effects". Here, (217,218) it was discovered that when two closely spaced *m/z* intensity ratios were measured across 30 *m/z* wide windows, the ratios were constant within the central regions of the window. However, this ratio decreased significantly if either of the two ions were close to the edge of the window. As such, overlapping windows were applied so the edges of each window could be removed.

The automated (nano)ESI-DIMS sample delivery was achieved using the Triversa Nanomate, manufactured by Advion Biosciences (Figure 9). This comprises of a cooled sample autosampler, and a robotic arm for sample aspiration. Sample delivery is achieved through movement of this robotic arm to aspirate samples using a conductive pipette tip. The pipette tip with the sample is then aligned with a (nano)ESI nozzle on the ESI-chip (Figure 8) and voltage applied to the conductive pipette tip to start the ESI process (1.5.1.2.1.1). A pressure is applied to the liquid sample to facilitate forward movement of analytes from the pipette tip into the MS orifice through the ESI-chip nozzle.

Figure 7 shows an example of a successful (nano)ESI SIM-Stitch DIMS analysis using the methods described in this section.

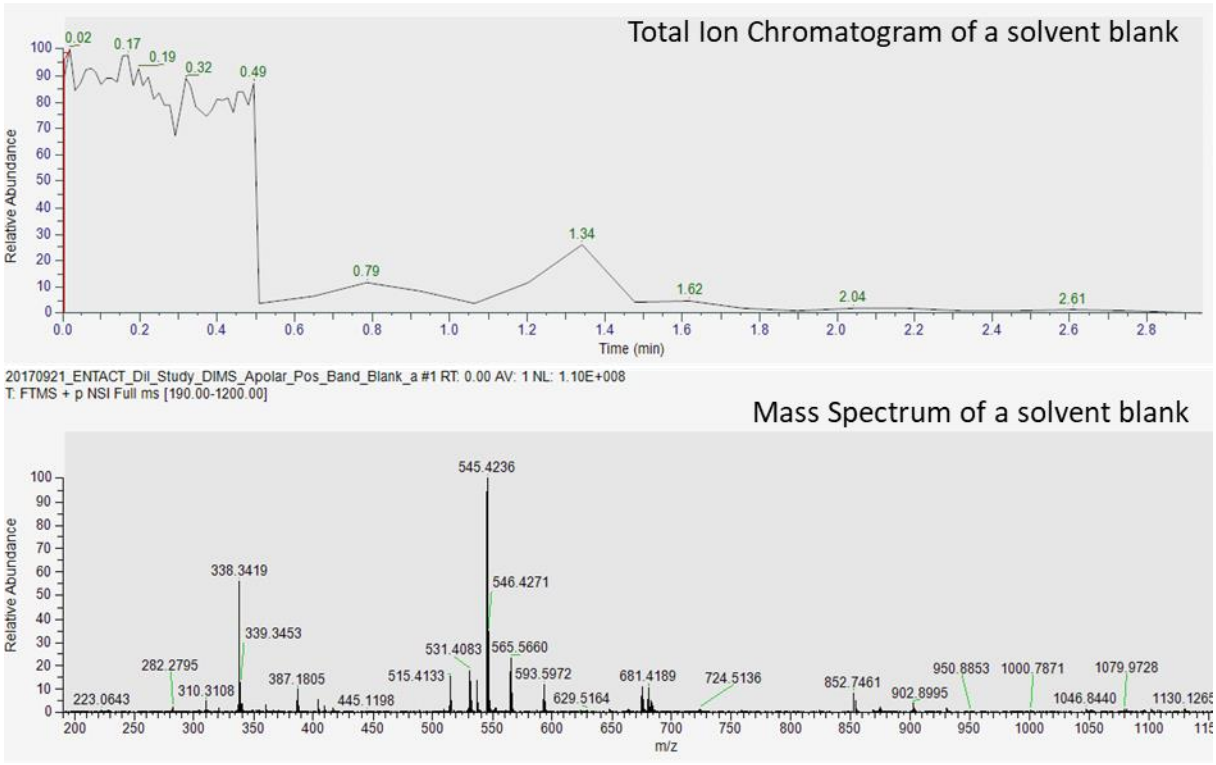


Figure 7: A successful (nano)ESI-DIMS analysis of a solvent blank. The top graph shows the total ion current (TIC), which is the summed intensity of all ions detected across the mass range. The TIC displays a stepwise change at ca. 0.5 min due to the method setup. The first 0.5 min are data collected for the full mass range thus higher ion intensities are observed. Thereafter, data are collected in smaller SIM windows, so the TIC decreases and varies slightly for each SIM window. The bottom graph shows the mass spectrum for the full *m/z* range, with no obvious or apparent sections of missing data. The (nano)ESI-DIMS method applied was developed and described in detail by (218).

(nano)ESI-DIMS methods have been successfully applied for metabolomics studies, yielding high quality, reproducible mass spectral data of high mass accuracy and high mass resolution, as exemplified here (219–222). Most chip-based (nano)ESI sources are manufactured by Advion Biosciences (Figure 8 and Figure 9).

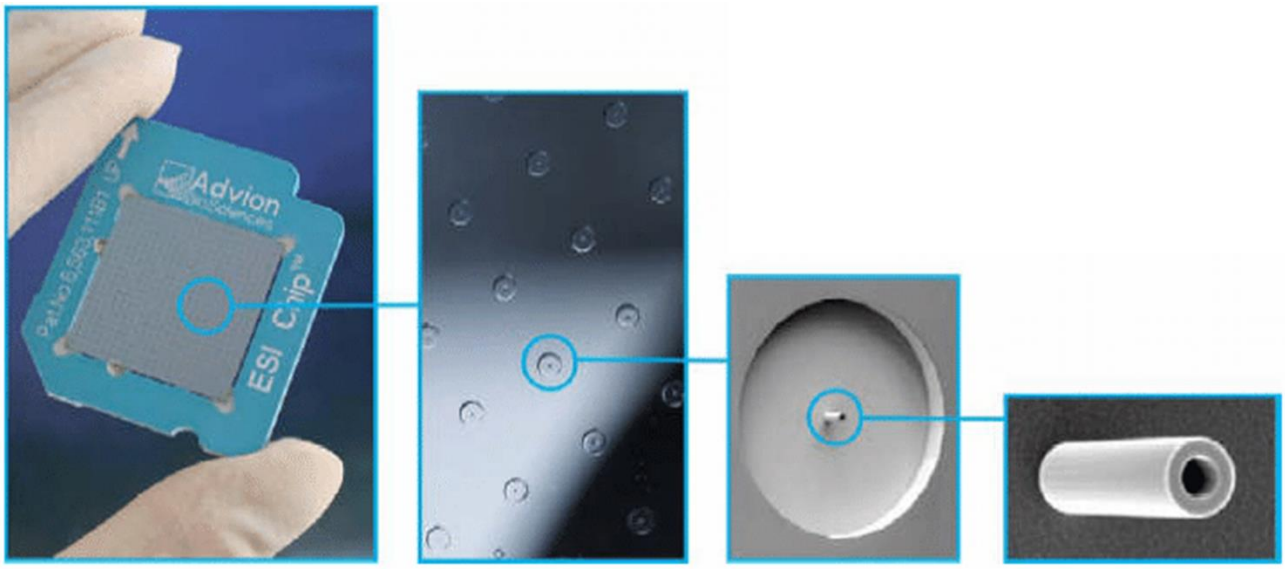


Figure 8: An Advion ESI chip is shown in this figure. A close look shows the small nozzles applied to the chip to facilitate sample spray and minimise spread of liquid onto the chip, reducing sample carry over. Image taken from (223).

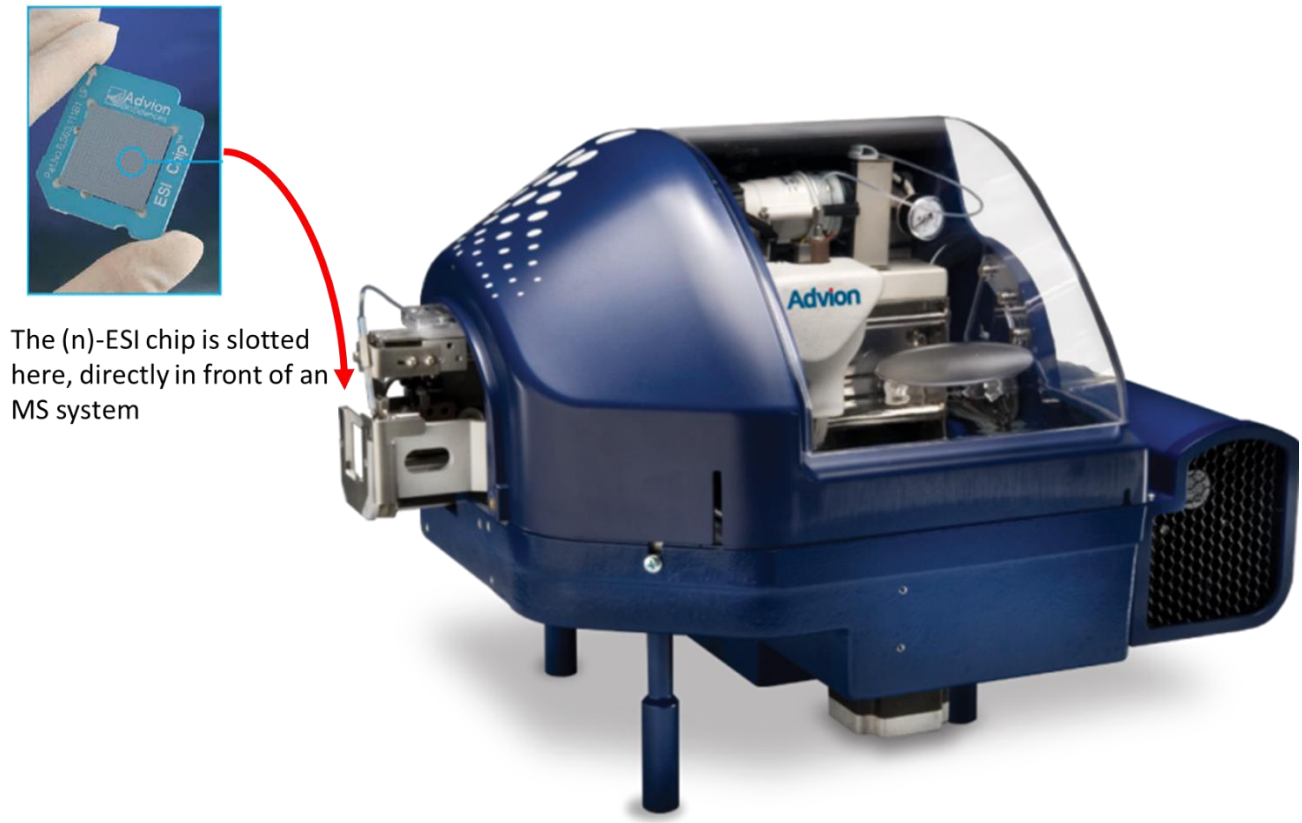


Figure 9: An Advion ESI source set-up is displayed. The ESI chip is slotted into an ESI-carrier located at the front of an autosampler, samples are aspirated and sequentially sprayed through the nozzles on the chip.

1.5.3 Hyphenated Techniques

Despite the advantages of (nano)ESI-DIMS methods, matrix effects which influence sensitivity can still be observed. Ergo, most MS methods used in metabolomics and exposomics studies are hyphenated with separation techniques such as gas chromatography (GC) (224–229), liquid chromatography (LC) (230–234), supercritical fluid (SFC) (235–237), ion mobility (IM) (238–243), and capillary electrophoresis (CE) (244,245). All offer pre-separation of analytes based on a range of physico-chemical properties before mass spectrometry analysis, which can reduce the impact of matrix effects such as ion suppression and/or enhancement and provides additional information in the form of retention or migration times with which the confidence of metabolite identifications can be increased (1.6.4). For example, GC analysis is applied for volatile analytes, and different types of LC have specificity for either polar, charged, or non-polar analytes. SFC is applied for non-volatiles and thermally labile analytes, IM is often employed for large molecules, and CE is applied for charged analytes. However, they also come with caveats. Their use can increase analysis times (for chromatography methods but not for ion mobility), decrease sample throughput and increase the analysis cost. Moreover, analytes of interest may irreversibly interact with the stationary phase (1.5.3.1.1) and many different stationary phase materials are required for increased coverage of whole metabolomes and exposomes. Regardless, the advantages of hyphenated techniques far outweigh any drawbacks, hence they have been employed widely in metabolomics and exposomics research areas. The next section introduces the most applied separation technique.

1.5.3.1 Liquid Chromatography (LC)

Chromatography separates analytes present in complex mixtures based on physico-chemical properties. Liquid chromatography (LC) applies a liquid mobile phase and solid stationary phase and is applied to analyse samples in solution. The data output from a chromatographic separation is a chromatogram constructed of retention time ((RT) time between injection and detection) as the x-axis and response as the y-axis, and in an ideal experiment each chromatographic peak represents a single compound. However, in complex sample mixtures it is more likely that a single chromatographic peak represents several un-separated compounds. Chromatographic separation is achieved through interaction of an analyte with

a stationary phase (SP) and a mobile phase (MP). There are many types of LC stationary phases (see 1.5.3.1.1) which can be packed into a column. Selection of an appropriate stationary phase depends on the chemical analytes to be separated, with different stationary phases offering specificity to different chemical classes. The MP is often a mixture of organic and aqueous solvents and as with the SP, different mobile phase solvents and their composition also offer specificity to certain classes of chemicals. To achieve an LC analysis, the sample is introduced into the MP. The MP carries the analytes towards the column, and once within each analyte partitions or adsorbs and desorbs through interactions between the MP and SP. Stronger interaction with the SP compared to the MP means the analyte elutes from the column slower than analytes that interact less with the SP. The MP entering the column is technically referred to as the eluent, and the MP exiting the column is the eluate. The interaction of compounds with the stationary phase is called retention, whilst the washing of compounds from the column is called elution.

A liquid chromatograph setup includes a pump system to push the solvents, stored in solvent reservoirs, around the system, travelling through the column and ultimately to the detector. As such, these pumps exert a high back pressure on the system, and thus the MP is pressurised. The pump system can be binary (two pumps) or quaternary (four pumps), with the former pumping two different mobile phases whilst the latter pumps up to four different mobile phases. Most metabolomics and exposomics analyses gain sufficient separation from binary pumps, although quaternary pumps can be useful for post-analysis clean up or more exotic analyses involving more than two types of solvents for separation. Historically, LC was referred to as high pressure liquid chromatography (HPLC) (246). This has evolved to ultra-high performance liquid chromatography (U(H)PLC) and ultra-performance liquid chromatography (UPLC), although it must be noted there is no difference between these two (247). Solvents arrive at the pump compartment in different solvent lines, often composed of peek tubing (248). Upon arrival at the pump compartment, solvents are mixed in a mixer (or mixing valve) prior to entering the pumps. Some solvent mixtures result in gas production, so the pump compartment is also often equipped with a degasser.

The column consists of an outer tube or container typically made of stainless steel with the stationary phase packed within. This outer container must withstand high pressure and temperature and must maintain its dimensions for reproducible results. The packing material

depends on the specificity required, so different materials are manufactured, with silica particles most popular. Packing material, regardless of chemical structure, is often manufactured as spherical porous beads of varying dimensions.

As the solvent is pumped towards the column, the sample, located in a compartment of the LC system called the autosampler, is injected using the system's injector, directly into the flow path of the mixed solvent as a narrow band just prior to arriving at the column. The autosampler is often temperature controlled to minimise sample degradation as medium to large-scale studies can take days for analysis and samples are stored in the autosampler whilst awaiting analysis. The injected sample then enters the column as a narrow band, where separation occurs, and is pumped, together with the MP, towards the detector. The mobile phase flow can be constant (isocratic elution) or gradually change with time (gradient elution). Changing the MP ratios during analysis can reduce the elution time for some analytes, or conversely, increase the elution time of other analytes, and as such is used to improve chromatographic resolution and shorten analysis times (249). Various detectors can be attached to the end of liquid chromatographs which are beyond the scope of this thesis, including ultraviolet detectors (UV) (250) and diode array detection (DAD) (251). Here, focus is on detection by MS. Figure 10 shows the basic components of a liquid chromatograph, and Figure 11 exemplifies a chromatogram. The samples that have been detected are disposed of as waste.

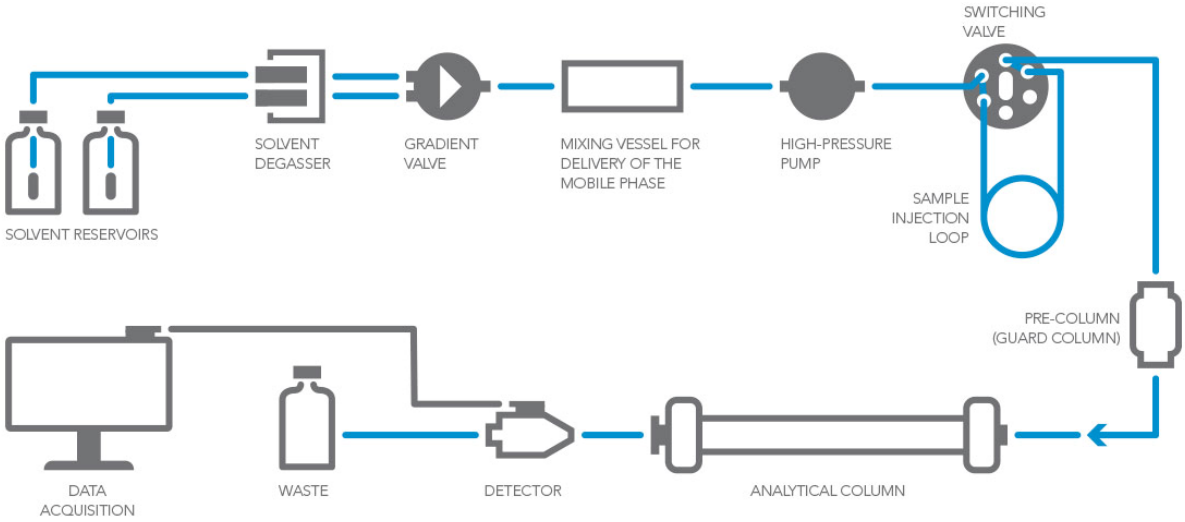


Figure 10: Basic components of a liquid chromatograph are shown. Solvent reservoirs store several solvents necessary for separation, commonly two, but more can and have been used. Solvents move in separate lines towards the solvent mixer, influenced by gravity (no pumping occurs at this stage). On the way they pass through a solvent degasser which removes any gases produced by each solvent. The gradient valve (mixer) is electronically controlled and proportions the mobile phases for gradient elution as specified by the user. The gradient valve is ignored when isocratic elution is applied. The pumps then push the solvent mixture towards the column, and samples are injected from the autosampler compartment into the solvent flow path. Sample separation occurs in the column, with samples partitioning between the mobile and stationary phases resulting in different elution times. Eluted compounds are then sent on to the detector. Image taken from (252).

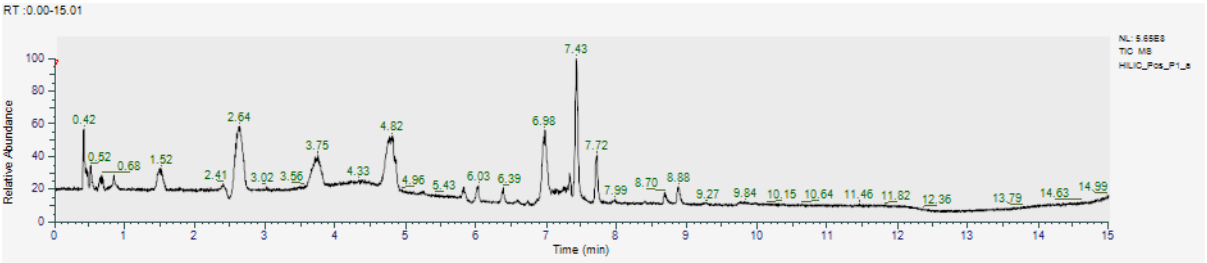


Figure 11: A chromatogram is shown (the detector used for this data collection was a mass spectrometer). The chromatogram comprised of RT along the x-axis and relative abundance along the y-axis. Separated compounds are shown as peaks in the chromatogram. The area between RT 10-13 min demonstrates a common scenario when analysing complex mixtures; insufficient resolution. This impedes baseline separation and results in "crowded" chromatograms.

1.5.3.1.1 Types of Liquid Chromatography (LC)

1.5.3.1.1.1 Reversed Phase (RP) Chromatography

LC can be divided into different types depending on both the MP and SP. The most frequently applied chromatography for various applications such as metabolomics and exposomics is reversed phase (RP), which has a non-polar SP and a polar MP. Commonly used RP column materials are comprised of hydrophobic hydrocarbon chains such as C_8 (octyl), C_{18} (octadecyl) or C_{30} chains bonded to silica particles (Figure 12) and use a combination of polar and non-polar organic and aqueous solvents for gradient elution such as water and acetonitrile or methanol including volatile salts.

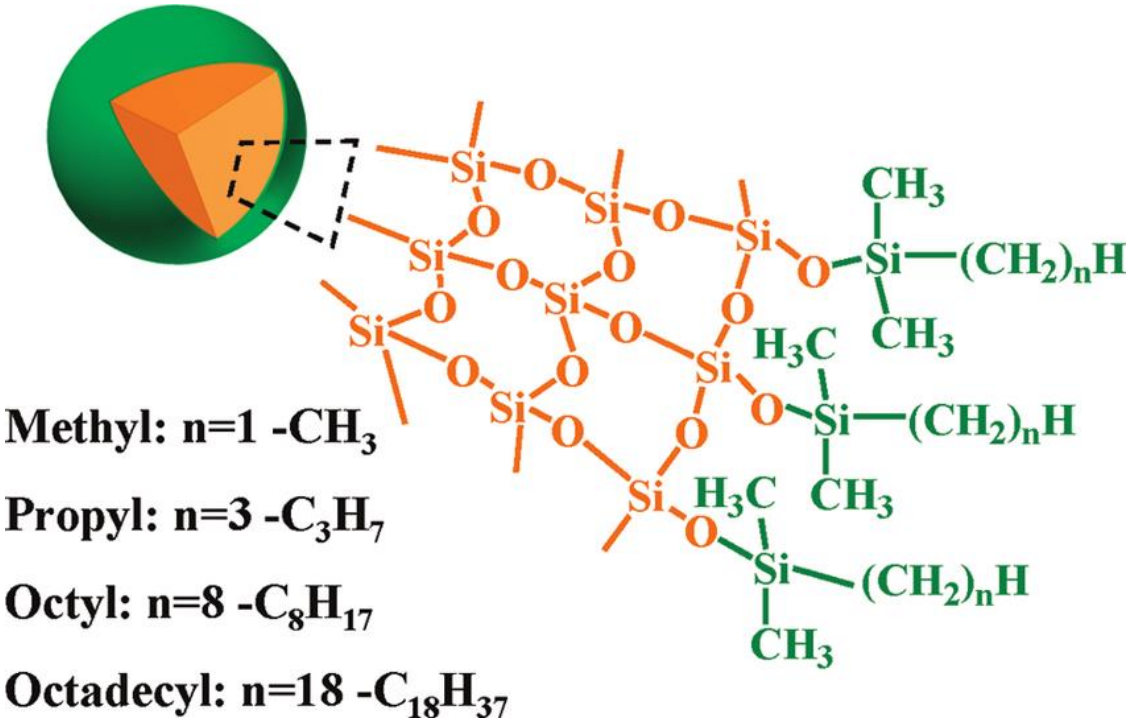


Figure 12: Reversed phase silica bonded particles are shown. The silica part of the particles contains silanol groups, to which the hydrocarbon chain is attached to achieve hydrophobicity of the packing material. The longer the hydrocarbon chain, the more hydrophobic the stationary phase is. Image taken from (253).

Additionally, these solvents can include buffer solutions by adding combinations of modifiers such as acetic acid and ammonium acetate because the pH of the mobile phase (and therefore whether analytes or neutral or are ionised) is an important parameter. At the right pH range, some of the molecules of each chemical of interest will dissociate and be ionised in solution, whilst others will remain unionised as has been described in 1.5.1.2.1.1. The degree of

ionisation is reliant on the dissociation constant K_a of the analyte, and the pH. Analyte retention on a reversed phase column is least when compounds are ionised, as ionised compounds are more polar. As such, the unionised forms of each analyte are preferable for good retention on a reversed phase column. Indeed, it is not uncommon to observe the ionised form of an analyte in the void volume (unretained) whilst the unionised form is retained and eluted much later. An important characteristic occurs during the interaction of the analytes with the stationary phase; the analytes adsorb to the surface of the stationary phase, held in place by weak van der Waal's forces (weak electric forces). The creation of buffer solutions minimises the impact of Bronsted-Lowry acid-base reactions, where protons are donated from one acid (negatively charged) to form its conjugate base (positively charged) and vice-versa, as buffer solutions can resist small increases in hydrogen ion concentration from the acidic proton donors.

For RP assays, greater retention is obtained by using longer hydrophobic alkyl chains as the column packing material. One of the reasons RP is so popular is that it uses water as a part of the MP composition and a wide range of compounds can be dissolved in water. In relation to metabolomics, RP is an ideal choice for its selectivity towards non-polar analytes such as lipids and any compounds with a carbon skeleton.

1.5.3.1.1.2 Normal Phase Chromatography

Normal phase chromatography (NP) was the first type of liquid chromatography developed and operates in the opposite process when compared to reversed phase chromatography, with a polar SP and non-polar MP. It is the oldest form of chromatography, and it too employs the use of modifiers to create buffer solutions with the MP and encourage proton donation or acceptance. The column packing material is often made from silica or alumina whilst the MP uses solvents such as hexane. Normal Phase chromatography is usually applied for ionic or very polar compounds that may interact too weakly with a hydrophobic RP stationary phase. Here, the ionised form of the analyte is desired to increase retention. One disadvantage is that a smaller number of compounds can be dissolved in these hydrophobic organic solvents. To resolve this, aqueous NP is used. This has a polar SP, but instead of the hydrophobic MP it has an aqueous-organic MP composition instead. The organic content is

usually high (< 50%). This still allows for retention of compounds weakly retained in RP but introduces water to allow dissolution of analytes.

1.5.3.1.1.2.1 Hydrophilic Interaction Chromatography (HILIC)

Aqueous NP is referred to as hydrophilic interaction chromatography (HILIC). Stationary phase materials for HILIC are composed of silica, amide, amine or cyano- functional groups, although other materials are also used for more specialist applications. Since the HILIC SP is polar in nature, the water molecules in the aqueous part of the MP form a hydrating layer around the SP (Figure 13). There are two conflicting theories on specifically how analyte retention is achieved. The first proposes that analytes adsorb to the surface of SP, just as occurs in RP and NP chromatography. However, others posit that the analytes do not adsorb to the surface of the SP, but rather partition between the organic MP and the hydrating layer of water around the SP molecules. Moreover, some groups suggest that a mixed approach occurs, with both adsorption and partition of analyte occurring (254,255).

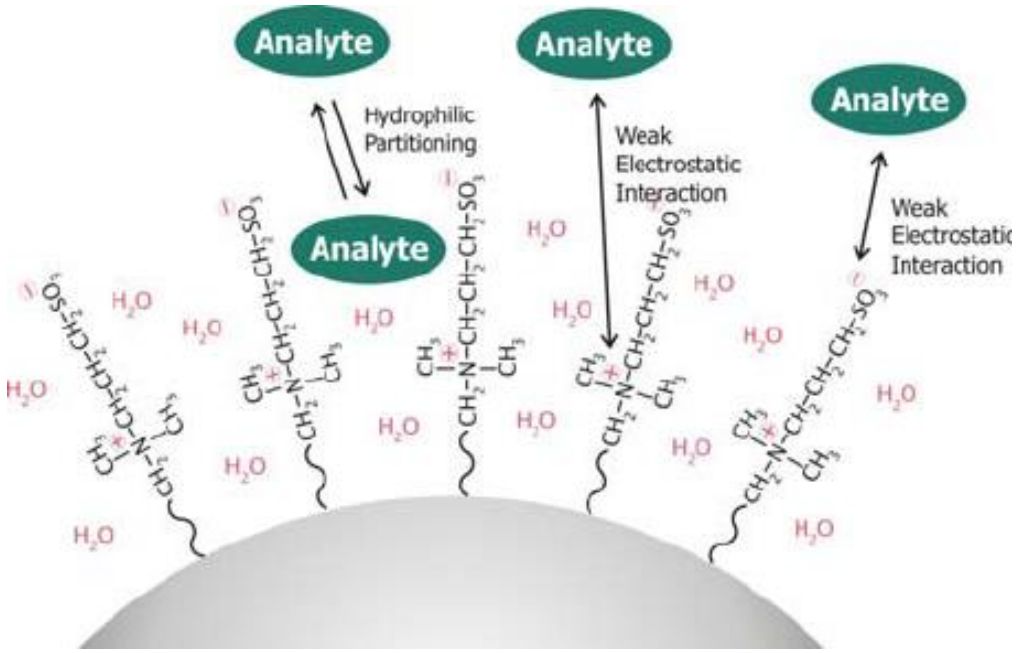


Figure 13: A HILIC particle is shown, with analytes partitioning between the organic MP and the water layer surrounding the HILIC SP particles. Taken from (256).

The marriage between LC and MS is a happy one, producing datasets rich with information that can be used to discern the qualitative and quantitative composition of any suitable sample type. Application of different types of chromatography provides even more comprehensive chemical coverage, allowing for more comprehensive characterisation of metabolomics and exposomics sample types, although these fields are still far from full characterisation due to vast physico-chemical properties of analytes.

Many variables are important in understanding the theory of chromatography. Key introductory terms and principles, such as definitions of RT, adjusted RT, retention volume, capacity factor, resolution, and separation efficiency are defined next. The time it takes an analyte to arrive at the detector after injection is the retention time (t_R). Ideally, with sufficient separation efficiency, each analyte will have a distict t_R, and this information can be used to assign identities more confidently to the peaks (1.6.4). The time taken for unretained solvent to elute is the dead time or dead volume (t_m). Knowing the dead volume allows calculation of an adjusted t_R', which is the time an analyte spends inside the column ($t_R' = t_R - t_m$). The retention volume (V_R) is the amount of solvent required to elute an analyte from the SP and can be calculated by considering the flow rate (F) ($V_R = t_R \times F$). A relative retention time (α) can be calculated between two peaks (α = tR2'/ tR1'), where α is always greater than 1. A capacity factor (k') can also be calculated for each peak. This is the degree to which an analyte is interacting with the stationary phase. The capacity factor is calculated relative to the dead time ($k' = t_R'/t_m$). A capacity factor between 1 and 5 is ideal, indicating some interaction with the SP, but one as large as 20 indicates that elution will take a very long time. The capacity factor impacts the resolution (R_s), which is the degree to which two peaks are separated at baseline from one another. The higher the R_s, the better the chromatographic results. Unresolved peaks complicate assignment of peak identity.

The separation efficiency of a chromatographic analysis is the success with which analytes in a mixture appear as distinct separate peaks in the chromatogram. It is determined by multiple parameters, including peak width. A single analyte travels through the column in a "band" of molecules of that analyte, which spreads into a Gaussian shape as it progresses, and is visualised in the chromatogram as a Gaussian shaped peak. The Gaussian band's, or peak's, standard deviation (σ) can be determined. It is standard to measure the peak width at half height ($w_{1/2}$) or at the baseline between the tangent drawn at the steepest part of the peak

(w). Therefore, with normally distributed peaks, $w_{1/2}$ =2.35σ and w =4σ. Calculation of peak widths allows computation of R_s, which can be calculated as $R_s = \Delta t_R/w_{av} = \Delta V_R/w_{1/2.av} = 0.589\Delta t_R/ w_{1/2.av}$, where Δt_R and ΔV_R are the separation between peaks and w_{av} is the average width of the 2 peaks. These equations demonstrate that the smaller the peak width the better the R_s, and separation is governed by the width of the peak and the distance between peaks, with larger distances between peaks reflecting greater column selectivity.

Peak width is increased by diffusion, which occurs as the band of analytes moving through the column spreads out from an area of high to low concentration. This is called band broadening and is what causes the very narrow band to emerge on the other side of the column as a Gaussian shaped band. The rate at which diffusion occurs (flux) can be calculated to give a diffusion coefficient. Flux is proportional to the band concentration gradient.

1.6 Metabolomics and Exposomics Workflows for nano(ESI)-DIMS and LC-MS

There are multiple steps involved in the metabolomics workflow and depending on the experiment's objectives these may vary between groups. The main steps in the metabolomics workflow are shown in Figure 14. N.B. such a workflow is like what would be applied in exposomics.

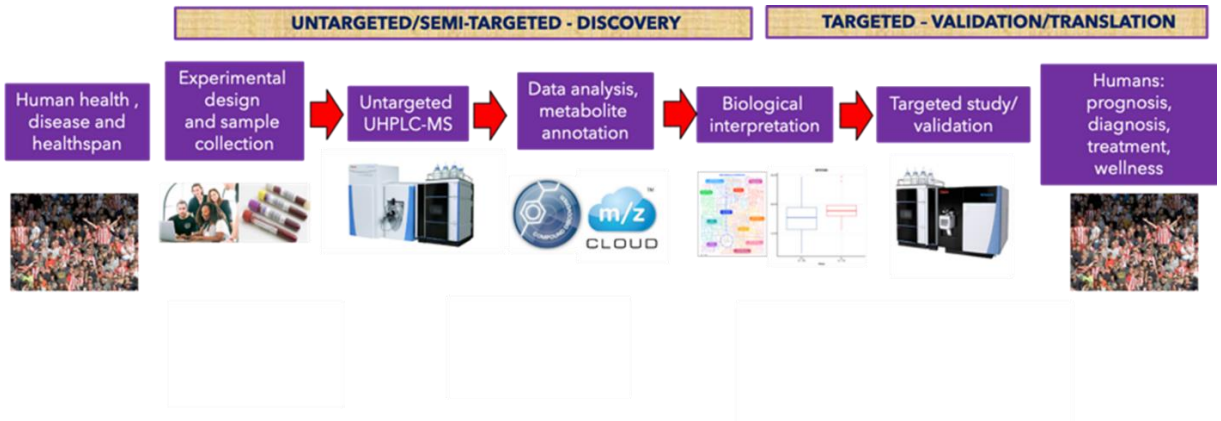


Figure 14: A general metabolomics workflow is presented. In this example, the focus is on human health, but the workflow applies for any sample types. Depending on the objectives of the study, an appropriate experimental design is selected, and samples collected and stored until ready for analysis. In this example, the analytical technique applied is U(H)PLC-MS, but this can be any suitable analytical technique such as (nano)ESI-DIMS. Data are collected, either in MS1, MS2, or both, and then analysed (e.g., using XCMS (257) for raw data processing). Metabolite annotation is then carried out, followed by biological interpretation. Essentially, targeted analyses are applied to confirm any biological interpretations made, and in this example these data can aid prognosis, diagnosis, or treatment progression.

1.6.1 Sample Collection and Preparation

Metabolomics and exposomics studies have many sample types. For clinical studies, urine, tissues, and blood products (plasma, serum) are the most frequently studied sample types because they are easy to collect and store. Environmental sample types depend on the biological question and study and are much more varied. Some sample types (cells and tissues) have both enzymes and metabolites present and are therefore always in undergoing biochemical processes. For such types of samples, arguably the most important aspect of collecting any metabolomics or exposomics data is metabolic quenching. As the presence of enzymes in cells means the metabolome is dynamic and always in flux, it is important to stop metabolic processes immediately when collecting samples (258), acting as fast as possible as metabolic changes can occur very rapidly (259). There are various methods for doing this depending on sample type. Samples can be frozen immediately upon collection using liquid nitrogen, addition of cold solvent, addition of acid, or fast heating (260–265). Once quenched, it may be necessary for some sample types to remove water from samples through drying. This is an important step in minimising bacterial growth in samples, particularly those intended for long term storage. Additionally, water can disrupt sample extraction using organic solvents (266). Dried samples must then be stored at low temperatures to avoid sample degradation and water reabsorption, with most samples stored at -80°C until ready for analysis (266). However, for some sample types such as urine where there are no enzymes, quenching is not required.

Sample extraction protocols are dependent on the sample type, type of chromatography, experimental design, objectives, and indeed targeted or non-targeted analysis or analytical assay. Prasad and Ferenci defined an efficient extraction protocol as one that will extract the largest number of metabolites and be non-destructive (267). Additionally, extraction reproducibility is important to ensure high quality data. To maximise extraction efficiency by making components within tissues or whole organisms accessible, metabolomics and exposomics tissue, organ, or whole organism sample types can be homogenised prior to extraction. There are various approaches for homogenisation, including the gold standard manual grinding technique using a pestle and mortar (268), the use of homogenisers, which are quicker (269), and freeze-thaw cycles (270) that can lyse cellular walls for plant and

bacterial sample types (266). Extraction of metabolites into extraction solutions can then be carried out.

Different extraction protocols are used for sample extraction, including liquid-liquid extraction (LLE) (271), solid-liquid extraction (SLE) (272), solid-phase extraction (SPE) (273), and supercritical fluid extraction (SFE) (274). Indeed, in an ideal study, given sufficient time, money, and sample quantity, many different extraction protocols would be applied to extract as many different metabolite classes as possible and increase the detectable metabolome coverage. The use of different solvents requires considerations for desirable and ideal characteristics of those solvents, such as low toxicity, high solubilising power, and selectivity (266), although all these desirables are not always achievable, and some are study-specific. Polar, hydrophilic and ionic or ionisable compounds are extracted using methanol/ethanol precipitations (275), whilst moderately polar, non-polar, and hydrophobic compounds are extracted using combinations of water-organic solvent mixtures, including chloroform which is toxic. It is important to note that chloroform extraction would require drying down as this solvent is not compatible with MS (276,277). The advantage of these organic solvents over water is that a wider portfolio of metabolites can be extracted this way, and these solvents are compatible with most analytical techniques (266). Indeed, selectivity in metabolomics or other NTA is less essential. Rather, what is required are extraction protocols that can target larger classifications of analyte to reduce analytical interference when collecting data. Most extraction protocols attempt to separate polar and non-polar analytes (278)

Extracted samples cannot always be analysed directly after extraction using (nano)ESI-DIMS (218) and require removal of solvents by drying down and subsequent resuspension in method specific solvents with added modifiers such as formic acid and ammonium formate for pH control (218). This is because the non-polar extracts for these methods use chloroform as an extraction solvent, which is not compatible with MS. This step can also be used for sample concentration through drying down the available extract and resuspending in a smaller solvent volume to increase metabolite concentrations in the final solution compared to in the sample extracted. For U(H)PLC-MS analysis, samples can usually be analysed directly after extraction as these organic solvents are often compatible (or the same) with the mobile phases used in LC.

1.6.2 Analytical Data Collection

Data collection using MS, whether hyphenated or direct, can be in different ion-modes. Positively and negatively charged ions are analysed using separate positive and negative ion-mode methods, respectively. Furthermore, data can be full-scan (MS1), tandem (MS2) or a combination of both.

A metabolomics study has various classes of samples, often including a control and an experimental group. A typical experiment aims to discern differences between these groups. Innate individual differences can mask the differences arising due to some experimental condition, and thus individual variability must be minimised where possible. For example, research has shown that there are gender specific differences in the serum metabolome (279). This is achieved by using a large sample size, and the appropriate size relies extensively on the study design and objectives, and the desired statistical tests. This is not always possible, indeed for clinical studies it is even more challenging than for plants and cells that can be grown and cultured. Data acquisition therefore often includes multiple biological samples per sample experimental class, with a minimum of three, although this is rarely sufficient and only used in participant or sample limited studies. These multiple samples per class are called biological replicates. Even variation introduced by instrument performance can be problematic when it comes to differential sample analysis, and therefore it is also common practise to collect repeated analyses of the same biological sample. This is called technical replication and serves to characterise the standard deviation of instrument response when repeated analyses of the same sample are carried out. When data analysis begins, biological and technical replicate standard deviations are measured, and samples with high variability are excluded from differential analysis.

To further control the origins of observed class differences, quality control (QC) samples are also analysed during data collection. These are often a pooled mixture of all samples from all classes in the study. This provides an average instrument response to the various analytes present within the samples, whilst also demonstrating instrument response in the relevant sample matrix. Quality control samples are analysed after every 5-6 injections of the study samples, and so also serve to characterise instrument stability throughout the run. QC protocols have been widely reviewed (280–283).

Analysis of solvent, or indeed, just a simulated injection followed by running the MP through the SP to the detector would still yield chromatograms and mass spectra with peaks in them. This is because the compounds present in the solvents, either as contamination from containers or from sample preparation equipment, will still be detected as any compounds would. This is the background level of the whole system and can complicate the annotation of real peaks during data processing. Blank samples are analysed to remove background features and are often injected at the start and end of the run, although there is no consensus on this analysis order in the community. There are different types of blank samples. Extraction blanks are created by simulating the extraction process but without any sample. This gives an indication of background features arising from the sample preparation process and materials. Solvent blanks can also be prepared, indicating only on background features in the solvents used (284).

1.6.3 Raw Data Processing

Data processing often starts with manual inspection of raw data files, inspecting files that appear smaller in file size and removing them from the list of files to be processed if they are empty, have missing sections of spectra, or have insufficient chromatographic resolution. These files would otherwise decrease the precision of the data and thus disrupt differential analysis. N.B. that no data files are removed without justification (i.e., empty spectrum due to lack of spray stability).

There are multiple open access tools that can be used to process raw mass spectral and chromatographic data such as MetAlign (285–287), MZmine (288–290), MetaboAnalyst (291), and XCMS (292–294). Vendor specific software also exists, such as Mass Profiler Professional from Agilent Technologies (295), Metaboscape from Bruker Daltonics (296), and Compound Discoverer from Thermo Fisher Scientific (297). With so many options to select from, data processing can become extremely difficult. Moreover, some of the tools in these fields require some computational expertise. Efforts have been made to simplify the process and make it more accessible to non-computational experts. XCMS, for example, offers the whole non-targeted workflow (298,299) with little expertise required. This has contributed to its popularity as it is one of the most used tools in metabolomics.

Work presented in this thesis used Galaxy-M (300)and MI-pack (301) for (nano)ESI-DIMS data processing and annotation, and Galaxy-M (using XCMS coding within Galaxy-M) (294,302) and the Birmingham mEtabolite Annotation for Mass Spectrometry (BEAMS) (available at: *https://more.bham.ac.uk/beams/*) for U(H)PLC-MS data processing and annotation. These will be described in detail in the next section.

1.6.3.1 Galaxy-M for (nano)ESI-DIMS and U(H)PLC-MS Data Processing

DIMS data processing uses SIM-Stitch and U(H)PLC-MS data processing uses XCMS as summarised in Figure 15. Each step will be described in detail, first for DIMS and then for U(H)PLC-MS, as the first steps of processing vary for these different types of data.

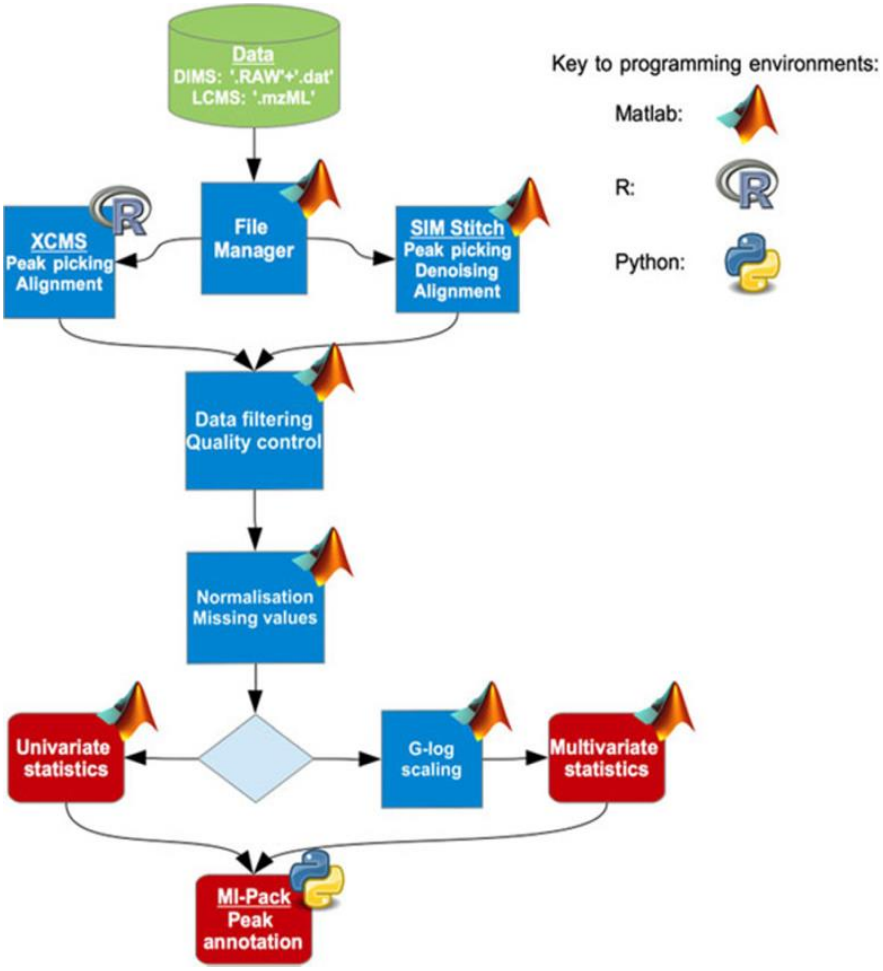


Figure 15: Figure taken from (302). A summary of the Galaxy-M workflow for both DIMS and U(H)PLC-MS data processing is shown. The environments used for each step are denoted, being either Matlab, R, or Python. Data processing steps are different for DIMS and U(H)PLC-MS for the peak-picking steps, but data filtering steps are the same for both types of analysis. MI-Pack is used for metabolite annotation for both DIMS and U(H)PLC-MS.

1.6.3.1.1 Data Handling using the File List Manager for DIMS and U(H)PLC-MS Data

Galaxy-M accepts raw (unprocessed) data for DIMS. Data conversion of raw data to mzML files (303) is required for U(H)PLC-MS data. There are various tools for converting to mzML file format. Proteowizard's MSConvert is open source and commonly used (304), although analysts can employ any file converter. Galaxy-M does not require manual uploading of data files, which can be time consuming. Rather, a file location is accepted, which points to where the data files are stored on the user's computer. Additional information must be submitted alongside the raw and mzML files. For DIMS, file-lists with metainformation about the study are required. These include information on the sample classes (control, experimental, blank, QCs etc.), and the number of technical replicates per biological sample (302).

1.6.3.1.2 Sum and Process Scans for DIMS Data

The SIM-Stitch workflow is applied for the first few steps of DIMS data processing. Multiple scans within each SIM window are averaged. Where multiple scan data is not available, the sum scans step is skipped. Next, the process scans step begins, starting with application of the Hann function (305) , often referred to erroneously as "Hanning" apodization. Frequency measurements in FTMS are taken over a finite amount of time. However, in theory, repeated measurement of the same species could go on infinitely, producing never ending frequency data. This phenomenon results in the observation of neighbouring frequencies when looking at any one species of interest. This is called spectral leakage, and results in small peaks (artefacts) that can be observed around a main peak (Figure 16). Simply put, the Han function is a smoothing function, removing frequency data arising from data collected on frequencies that go on much longer than the allowed time (302,306). Apodized peaks have better peak shapes, which in turn improves resolution and sensitivity (307).

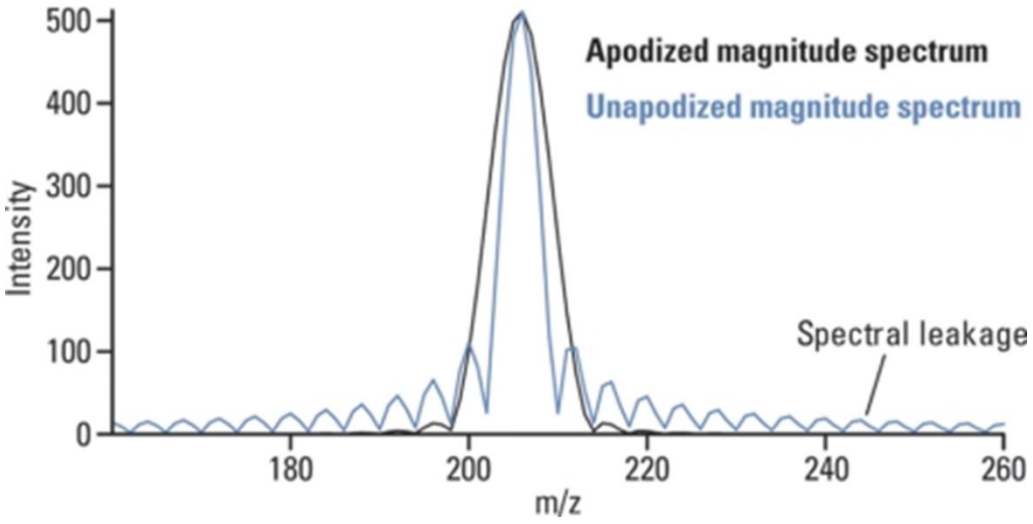


Figure 16: The figure is taken from (306), showing the phenomenon of spectral leakage.

Next, zero-filling is applied. This is another type of smoothing function that involves addition of zero-filled intervals to the scans (306,307). Zero-filling improves peak shape and increases sensitivity of peak discernment (307). Fourier transformation is then applied, converting image current to *m/z* (300,306), followed by baseline correction (300).

1.6.3.1.3 Mass Calibration and SIM-Stitch for DIMS Data

Peaks above a user defined signal to noise ratio (SNR or S/N) are picked (300), using a minimum of S/N = 3. The higher the S/N applied, the less chance there is of selecting noise peaks as part of the analysis, which increase incorrect annotation (FPs). However, using too high a S/N can result in lost information where low intensity metabolites exist. Each SIM window is then calibrated (300), and overlapping windows then stitched together to form a peak list (300).

1.6.3.1.4 Replicate Filtering and Align Samples

Peaks present in x-out of n technical replicates are removed. This step is called replicate filtering, and the ratio to use for filtering can be user adjusted. The more stringent the replicate filtering, the less noise peaks have a chance of being included in the final peak matrix. However, this also increases the chances of missing out on important information.

Once data have been filtered this way, all samples, with their various experimental classes, are aligned in the align samples step (300).

1.6.3.2 XCMS for U(H)PLC-MS Data

U(H)PLC-MS data can also be processed using the Galaxy-M environment and the workflow is summarised in Figure 17.

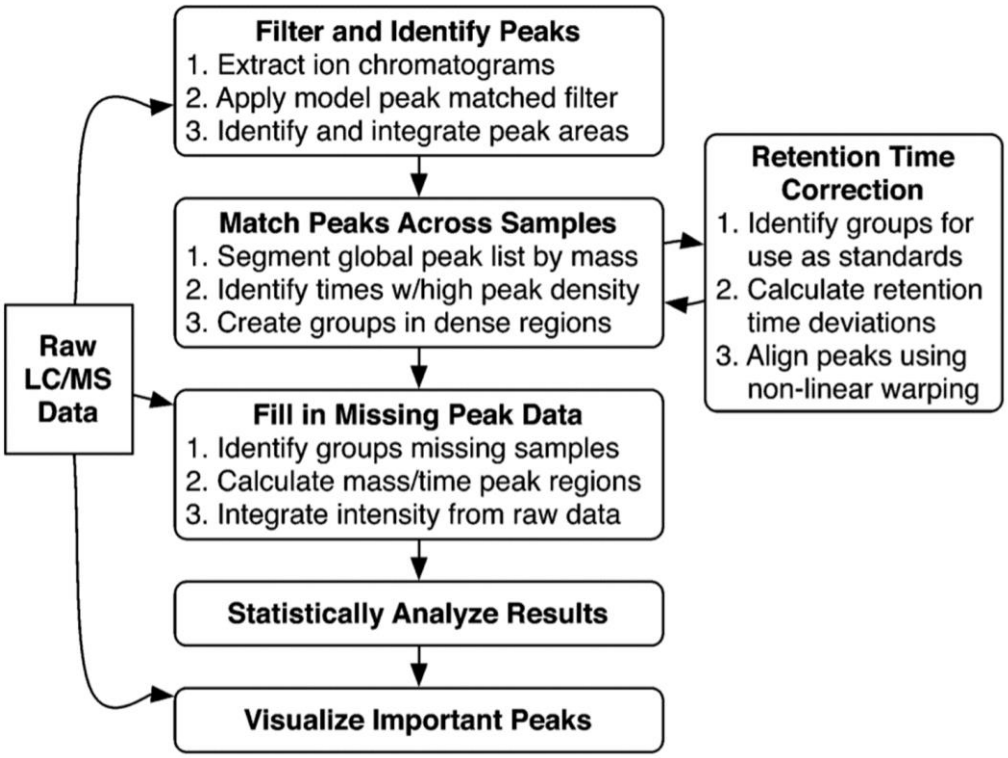


Figure 17: Figure taken from (257), showing the entire data processing steps, including preparation of data prior to statistical analysis (fill in missing data) using XCMS. For U(H)PLC-MS data processing using Galaxy-M, data are filtered, and peaks identified as shown in the first step titled "Filter and Identify Peaks". Peaks are then matched across samples as per the second step titled "Match Peaks Across Samples".

1.6.3.2.1 Peak Detection for U(H)PLC-MS Data

The XCMS algorithm detects peaks by cutting mass spectral data into slices, often 0.1 *m/z* wide. It then focuses on these individual slices in the chromatographic time domain. To determine a signal, the maximum intensity at each time-point in the slice is taken. This is called an extracted ion base peak chromatogram (EIBPC) (257). There are drawbacks to this approach. If the slices are cut too thinly, then mass spectral peaks may be cut in half, divided

between two adjacent bins. Spectral peaks cut in half can impact data alignment, introducing missing values. This in turn can impact differential statistical analyses. Additionally, multiple analytes may contribute to the most intense signal within a bin, making it challenging to differentiate between these different analyte signals (257). Each slice is then filtered using a method called matched filtration, which fits a second derivative Gaussian as a model for peak shape. Application of this model peak shape transforms each peak, causing it to "dip below the x-axis at the peak inflection points" (257). The points at which the peak dips below the x-axis are then used to calculate the peak area (integration) (257). The peak detection process is summarised in Figure 18. After matched filtration and peak integration, peaks are then selected, applying a user defined S/N ratio and excluding any peaks with less than 5 mass spectral peaks per slice (257). Peaks are then aligned across all samples to give a peak list.

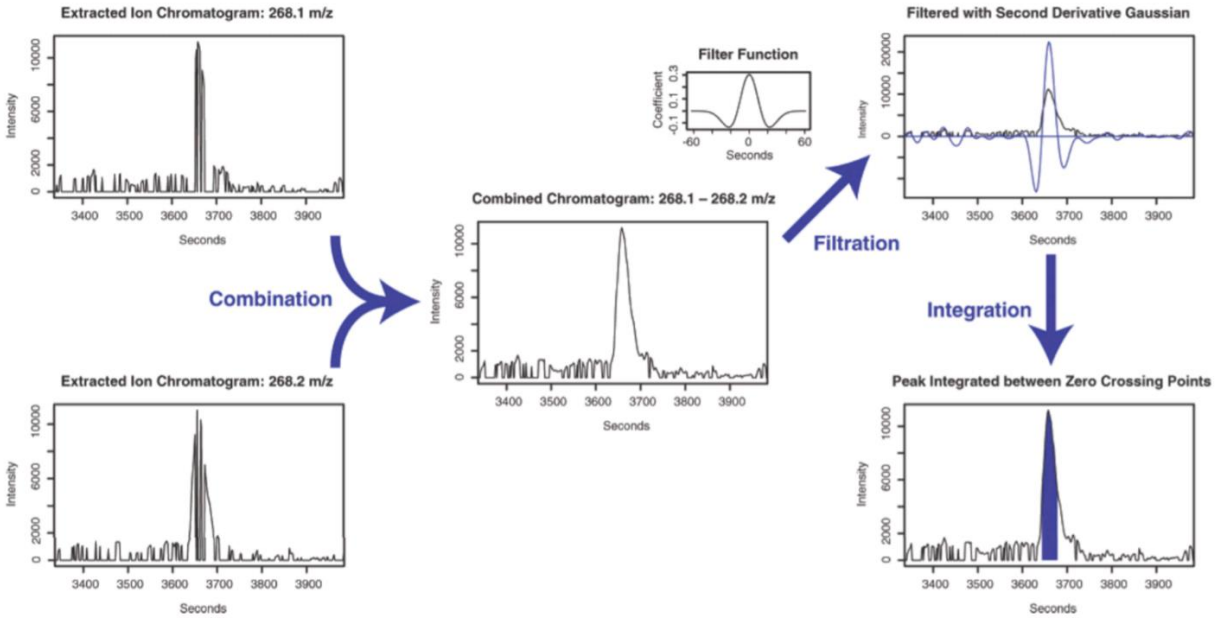


Figure 18: Figure taken from (257), showing the peak detection process for U(H)PLC-MS data using XCMS. To detect chromatographic peaks, the mass spectrum is divided into bins of width 0.1 *m/z*. It then focuses on these individual slices in the chromatographic time domain. A signal is determined by taking the maximum intensity at each time point. This is denoted on the left-hand side of the figure shown here. A matched filtering function is then applied, which causes the resultant combined chromatogram to dip below the x-axis at the peak inflection points. These points on the x-axis are then used to integrate the peak (257).

1.6.3.3 Galaxy-M Blank and Sample Filtering for DIMS and U(H)PLC-MS Data

At this point, both DIMS and U(H)PLC-MS data have been processed up to the align samples step as described above. The following Galaxy-M steps are now the same for both data types. Blank filtering is applied by comparing blank samples to biological samples and removing any overlapping peaks. Blank filtering can exclude all peaks that overlap between samples or can be more conservative and remove peaks that are present above a certain fold change threshold in the blank. For example, using this strategy, a peak present in the biological sample and the blank, but with a 10-fold intensity increase in the blank, must be removed (300). These parameters are left at to the user's discretion. Finally, sample filtering removes peaks that are not found in x-out of n samples (300), again left to the user's discretion.

1.6.4 Metabolite Annotation

Full-scan MS1 data is typically the first data type applied for annotation of metabolites in NTA (NTA). MS1 data is collected for all metabolites detected, as well as for other non-metabolites such as chemicals from the environment which are detected, and it provides a route to significantly decrease the size of the search space of annotations for each metabolite. Annotation of full-scan data is one of the most challenging aspects of MS based NTA, regardless of application. This is because the NTA datasets collected are large, comprising of hundreds of thousands of metabolites as well as non-biologically relevant chemicals (308). Metabolites are often common across species due to shared core metabolic pathways such as the glycolysis metabolic pathway (309,310). This commonality impedes the creation of comprehensive species or sample type-specific databases, and thus far the main strategy for annotation of MS1 data is to search very large reference lists or databases such as the Human Metabolome Database, which contains ~114,000 metabolites (75) while typically <3,000 metabolites are detected in human biofluids and tissues. Not only are there a large number of metabolites in any given sample type, but they also have a wide variety of physicochemical properties and concentrations that span several orders of magnitude (311). The variability in physicochemical properties means their behaviours in U(H)PLC and MS instruments are varied too, making it challenging to predict aspects of their behaviours such as RT, ionisation efficiency and signal linearity (312). The concentration differential between different

metabolites also poses analytical challenges. These will be introduced and discussed briefly below.

The current levels of identification, described in 1.6.5.1, for metabolomics prioritise tandem mass spectrometry (MS/MS) for high confidence identification, with many community members referring to lower confidence assignments as annotations and higher confidence ones as identifications (313,314). Identification requires analysis of authentic chemical standards to create MS/MS libraries, limiting the number of confidently identified metabolites that can be achieved as authentic chemical standards are either too expensive to purchase thousands of them or are simply not available. To circumvent this, many research groups have investigated the feasibility of sharing MS/MS libraries across groups or creating open-source libraries, which is challenged by the different data acquisition strategies such as instrument type and different collision energies (315) applied in collecting MS2 data resulting in different fragmentation spectra (309,316). Moreover, some groups have investigated the use of computational approaches to create in-silico MS/MS libraries (317–320). This has had success in improving metabolite annotation and identification, but also has its limitations. For any MS/MS library to yield a match, high-quality fragmentation spectra are required to be collected for authentic chemical standards to construct the mass spectral library and for metabolites present in biological samples which require identification. High-quality fragmentation spectra require sufficient ion intensities for the precursor ion and the high concentration differences of metabolites in biological samples handicap the collection of good quality MS/MS spectra when metabolite concentrations are too low. Indeed, 60-70% of metabolites typically detected can be classified as being present at low concentrations (Figure 19), thus are unlikely to yield good quality fragmentation mass spectra (309). This means 60-70% of metabolomics datasets rely on MS1 data processing strategies, but relatively less focus has been given to optimisation of software tool parameters for metabolite annotation using MS1 full-scan data.

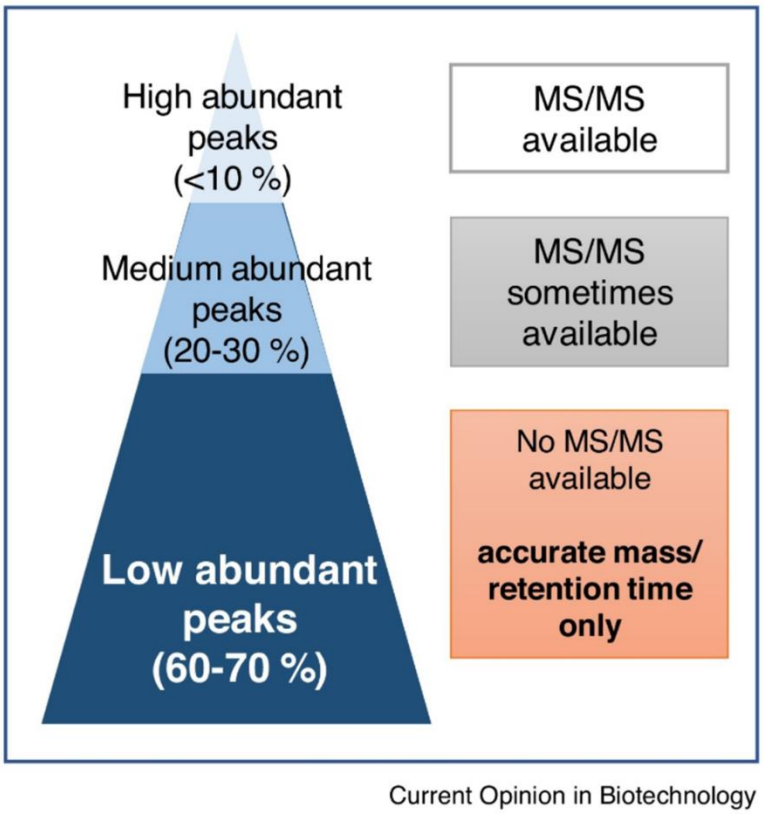


Figure 19: Distribution of metabolite abundances and availability of MS/MS spectra in relation to metabolite abundances. Taken from (309).

It has been reported that only ~1.8% of metabolomics mass spectral peaks are annotated (321). This shows the importance of continually improving strategies for the use of MS1 data for metabolite annotation. However, this is no simple task. In addition to the huge variety of metabolites present across several orders of magnitude of concentration, MS analysis does not yield a single peak for each metabolite. These data are instead comprised of many different peaks corresponding to a single metabolite. If relevant, analytes are initially separated chromatographically, yielding chromatographic peaks in the time domain. The analytes eluting at the same time are then recorded as mass spectral peaks when MS detection is used. Each chromatographic peak therefore 'contains' any mass spectral peaks that have the same or very similar RTs (and this information is missing for DIMS assays). Furthermore, a single metabolite can yield many different mass spectral peaks, often referred to as metabolite features, ion species, ion forms, or degenerate features (322,323). These comprise of adducts, isotopes, in-source fragments, and neutral losses. To further complicate these data, biotransformations are also recorded in the data. This results in peak matrices

with tens or even hundreds of thousands of *m/z* features (often defined as unique *m/z*-RT pairs where chromatography is applied). The presence of these degenerate features increases U(H)PLC-MS data complexity greatly, can reduce the number of metabolites annotated, can decrease the number of TPs, and overall makes biological interpretation difficult, whilst hiding the true number of unique metabolites within each dataset. Moreover, inappropriate handling of these features not only deflates the estimated annotation rate in metabolomics datasets, with unannotated peaks considered as separate and unique metabolites when in-fact they are not always, but it also potentially disrupts statistical analyses (324). Degenerate features must be grouped together to maximise MS1 annotation of metabolomics datasets. Indeed, one group demonstrated that by grouping degenerate features belonging to a single metabolite using a "context-driven approach", and using an orthogonal isotopic labelling strategy, a 90% reduction in the data was observed (324,325).

Therefore, the process of annotation often involves assigning peak patterns (isotopes) and adducts, looking for neutral losses and in-source fragments, and grouping all features representative of the metabolite as a single spectrum. There is certainly no shortage of tools used for annotation in metabolomics, both commercial and open source. Specific to the issues introduced thus far, there is a wealth of new computational approaches for the identification and grouping of degenerate features (322,326–329). The different annotation approaches can be divided into graph-based clustering of co-eluting features by peak shape, unsupervised clustering, and Bayesian probabilistic sampling (322). Graph-based clustering groups co-eluting features together based on similarities between co-eluting profiles. The more similar chromatographic peak shapes are, the higher the likelihood that they belong to the same metabolite. For example, CliqueMS uses this graph-based approach first to determine the peak shapes of all chromatographic peaks in the MS1 data, and then determine similarities in this peak shape of all pairs of features in the data. Groups, or "cliques", of features are created using the network of similarities. Each group of features deemed to belong to a single metabolite is then used to calculate the neutral mass, which seriatim is used for annotation of compound name. Adducts, isotopes, and fragments are also annotated within each group through calculation of *m/z* differences between ion species within the group and annotation against a reference list of common adducts, isotopes, and fragments (330). CAMERA, a popularly applied annotation tool in metabolomics, also uses graph-based clustering,

alongside other approaches, to create compound-spectra (331). One disadvantage of graph-based clustering is that it does not utilise all the available information in the dataset. For example, CAMERA is biased towards high intensity ions and CliqueMS relies on co-elution, thus discards other useful information that does not meet these criteria. Unsupervised clustering attempts to fill in these gaps by utilising as much of the available information in the datasets as possible. For example, RAMClust considers the differences in feature RTs and the correlation between two features across all samples. To be grouped, RTs must be similar (within a user specified window) and features must correlate. This clustering is done for all features present in a peak matrix, thereby using all available data (332). MSClust uses a similar approach, considering RT similarity and quantitative similarity of ion-fragment patterns across all samples in the dataset (333), whilst PUTMEDID also uses RT similarity and correlation analysis to group features (334). Finally, Bayesian probabilistic sampling considers the inter-relatedness of peaks within a peak matrix, creating connectivity networks and assigning metabolite names based on the most probable answers. For example, an *m/z* assigned a metabolite name (by matching against a reference list or database) by mass alone will get a lower probability score than one for which other related components, such as adducts, isotopes, and neutral losses, and biochemical transformations are found. This probability scoring was shown to improve the reliability of assigning molecular formulae to *m/z* features (335–337). Most annotation software uses one or a combination of these different approaches.

Although more commercial and open-source annotation tools are searching for, grouping, and annotating degenerate features, there is still a lack of standardisation on which adducts, isotopes, and neutral losses to include. This is partially because this information is not comprehensively known. Moreover, it is challenging to determine which adducts, isotopes, and neutral losses are prevalent enough to be regularly included during annotation, and which ones are sample type and sample preparation specific. Some tools allow for customised adduct, isotope, and neutral loss lists to be created and added by the user, but this too faces a caveat. Metabolomics requires expertise in analytical chemistry, statistics, informatics, and biology. A significant proportion of researchers in metabolomics must therefore not have expertise in computational approaches, thus are unlikely to utilise customised lists, opting for

the simpler option of applying default parameters. Indeed, the variation and distribution of adducts included in 11 annotation software/tools (anonymised) is shown in Table 5.

Anonymised Software ID	Cumulative Frequency of Adducts Present (positive ion-mode)	Cumulative Frequency of Adducts Present (negative ion-mode)
1	20	16
2	51	40
3	54	40
4	61	42
5	65	43
6	69	46
7	73	48
8	80	50
9	81	52
10	83	54
11	89	56

Table 5: The overlap in the adducts included in 11 software tools applied in metabolomics is shown. Only 6 adducts in positive ion-mode, and 2 adducts in negative ion-mode, are included in all 11 software tools. By contrast, 20 positive ion-mode adducts are included in only 1 software tool out of the 11, whilst 16 negative ion-mode adducts are included in one out of 11 software tools. This demonstrates that although degenerate features are increasingly considered, there is still a lack of standardisation on which ones to include and exclude, and there is a wide variation in the adduct lists applied.

Nonetheless, searching for degenerate features is more prevalent in the newer metabolomics annotation tools released. Increasing attention is paid to searching more extensive adduct, isotope, and neutral loss lists. However, annotation accuracy could be markedly improved by customising the lists used by each tool to the data in question (338). Moreover, grouping parameters during annotation, such as RT window, can also improve annotation accuracy.

1.6.5 Complexities of Metabolite Annotation

No matter what specific tools are used for the annotation of metabolomics or exposomics MS datasets, and irrespective of whether a separation technique is applied or not, the basic and complexities remain. Recorded *m/z* features must be converted to neutral molecular weights before they can be assigned a molecular formula or compound name. A single metabolite or chemical detected using MS can yield many different ion forms, such as adducts, isotopes, in-source fragments, and neutral losses, often referred to as degenerate features (339). This adds complexity to the process of determining the neutral mass of the analyte in question. Various aspects of chemical knowledge can be used to help the determination of the neutral masses detected in various ion forms. One such example is that certain ion species are more commonly observed using different ionisation methods. For example, ESI commonly yields protonated (positive ion-mode) and deprotonated (negative ion-mode) ions. At the most basic level, it can be assumed that all *m/z* observed in positive ion-mode are the protonated forms of their neutral compounds [M+H], or the deprotonated forms for negative ion-mode [M-H]. As such, adding or removing the mass of the proton gives the neutral mass. It is far more complex than this as compounds ionise to yield many different ion forms. However, the response of each ion form is still expected to reflect the concentration of the compound in question within the sample. This means that, for example, a compound at a given concentration, that forms two adducts at a 50: 50 ratio, will have two *m/z* features of that compound that have equal intensities. If a dilution series was created of this sample, then the intensities of the two *m/z* features would decrease equally with increased dilution, and the intensity ratios of these *m/z* feature pairs would be equal. This response-relationship between ion forms is often exploited through correlation analysis to determine features that belong to a single compound, which improves the false discovery rate, and to improve accuracy of neutral mass calculations. Unfortunately, the ratio of different ion forms of the same compound can be impacted by sample preparation and ionisation parameters, and therefore does not always behave as expected. Nonetheless, neutral masses are calculated using some variation of this strategy, either with or without correlation analysis. Annotation of different ion forms is also often carried out. This is done by calculating *m/z* defects between suspected ions, such as calculation of the difference between the protonated and sodiated ion forms and matching against a reference list of commonly observed *m/z* defects.

With a neutral mass in hand, the next step is either assignment of a compound name or molecular formula. Some chemistry knowledge can be used to determine a molecular formula. A given elemental composition has a specific molecular weight (mass). Given this, calculating the neutral mass from MS data can give clues about the elemental composition of the measured analyte. Unfortunately, a given mass (molecular weight) does not have a single unique elemental composition, and thus a single calculated neutral mass from the MS data can yield several hundred candidate elemental compositions, although some heuristic rules have been proposed to limit the elemental compositions selected as candidates as described here (340). The candidate molecular formulae calculated can then be matched against a relevant database or reference list, thus yielding annotation of mass spectral data. Some approaches do not compute the elemental composition, opting instead to match the neutral mass against the masses listed in the reference list or database directly. Indeed, sometimes neutral masses are matched against molecular formula reference lists, and then molecular formulae matched against metabolite or chemical reference lists or databases. The database or reference list selected is therefore of utmost importance. Analysis of a biological sample and subsequent annotation against a reference list of chemicals will yield results, but these will likely be predominantly incorrect as the database does not contain biological compounds. Moreover, a database high in isomers as most biological databases are, will result in tens to hundreds of annotations for a single neutral mass. (nano)ESI-DIMS methods are particularly challenged by this, as they lack pre-separation of analytes prior to detection.

Indeed, the use of a separation technique adds great value to the annotation of MS data. Separation techniques such as U(H)PLC-MS separate analytes based on physicochemical properties, thus yield orthogonal information that can aid in the differentiation of isomers (although isomers are not always resolved using chromatography). Moreover, different ion forms of the same compound are expected to elute simultaneously, and therefore RT information can be used to improve determination and annotation of degenerate features. However, to be of most value, databases must exist that contain both the molecular weight information and RT. Such databases, called RT databases for U(H)PLC-MS studies, are challenging to construct as RTs are heavily reliant on instrument setup and plumbing (arrangement of tubing and capillaries) and therefore cannot be shared easily between laboratories.

1.6.5.1 Levels of Identification

Because metabolomics (and exposomics) data contain many isomers, assignment of compound name to MS1 data is referred to as putative annotation, where a single peak can be named as multiple compounds that share a common molecular weight. This decreases the level of confidence in MS1 annotations. Indeed, levels of confidence in annotation and identification have been defined by the Metabolomics Standards Initiative (313). These span from level 1-4, with 1 being the highest achievable level of confidence (Table 6.

A separate but similar system defining levels of confidence in annotation and identification exists. Here, there are 5 levels (Figure 20). Experimental design must consider the level of confidence that the data will yield, MS1, MS2, or both.

Level of Identification	**Requirements for Qualification**
1	An unambiguous identification using 2 or more orthogonal properties of data collected compared to data for the standard collected in the same lab using the same analytical set up.
2	A putative annotation of a compound based on the physicochemical properties of the compound displayed in the data.
3	A putative annotation to a class of compounds based on the physicochemical properties of the compound displayed in the data.
4	Unknowns

Table 6: Table summarising the levels of identification in metabolomics as described by (313). Level 1 identification requires 2 or more orthogonal analytical characteristics of an authentic chemical standard be analysed, and that experimental data be collected using these same methods and compared to the authentic chemical standard data (341). Level 1 is the only one referred to as identification and defines compound identity that involves the use of two of more types of data collected for authentic chemical standards applying the same analytical assays. Any other work must be called annotation. Level 2 and 3 are compound name assignments made based on the physicochemical properties of the collected data, but not matched against data collected for authentic chemical standards applying the same analytical assay, and thus offering less confidence. Level 2 is assignment to a specific candidate compound, whilst level 3 is assignment to a chemical class. Level 4 are compounds simply unknown.

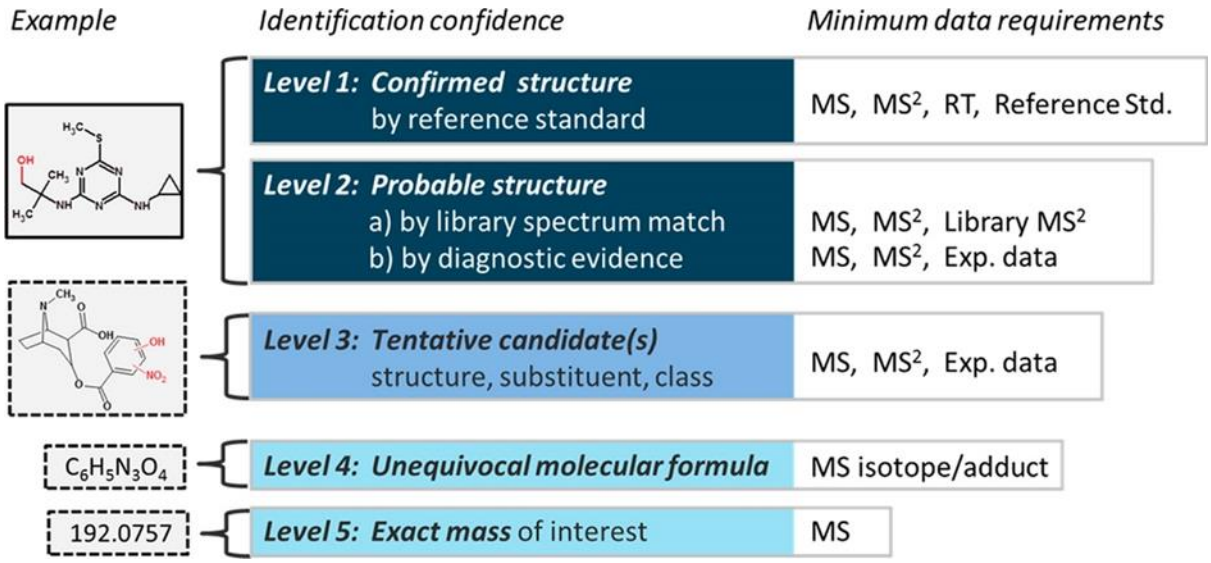


Figure 20: Figure taken from (314), showing levels of identification for metabolomics and exposomics studies. Level 1 requires the use of authentic chemical standards, and results in a confirmed chemical structure. The second level, however, has been expanded into 2 subsections (a) and (b). Matching against a library spectrum (i.e., the spectrum of a compound analysed previously and stored in a repository called a library for reference in later studies), albeit not against an authentic chemical standard, offers Level 2(a) annotation, where a chemical structure is probable as the data match literature and other reference libraries. Level 2(b) arises where no other candidate matches the data, there is no literature or standard to confirm, but diagnostic data such as fragments and parent ion information exists (314). Where multiple candidates exist but insufficient information is available to further narrow down the options, this is level 3. Unambiguous assignment of a molecular formula is level 4, and level 5 is peaks simply unknown (314).

1.6.6 Statistical Data Analysis

Metabolomics and exposomics studies require statistical analyses to identify which peaks in a data matrix are associated with the biological question being asked, i.e., are statistically significant features or metabolites. Statistical analyses can be classified into univariate methods or multivariate methods.

Univariate analyses consider one variable at a time, and do not look at relationships or associations between different features. These analyses are still useful in NTA, comparing each signal between classes one at a time. Univariate analyses use traditional statistical methods such as t-tests, comparing the mean or medians of a single variable present in 2 sample classes. They can be used for hypothesis testing. Here, a null hypothesis is required, which states that there is no difference between, for example, the peak intensities of a single *m/z* between a group of control and a group of experimental samples. A probability threshold

is then set at which the null hypothesis can be rejected when it is in-fact true (342). This threshold is the chance that a type I error, or FP, will occur. Probabilities of rejecting the null hypothesis (p-values) are then calculated. If below the 5% threshold, the null hypothesis is rejected, and it is concluded that there is a statistically significant difference between the single variable of interest in the sample classes in question. If above the threshold, the null hypothesis is accepted, and it is concluded that there is no statistically significant difference. Selecting the appropriate univariate statistical test depends on the type of data collected and the experimental design (342). There are tests that work best for parametric, or normally distributed data, whilst others work best for non-parametric, or skewed data. Moreover, data can be paired or unpaired. Paired data considers the difference between the same group under different experimental conditions, whilst unpaired data considers the difference between two different and unrelated groups. A summary of univariate tests best applied for parametric and non-parametric data is shown in Table 7.

A major drawback in univariate testing is that non-targeted datasets are multivariate in nature often having thousands of peaks. Repeated testing for significance of each variable increases the chances of type I errors. This is the multiple testing problem (343) and can be resolved by calculating the false discovery rate (FDR) associated with the results, (344,345) and/or application of a multiple testing correction such as the Bonferroni correction (346).

Experimental design	**Normal distribution (parametric)**	**Skewed distribution (non-parametric)**
	Compare Means	**Compare Medians**
Compare two unpaired groups	Unpaired t-test	Mann-Whitney U test
Compare two paired groups	Paired t-test	Wilcoxon signed rank test
Compare more than two unmatched groups	One-way ANOVA with multiple comparison	Kruskal-Wallis one-way ANOVA
Compare more than two matched groups (repeated measures)	Repeated-measures ANOVA	Friedman test

Table 7: Table adapted from (342). A summary of statistical tests that can be applied to normally distributed and skewed data.

Multivariate analysis (MVA) methods consider multiple variables simultaneously. Additional to considering the means and medians of the data, these types of analyses consider covariances and correlations within the data (347). A multivariate statistical analysis aims to give a summarised version of all (or a subset) of the variables in a large dataset into fewer simulated variables. This is done by applying combinations of variables from the original dataset that explain the largest amount of variability amongst measurements. MVA can be unsupervised or supervised, with the former being useful for exploratory work whilst the latter is guided by a designated response variable. The unsupervised MVA look for unknown patterns within the entire dataset, irrespective of class. The supervised MVA are assigned a response variable and thus look for relationships associated with the response. Examples of unsupervised MVA include Principal Component Analysis (PCA) (348) and Correspondence Analysis (CA) (349). Since these methods compress the level of multi-dimensionality of the data representation, they are often referred to as "dimension reduction methods". PCA is the most commonly used multivariate statistical analysis in omics datasets, offering advantages by catering to both continuous and discrete data, and can be applied to variables with both similar and different scales (centred and scaled PCA respectively) (350,351). According to M. Chadeau-Hyam, the technique does have a pitfall in assuming that the principal components (PCs) represent all diversity in the data, neglecting to factor in inter versus intra data variation (352). This can be minimised by using supervised MVA such as Discriminant Analysis (DA) (353), which attempts to find PCs based on a larger ratio of inter versus intra group variation (352).

Any MVA statistical analyses must be cross validated (354,355). This is because when data are too complex, as often NTA datasets are, containing biological/xenobiotic information, and perhaps different time points for data collection and multiple biological replicates, overfitting can occur. This means that the statistical analysis starts mapping noise peaks as real peaks and modelling relationships from noise rather than real data. As such, the results derived lose their generalisability, meaning they are specific only to the dataset tested and cannot be used to generalise across populations (e.g., smokers vs non-smokers) (356–359).

Multivariate analyses were not applied in the work presented in this thesis; thus, in-depth description is beyond its scope.

1.7 Project Aims and Objectives

1.7.1 Research Chapters 3 and 4

The University of Birmingham (UoB) metabolomics team collaborated with the US Environmental Protection Agency (US-EPA) on the work presented in the first two research chapters of this thesis (sections 3 and 4). A global ring trial involving 20 industrial and academic research groups was designed by the US-EPA to investigate the suitability of various non-targeted methods used globally for the analysis of xenobiotics in increasingly more complex sample matrices. The chemicals investigated were deemed by the EPA to have exposure potential, both for humans and the environment. The experiment was divided into two phases. The first was a blinded phase, in which the composition of chemical mixtures in different matrices was unrevealed prior to analysis, followed by an unblinded phase where sample compositions were revealed. This experimental design allowed for investigations into how method performance could be improved when information, or indeed suspected information, was known about the samples being analysed. As such, the overarching project aims are as follows:

- Four in-house nano(ESI)-DIMS and U(H)PLC-MS metabolomics methods will be applied for the analysis of 10 chemical standard mixtures in DMSO. These same methods will also be applied for analysis of a pair of serum, house-dust, and wrist-band extracts, to determine if these assays can be applied for the detection of xenobiotics in increasingly complex matrices. In the experimental design, an unknown standard mixture was selected from the 10 ENTACT standard mixtures in DMSO and spiked into one of each pair of the matrices. Analysis with the four (nano)ESI-DIMS and U(H)PLC-MS assays will be used to determine which of each pair of matrices (serum, house-dust, and wrist-band extracts) is fortified, and with which of the 10 ENTACT standard mixtures.
    - For analysis of xenobiotics in DMSO, method success will be determined through calculation of true and false positive rates of annotation for each of the 10 ENTACT standard mixtures upon revelation of sample compositions at the unblinded stage of experiments.
    - For analysis of xenobiotics in environmental (house-dust and wristband extracts) and biological (serum) matrices, method success will be determined through correct determination of which of each pair of sample matrices has

been fortified, and with which of the 10 ENTACT standard mixtures. To determine which sample is fortified and with which standard mixture, true and false positive rates of annotation will be calculated for each pair of samples (e.g., serum 1 and serum 2), assuming in sequence that each of 10 ENTACT standard mixtures has been spiked in. The standard mixture yielding the highest true positive rate will be assumed to be the correctly fortified standard mixture.

- At the blinded stage of experiments, a large database containing 4, 462 chemicals will be used for annotation, whilst at the unblinded stage, smaller databases containing only chemicals revealed to be in each standard mixture will be used for annotation. True and false positive rates of annotation will be calculated for both stages and compared against each other to determine if the use of smaller databases is effective at reducing false positive rates of annotation, and if this effectiveness persists in increasingly complex sample matrices.

1.7.2 Research Chapter 5

Data processing parameters such as feature intensity correlation coefficients (r), calculated using regression statistics, and annotation search parameters such as adduct, and isotope reference lists can play a crucial role in false positive annotation. Although much research has gone into optimising data processing parameters applied in metabolomics, there is no clear consensus on best practice, whilst primary focus has not been on annotation of MS1 data, but rather on MS2 spectral identifications. The Birmingham mEtabolite Annotation for Mass Spectrometry (BEAMS), a software tool dedicated to annotation of metabolomics U(H)PLC-MS and DIMS data, will be used to characterise the impact of changing annotation parameters in research chapter 3 (section 5), including the length of various annotation reference lists in relation to true and false positives. These optimised parameters will be recommended for use within the ACMG.

The following overarching objectives will be addressed in chapter 5 (section 5):

- A metabolomics untargeted analysis study of serum samples will be used for evaluation of BEAMS annotation parameters. Four assays will be evaluated, including HILIC and Lipids positive and negative ion-modes. A subset of these datasets,

containing MS/MS identifications obtained through in-house library matching, will be used to calculate true and false positive rates.

    - For each of these four assays, the maximum retention time difference, correlation threshold, and mass error tolerances will be optimised.

- For each of four assays, default adduct, isotope, and neutral loss lists will be applied, followed by longer customised adduct, isotope, and neutral loss lists created using an R-package to determine which one of these is optimal.

2 CHAPTER 2: Methods and Materials

2.1 Materials and Instrumentation

All materials and instruments used for all data collected in this thesis are listed in Table 8 and Table 9, respectively.

Materials
Optima U(H)PLC Grade Acetonitrile (ACN) (Fisher Chemicals)
Optima U(H)PLC Grade Water (Fisher Chemicals)
Optima U(H)PLC Grade Isopropanol (IPA) (Fisher Chemicals)
Optima U(H)PLC Grade Methanol (MeOH) (Fisher Chemicals)
Optima U(H)PLC Grade Chloroform ($CHCl_3$) (Fisher Chemicals)
U(H)PU(H)PLC-MS Grade Formic Acid (FA) (Fisher Chemicals)
Ammonium Formate (Fisher Chemicals)
LC-MS Grade Acetic Acid (AA) (Fisher Chemicals)
LC-MS Grade Ammonium Acetate (Fisher Chemicals)
HILIC Accucore Amide 100mm, 2.1mm i.d., 2.6 µm particle size (Thermo Fisher Scientific)
Hypersil Gold aQ Column 100mm, 2.1mm i.d., 1.9 µm particle size (Thermo Fisher Scientific)
Hypersil Gold Column 100mm, 2.1mm i.d., 1.9 µm particle size (Thermo Fisher Scientific)
Polypropylene 384 Well Plates (Thermo Fisher Scientific)
(nano)ESI silicone chips, 2.5µm nozzle emitter size (Advion)
(nano)ESI pipette tips (Advion, cat. no. 1004763)
LC Vials (Chromatography Direct), 10, 100, 200, and 1 000µL Gilson pipettes
2-, 7-, and 28-mL Glass Vials (Scientific Supplied Limited)
5, 10, 100, 200, and 1 000µL pipette tips (VWR)
100, 200, 500, 1 000mL conical flasks, measuring cylinders, and Duran bottles (VWR)
1.5 mL Eppendorfs (Fisher Chemicals)
Easy-pierce heat-sealing foil covers for well plates (Thermo Scientific)

Table 8: Materials used for all methods in the work presented in this thesis are listed.

Instruments
Vortex mixer (Fisher Scientific, cat. no. 13214789)
Refrigerated centrifuge (Microcentrifuge Biofuge Primo R (Thermo Scientific, cat. no. 75005440) with bucket rotor (Thermo Scientific, cat. no. 75007591) and fixed-angle rotor (Thermo Scientific, cat. no. 75007593))
SpeedVac sample concentrator, including ultra-low temperature vapour trap (Thermo Scientific Savant, cat. no. RVT5105230), SpeedVac (Thermo Scientific Savant, cat. no. SPD111V230) and pump (KNF Laboport, cat. no. N 820.3 FT.18)
Well plate heat sealer (Thermo Scientific, cat. no. ALPS 50 V)
High-resolution, high-mass-accuracy Fourier transform mass spectrometer, Thermo Scientific Q Exactive Hybrid Quadrupole Orbitrap mass spectrometer
High-resolution, high-mass-accuracy Fourier transform mass spectrometer, Thermo Scientific LTQ-Orbitrap Elite Hybrid Ion Trap-Orbitrap mass spectrometer
Thermo Fisher Scientific Vanquish LC Core System

Table 9: Instruments used for data collection of data presented in this thesis (216).

2.2 Assessing the Performance of (nano)ESI-DIMS During the Non-targeted Analysis of Xenobiotics in Different Matrices as Part of the US-EPA's Non-Targeted Analysis Collaborative Trial (ENTACT)

2.2.1 Preparation of ToxCast Standard Mixtures by Evotec and Fortified Reference Materials by EPA Laboratories

Chemical selection was discussed and agreed upon by EPA staff, considering availability, and exposure potential (360). Chemicals were selected from the ToxCast chemicals list (361). Sample preparation was performed as described in (360) and is summarised next. For the standards, "each ToxCast stock solution was nominally 20mM in dimethyl sulfoxide (DMSO); the final concentration of each substance varied depending on the concentration in the ToxCast stock solution, but most were at the nominal concentration. A total of 7.3mL of each mixture was created by adding 20µL of each requested ToxCast stock solution into a clear glass vial. Each mixture was then diluted to the final volume in DMSO for a final nominal concentration of 0.05 mM for each mixture constituent; the DMSO volume added varied (0–5400µL) depending on the number of substances included in the mixture. EPA prepared 30 aliquots of 100µL of each mixture using an Agilent 7696 Sample Prep Workbench (Santa Clara, CA) and a solvent blank using a different stock of DMSO than was used to prepare the mixtures.

The reference materials (serum, house-dust, and silicone wristbands, were prepared as described in (360), and methods are summarised next. "Extracts of house dust were prepared at EPA laboratories (Research Triangle Park, NC) using SRM 2585 Organic Contaminants in

House Dust from NIST (Gaithersburg, MD, USA). Twenty aliquots of 300 mg ± 10 mg dust were weighed in Falcon tubes (Becton Dickinson, Franklin Lakes, NJ). Ten of the tubes were spiked with 10µL of a nominally 0.05mM ENTACT mixture 7. Four empty test tubes were included as method blanks to undergo the same procedure as the dust samples. Methanol was added to each sample until the 13mL mark. Samples were vortexed for approximately 1min, until the dust at the very bottom tip could be seen moving in solution. Samples were then sonicated for 30min and vortexed again for approximately 1min. They were then centrifuged at 10,000 rpm for 10min. Aliquots of 4mL were applied to precleaned 3cm^3 liquid chromatography/silica (LC-Si) cartridges (Supelco, Bellefonte, PA, USA) and the eluent was collected. An additional 2 mL of methanol was added for further elution. Samples were evaporated to approximately 0.5mL under gentle nitrogen except blank samples which were evaporated to 1.0–1.25mL. Samples of the same type (spiked, unspiked, or blanks) were combined and the volume of each was adjusted to 15mL with methanol. Samples were stored at − 20 °C. Prior to preparing aliquots to send to trial participants, and after storing in the freezer for 3 days, a precipitate was noticed so the samples were centrifuged again to remove solids. The supernatant was poured into a new tube and the volume adjusted to 15 mL with methanol (less than 2 mL was needed). The final dust samples and blanks were distributed as 400µL aliquots. Aliquots were stored at − 20 °C prior to shipment to ENTACT participants".

"Serum extracts were prepared at EPA laboratories by reconstituting NIST SRM 1957 Organic Contaminants in Non-fortified Human Serum in 10.7mL deionized water. Twenty-six aliquots of 750µL serum each were added to Falcon tubes. Thirteen samples were spiked with 10µL of a nominally 0.05mM ENTACT mixture 1. Three empty tubes were included for method blanks to undergo the same procedure as the serum samples. The samples received 1500µL of 0.1 M formic acid and were vortexed for approximately 1min. The samples then received 10mL cold acetonitrile (kept at −20 °C until used) and were vortexed briefly. They were centrifuged at 10,000 rpm for 10min, and the supernatant was poured into new tubes. Samples of the same type (spiked, unspiked, or blanks) were combined and the volume of each was adjusted to 19.5mL (except the blank which was adjusted to 15mL). Prior to preparing the aliquots, and after storing in the freezer for 3 days, a precipitate was noticed so the samples were centrifuged again to remove solids. The supernatant was poured into a new tube and the volumes were adjusted to their original volume before centrifugation. The final serum

samples and blanks were distributed as 400µL aliquots. Aliquots were stored at − 20 °C prior to shipment".

"Silicone band extracts were prepared at Oregon State University (Corvalis, OR). Sixteen silicone bands were cleaned by a water rinse and thermal conditioning. Cleaned bands were stored in airtight jars or bags until use. Bands were then deployed as passive air samplers in a semi-rural outdoor environment for 18 days. After the sampling period, bands were sealed and transported in polytetrafluoroethylene (PTFE) bags and stored at − 20 °C until further processing. Before extraction, bands were cleaned by sequential rinses in high purity deionized water and isopropanol then placed individually in extraction jars for dialysis. Eight bands were spiked with 20µL of a nominally 0.05mM ENTACT mixture 5 by applying the mixture to the surface of each band. To remove the DMSO solvent from the ENTACT mixture, spiked bands were sealed in a glass jar and heated for 15min then cooled to room temperature. All bands were spiked with 500ng each of ten isotopically labelled and three non-labelled standards (DTXSID indicates the substance identifier in the DSSTox database (vide infra)): naphthalene-D8 (DTXSID10894058), acenaphthylene-D8 (DTXSID00109466), acenaphthene-D10 (DTXSID40893473), phenanthrene-D10 (DTXSID60893475), fluoranthene-D10 (DTXSID20893476), chrysene-D12 (DTXSID00893474), benzo[a]pyrene-D12 (DTXSID00894062), benzo[ghi]perylene-D12 (DTXSID40894066), polychlorinated biphenyl (PCB) 100 (DTXSID8073504), PCB 209 (DTXSID4047541), 9-fluorenone-D8 (DTXSID60894068), 2-methyl-1,4-naphthalenequinone-D8 (DTXSID90703033), and tetrachlorometaxylene (DTXSID6075433). Each band was then submerged in 100mL of ethyl acetate and placed on an orbital shaker for at least 2h. The extraction solvent was removed, the extraction was repeated, and the two extraction solvents were combined. The volume was then reduced to 1mL using a large volume closed cell TurboVap (Biotage, Charlotte, NC) and a small volume nitrogen blowdown TurboVap. Samples of the same type (spiked, unspiked, and blanks) were combined and the volume of each was adjusted to approximately 12mL. Extracts were shipped overnight to the EPA lab and were kept frozen at − 20 or − 80 °C until aliquots were prepared. The final samples were distributed as 400µL aliquots. Samples were stored at − 20 °C until shipment. Band blank samples from a different stock of ethyl acetate were prepared at EPA laboratories using 400µL of ethyl acetate dispensed into the same vial type as samples, using the same pipette and tip stock".

2.2.2 Sample Shipping, Receipt and Storage

All samples were shipped from the US-EPA's main site to the University of Birmingham on dry ice at -78°C. Upon receipt, samples were immediately transferred to a -80°C freezer for storage. Summary information about the samples shipped is shown in Table 10.

Sample Type	**Sample Concentration**	**Carrier Solvent**	**Volume Sent**	**Sample Storage**
Mixture of Standards	0.05mM per chemical	Dimethyl sulfoxide (DMSO)	100µL	-80°C Freezer
Serum extract spiked with single std mixture	100pg/µL per 100g/mol chemical	Acetonitrile (ACN)	400µL	-80°C Freezer
House-dust extract spiked with single std mixture	100pg/µL per 100g/mol chemical	Methanol (MeOH)	400µL	-80°C Freezer
Wrist-band extract spiked with single std mixture	100pg/µL per 100g/mol chemical	Ethyl Acetate ($C_4H_8O_2$)	400µL	-80°C Freezer
Serum blank	N/A	Acetonitrile (ACN)	400µL	-80°C Freezer
House-dust blank	N/A	Methanol (MeOH)	400µL	-80°C Freezer
Wrist-band extract blank	N/A	Ethyl Acetate ($C_4H_8O_2$)	400µL	-80°C Freezer

Table 10: Summary of standards, serum, house-dust, and wristband extracts shipped to the University of Birmingham Metabolomics and Systems Toxicology Laboratory by the US Environmental Protection Agency (US-EPA) as part of the EPA's Non-Targeted Analysis Collaborative Trial (ENTACT).

2.2.3 Preparation of Re-suspension Solvents for (nano)ESI-DIMS Analysis

2.2.3.1 Polar Metabolite Assay - Positive and Negative Ion-modes

For the positive ion-mode assay, 15mL of 4:1 (v/v) methanol: water was prepared, then 37.5µL of the solution removed and replaced with 37.5µL of formic acid to create a formic acid concentration of 0.25%. For the negative ion-mode assay, a stock solution of 100mM aqueous ammonium acetate was prepared by adding 0.077g of ammonium acetate to 10mL water. 40mL of methanol was then added, and the solution shaken to mix and give a 4:1 (v/v) methanol: water solution with 20mM final concentration of ammonium acetate.

2.2.3.2 Non-polar Metabolite Assay - Positive and Negative Ion-modes

7.5mM methanolic ammonium acetate was prepared by adding 0.0058g of ammonium acetate to 10mL of methanol shaking until the ammonium acetate was fully dissolved. Addition of 5mL of chloroform gave a total volume of 15mL 2:1 (v/v) methanol: chloroform and an ammonium acetate concentration of 5mM.

2.2.4 Preparation of ToxCast Standard Mixtures for (nano)ESI-DIMS Dilution Studies

The ten standard mixtures and DMSO blank were removed from -80°C storage, allowed to thaw on ice for 1h and then vortex-mixed for 30s. Two dilution series were created, one for a polar assay and another for a non-polar assay, each re-suspended in the appropriate solvent for its corresponding method (section 2.2.3). Ten 15µL aliquots of each standard mixture and one 15µL aliquot of the DMSO blank were placed into 1.5mL Eppendorf tubes for the polar assay and 1.5mL glass vials for the non-polar assay. All aliquots were then dried in a vacuum concentrator (Thermo Scientific Savant) until no visible solvent remained (ca. 3h). Each aliquot was then re-suspended in 15µL of the appropriate solvent for each assay (section 2.2.3). From 15µL of each standard mixture, 10µL aliquots were taken and pooled into a single 1.5mL Eppendorf tube for the polar assay or 1.5mL glass vial for the non-polar assay to give a total volume of 100µL in a pooled mixture. The blank aliquot was re-suspended in 1,500µL of solvent to give a 100-fold diluted blank sample, which is equal to the smallest dilution in the standards' dilution series. Re-suspended standards and blanks were vortex-mixed for 30s each.

Dilutions were carried out in 1.5mL Eppendorf tubes for the polar assay, and 1.5mL glass vials for the non-polar assay, and preparation volumes are shown in Table 11. All sample preparation was carried out on ice.

From the 100µL pooled standard mixtures (~5µM), 10µL was taken and added to a 1.5mL vial labelled for 100-fold dilution. 990µL of the relevant solvent (section 2.2.3) was added to give 1mL of ~50nM pooled standard mixture. This was vortex-mixed for 30s. From the 1mL of 100-fold dilution standard mixture, 100µL was taken and added to vial labelled 300-fold dilution. 200µL of the relevant solvent (section 2.2.3) was added to give 300µL of 16.67nM pooled standard mixture. This was vortex-mixed for 30s. From the 300µL of 300-fold dilution standard mixture, 100µL was taken and added to vial labelled 1,000-fold dilution. 233µL of the relevant solvent (section 2.2.3) was added to give 333µL of 5nM pooled standard mixture. This was vortex-mixed for 30s. From the 333µL of 1 000-fold dilution standard mixture, 100µL was taken and added to a vial labelled 3,000-fold dilution. 199µL of the relevant solvent (section 2.2.3) was added to give 299µL of 1.67nM pooled standard mixture. This was vortex-mixed for 30s. From the 299µL of 3 000-fold dilution standard mixture, 100µL was taken and added to a vial labelled 10,000-fold dilution. 234µL of the relevant solvent (section 2.2.3) was added to give 334µL of ~500pM pooled standard mixture. This was vortex-mixed for 30s. All polar assay dilution mixes were centrifuged (Microcentrifuge Biofuge Primo R, Thermo Scientific) at 4°C for 10min with 20,000g and non-polar assays at 4°C for 10min with 2,500g.

Initial concentration (nM)	Volume from initial dilution (µL)	Solvent added (µL)	Dilution factor	Final concentration (nM)
5,000 (stock pooled standard mixtures)	10	990	100	50
50	100	200	300	16.67
16.67	100	233	1,000	5
5	100	199	3,000	1.67
1.67	100	234	10,000	0.5

Table 11: Dilutions were prepared as shown in the table and described above.

After centrifugation, the dilutions for each respective method were pipetted into a 384 polypropylene well plate. For the polar positive ion-mode method 7µL were plated and for the non-polar positive ion-mode method 10µL were plated per sample well. Each sample dilution and blank sample was plated into 4 separate wells to create 4 technical replicates. The plate was sealed with aluminium foil to prevent evaporation (Thermo Scientific) once plated using a plate sealer (Thermo Scientific) set to 170°C.

2.2.5 Preparation of Fortified and Unfortified SRMs for (nano)ESI-DIMS Dilution Studies

The fortified and unfortified serum, house-dust, and wristband extracts, and respective solvent blanks were removed from -80°C storage, allowed to thaw on ice for 1h and then vortex-mixed for 30s each. Two dilution series were created, one each for a polar assay and for a non-polar assay, each re-suspended in the appropriate solvent for its corresponding method (section 2.2.3). Dilution series were created for each sample, with fortified and unfortified treated as separate samples. 50µL aliquots of each sample and 15µL aliquots of each respective solvent blank were placed into 1.5mL Eppendorf tubes for the polar assay and 1.5mL glass vials for the non-polar assays. All aliquots were then dried in a vacuum concentrator (Thermo Scientific Savant) until no visible solvent remained (ca. 1.5h) and then re-suspended in 50µL of the appropriate solvent for each assay (section 2.2.3). The blank aliquot was re-suspended in 150µL of the appropriate assay solvent to give a 10-fold diluted blank sample, which is equal to the second smallest dilution in the sample dilution series. Undiluted blanks could not be used as there would be insufficient volumes for all planned experiments. All re-suspended samples and blanks were then vortex-mixed for 30s each.

Dilutions were carried out in 1.5mL Eppendorf tubes for the polar assay, and 1.5mL glass vials for the non-polar assay, and sample volumes are shown in Table 12. All sample preparation was carried out on ice.

From the 50µL of each sample stock (~1µM), 40µL was taken and added to a 1.5mL vial labelled for no-dilution. From the remaining 10µL stock for each sample type, 5µL was taken and added to a vial labelled for 10-fold dilution. 45µL of the relevant solvent (section 2.2.3) was added to give 50µL of ~100nM of each sample. This was vortex-mixed for 30s. From the

50µL of 10-fold dilution sample, 5µL was taken and added to a vial labelled 100-fold dilution. 45µL of the relevant solvent (section 2.2.3) was added to give 50µL of ~10nM of each sample. This was vortex-mixed for 30s. From the 50µL of 100-fold dilution sample, 5µL was taken and added to vial labelled 1,000-¬fold dilution. 45µL of the relevant solvent (section 2.2.3) was added to give 50µL of ~1nM of each sample. This was vortex-mixed for 30s. From the 50µL of 1,000-¬fold dilution sample, 5µL was taken and added to a vial labelled 10,000-fold dilution. 45µL of the relevant solvent (section 2.2.3) was added to give 50µL of ~0.1nM of each sample. This was vortex-mixed for 30s. All polar assay dilution mixes were centrifuged (Microcentrifuge Biofuge Primo R, Thermo Scientific) at 4°C for 10 mins with 20,000g and non-polar assays at 4°C for 10 mins with 2 500g.

Initial concentration (nM)	Volume from initial dilution (µL)	Solvent added (µL)	Dilution factor	Final concentration (nM)
1,000 (stock serum, house-dust, or wristband)	40	0	0	1,000
1,000 (stock serum, house-dust, or wristband)	5	45	10	100
100	5	45	100	10
10	5	45	1,000	1
1	5	45	10,000	0.1

Table 12: Dilutions were prepared as shown in the table and described above.

After centrifugation, the dilutions for each respective method were pipetted into a 384 polypropylene well plate. For the polar positive ion-mode method 7µL were plated and for the non-polar positive ion-mode method 8µL were plated per sample well. Each sample

dilution and blank sample was plated into 4 separate wells to create 4 technical replicates. The plate was sealed with aluminium foil to prevent evaporation (Thermo Scientific) once plated using a plate sealer (Thermo Scientific) set to 170°C.

2.2.6 Preparation of ToxCast Standard Mixtures for (nano)ESI-DIMS Analysis

The 10 standard mixtures and DMSO blank were removed from -80°C storage, allowed to thaw on ice for 1h and then vortex-mixed for 30s each. 5µL aliquots of each standard mixture and the DMSO blank were placed into 1.5mL Eppendorf tubes for the polar assay and 1.5mL glass vials for the non-polar assay. All aliquots were then dried in a vacuum concentrator (Thermo Scientific Savant) until no visible solvent remained (ca. 1h) and then re-suspended in 500µL (both the polar and non-polar assays) of the appropriate solvent for each assay (section 2.2.3). Re-suspended standards and blank were then vortex-mixed for 30s each. All samples were centrifuged (Microcentrifuge Biofuge Primo R, Thermo Scientific) at 4°C for 10 minutes. The polar positive ion-mode samples were centrifuged (Microcentrifuge Biofuge Primo R, Thermo Scientific) at 20,000g in Eppendorf tubes whilst the non-polar positive ion-mode ones were centrifuged at 2,500g in glass vials.

After centrifugation, samples and blanks were pipetted into a 384 polypropylene well plate. For the polar positive ion-mode method 7µL were plated and for the non-polar positive ion-mode method 10µL were plated per sample, as per the optimised methods. Each sample and blank were plated into 4 separate wells to create 4 technical replicates. The plate was sealed with aluminium foil (Thermo Scientific) once plated using a plate sealer (Thermo Scientific) set to 170°C.

2.2.7 Preparation of Fortified and Unfortified SRMs for (nano)ESI-DIMS Analysis

The fortified and unfortified serum and house-dust extracts alongside their respective solvent blanks were removed from -80°C storage, allowed to thaw on ice for 1h and then vortex-mixed for 30s each. 5µL aliquots of each sample and respective solvent blank were placed into pre-labelled 1.5mL Eppendorf tubes for the polar assays and 1.5mL glass vials for the non-polar assays. All aliquots were then dried in a vacuum concentrator (Thermo Scientific Savant)

until no visible solvent remained (ca. 1h) and then re-suspended in 50µL for the serum samples and their respective blank (for both the polar and non-polar assays) and 500µL for the house-dust samples and their respective blank (for both the polar and non-polar assays) of the appropriate solvent for each assay (section 2.2.3). All re-suspended samples and blanks were then vortex-mixed for 30s each.

All samples were centrifuged (Microcentrifuge Biofuge Primo R, Thermo Scientific) at 4°C for 10 minutes. The polar positive ion-mode samples were centrifuged (Microcentrifuge Biofuge Primo R, Thermo Scientific) at 20,000g in Eppendorf tubes whilst the non-polar positive ion-mode ones were centrifuged (Microcentrifuge Biofuge Primo R, Thermo Scientific) at 2,500g in glass vials.

After centrifugation, the samples and respective blanks were pipetted into a 384 polypropylene well plate. For the polar positive ion-mode method 7µL were plated and for the non-polar positive ion-mode method 10µL were plated per sample and blank. Each sample and blank were then plated into 4 separate wells to create 4 technical replicates. The plate was sealed with aluminium foil (Thermo Scientific) once plated using a plate sealer (Thermo Scientific) set to 170°C.

2.2.8 Advion Triversa NanoMate

2.2.8.1 Polar and Non-polar Metabolite Assays - Positive and Negative Ion-modes

The Advion Triversa Nanomate was used for infusion of samples for (nano)ESI-MS analysis, operated by a software called Chipsoft. The Triversa method parameters used for data collection of the polar and non-polar samples are detailed in Table 13 and in (216).

Parameter	Polar Metabolite Assay-Negative Ion-mode Settings	Polar Metabolite Assay - Positive Ion-mode Settings	Non-polar Metabolite Assay-Negative Ion-mode Settings	Non-polar Metabolite Assay -Positive Ion-mode Settings
Sample aspiration volume	5µL	5µL	7µL	7µL
Return unused sample to well	no	no	no	no
Vent headspace	Toggle on	Toggle on	Toggle on	Toggle on
Aspirate air after sample	Toggle on	Toggle on	Toggle on	Toggle on
Volume of air to aspirate	0.25µL	0.25µL	1.5µL	1.5µL
Air gap before chip	yes	yes	yes	yes
Contact closure	0s	0s	0s	0s
Contact closure	after	after	after	after
Voltage timing delay	0s	0s	0s	0s
Equalisation delay	0s	0s	0s	0s
Aspiration depth	0.8mm	0.8mm	0.8mm	0.8mm
Pre-piercing	yes	yes	yes	yes
Mandrel pre-piercing depth	5.2mm	5.2mm	5.2mm	5.2mm
Pre-wetting	yes	yes	yes	yes
Pre-wetting mix repeat	1	1	1	1
Sample acquisition duration	3 mins 12s	3 mins 12s	3 mins 12s	3 mins 12s
Trigger acquisition when input signal received	Toggle on	Toggle on	Toggle on	Toggle on
Input signal	DigIn2	DigIn2	DigIn2	DigIn2
Gas pressure	0.3psi	0.3psi	0.3psi	0.3psi
Voltage	1.7kV	1.5kV	1.4kV	1.4kV
Ion-mode	negative ion-mode selected	positive ion-mode selected	negative ion-mode selected	positive ion-mode selected
Output contact closure	Rel1	Rel1	Rel1	Rel1
Duration	2.5s	2.5s	2.5s	2.5s
Spray sensing	off	off	off	off

Table 13: (nano)ESI is facilitated by an ESI chip, and an automated sample delivery system called the Triversa Nanomate. The Triversa Nanomate is controlled by ChipSoft software. Settings for the (nano)ESI-DIMS ChipSoft methods are shown for polar and non-polar positive and negative assays.

2.2.9 Thermo Scientific Orbitrap Elite Hybrid Ion Trap-Orbitrap mass spectrometer SIM-Stitch Methods

2.2.9.1 Polar and Non-polar Metabolite Assays - Positive and Negative Ion-modes

The MS polar and non-polar metabolite assay methods were divided into 2 MS segments and an associated tune method. These are summarised below in Table 14 to Table 17.

	Parameter	**Polar Metabolite Assay-Negative Ion-mode Settings**	**Polar Metabolite Assay - Positive Ion-mode Settings**	**Non-polar Metabolite Assay-Negative Ion-mode Settings**	**Non-polar Metabolite Assay - Positive Ion-mode Settings**
MS segment 1	source type	API	API	API	API
	type	Full-scan	Full-scan	Full-scan	Full-scan
	m/z range	50-620	50-620	190-1200	190-1200
	FT resolution	240,000	240,000	240,000	240,000
	duration	0.5min	0.5min	0.5min	0.5min

Table 14: LTQ-Orbitrap Elite method parameters for the first segment of the (nano)ESI-DIMS metabolomics methods. The MS methods are divided into segments. The first segment collects full scan data (50-620 for the polar assays and 190-1200 for the non-polar assays). This segment's data are not used for data analysis as the SIM data is preferred for its advantages in improving sensitivity. The first segment is added simply to allow equilibration of electrospray at the start of each sample analysis.

	Parameter	Polar Metabolite Assay- Negative Ion-mode Settings	Polar Metabolite Assay - Positive Ion-mode Settings	Non-polar Metabolite Assay- Negative Ion-mode Settings	Non-polar Metabolite Assay - Positive Ion-mode Settings
MS segment 2	scan events	10	10	18	18
	FT resolution	240,000	240,000	240,000	240,000
	duration	0.5min	0.5min	0.5min	0.5min
	m/z ranges	Each scan event was a 75 *m/z* wide window overlapping by 20 *m/z* with the next window. E.g., 50-125 *m/z* followed by 105-180 *m/z*	Each scan event was a 75 *m/z* wide window overlapping by 20 *m/z* with the next window. E.g., 50-125 *m/z* followed by 105-180 *m/z*	Each scan event was a 75 *m/z* wide window overlapping by 20 *m/z* with the next window. E.g., 190-265 *m/z* followed by 245-320 *m/z*	Each scan event was a 75 *m/z* wide window overlapping by 20 *m/z* with the next window. E.g., 190-265 *m/z* followed by 245-320*m/z*

Table 15: LTQ-Orbitrap Elite method parameters for the second segment of the (nano)ESI-DIMS metabolomics method*s*. The MS method*s are* divided into segments. The first segment collects full scan data (50-620 for the polar assay*s* and 190-1200 for the non-polar *assays*). This segment's data are not used for data analysis as the SIM data is preferred for its advantages in improving sensitivity. The first segment is added simply to allow equilibration of electrospray at the start of each sample analysis. The second segment of the MS methods contains SIM windows that are 75*m/z* wide and overlap by 20*m/z* either side. Each SIM window is a separate scan event. The polar metabolite assays have 10 scan events and the non-polar metabolite assays have 18 scan events.

	Parameter	Polar Metabolite Assay-Negative Ion-mode Settings	Polar Metabolite Assay - Positive Ion-mode Settings	Non-polar Metabolite Assay-Negative Ion-mode Settings	Non-polar Metabolite Assay - Positive Ion-mode Settings
Tune method	Source type	NSI	NSI	NSI	NSI
	Capillary temperature	200°C	200°C	200°C	200°C
	Injection waveforms	on	on	on	on
	Ion trap zoom AGC target	4,000	4,000	4,000	4,000
	Ion trap full AGC target	30,000	30,000	30,000	30,000
	Ion trap SIM AGC target	8,000	8,000	8,000	8,000
	Ion trap MS_n AGC target	10,000	10,000	10,000	10,000
	FTMS injection waveforms	on	on	on	on
	FTMS full and SIM AGC targets	500,000	500,000	500,000	500,000
	FTMS MS_n AGC target	50,000	50,000	50,000	50,000
	Polarity	negative	positive	negative	positive
	Source voltage	1.5kV	1.5kV	1.5kV	1.5kV
	Source current	100µA	100µA	100µA	100µA
	S-lens RF level	67.3%	68.65%	68.5%	68.5%
	Multipole 00 offset	-2.03V	3.60V	-0.75V	3.60V
	Lens 0 voltage	-3.43V	5.25V	-9.72V	5.25V
	Multipole 0 offset	-9.75V	11.01V	-9.72V	11.01V
	Lens 1 voltage	-14.51V	13.32V	-13.60V	13.32V
	Gate lens offset	-90.00V	90.00V	-90.00V	90.00V
	Multipole 1 offset	-16.10V	14.66V	-11.96V	14.66V
	Front lens	-9.76V	11.60V	-9.38V	11.60V

Table 16: LTQ-Orbitrap Elite method parameters for the MS for (nano)ESI-DIMS metabolomics methods. Tune parameters are shown for the polar and non-polar methods.

	Parameter	Polar Metabolite Assay-Negative Ion-mode Settings	Polar Metabolite Assay - Positive Ion-mode Settings	Non-polar Metabolite Assay-Negative Ion-mode Settings	Non-polar Metabolite Assay -Positive Ion-mode Settings
Tune method	Ion zoom micro-scans	1	1	1	1
	Ion trap zoom max ion time	50ms	50ms	50ms	50ms
	Ion trap full micro-scans	1	1	1	1
	Ion trap full max ion time	1000ms	10ms	1000ms	10ms
	Ion trap SIM micro-scans	10	1	10	1
	Ion trap SIM max ion time	1000ms	50ms	1000ms	50ms
	Ion trap MSn micro-scans	4	1	4	1
	Ion trap MSn max ion time	100ms	100ms	100ms	100ms
	FTMS full micro-scans	1	1	1	1
	FTMS full max ion time	1000ms	10ms	1000ms	10ms
	FTMS SIM micro-scans	10	1	10	1
	FTMS SIM max ion time FTMS	1000m	50ms	1000ms	50ms
	MS^n micro-scans	1	1	1	1
	FTMS MS^n max ion time	100ms	100ms	100ms	100ms

Table 17: LTQ-Orbitrap Elite method parameters for the MS for (nano)ESI-DIMS metabolomics methods. Tune parameters are shown for the polar and non-polar methods.

2.2.10 Galaxy Data Processing

Prior to data processing, the (nano)ESI-DIMS data file sizes were manually inspected. All files that had noticeably smaller file sizes than the average file size for the same assay type and ion-mode as viewed using Thermo Fisher Scientific's "Qual Browser" in Xcalibur were deleted. The Galaxy-M (nano)ESI-DIMS workflow (1.6.3.1) was applied for data processing. It comes pre-installed in a virtual machine (VM) that can be downloaded from the GigaDB repository (Linux username = galaxym; Linux password = galaxym; Galaxy username = galaxym@galaxym.org; Galaxy password = galaxym; both case studies are available as published histories and published workflows in the Galaxy installation, or in the 'galaxym' user's private history) (302). Source code and installation instructions can be viewed in GitHub (available at: *https://github.com/viant-metabolomics*).

The first step was to sum scans across each selected ion monitoring (SIM) window. The Hann function was then applied to the data, followed by zero-filling and Fourier transformation of data from the time to the frequency domain. Peak-picking above a signal-to-noise ratio (SNR) of 10 was carried out, followed by calibration of each SIM window to convert the frequency domain to *m/z* measurements and stitching together of SIM windows (218,302). Next, replicate filtering (removal of features found in < 66%, or 2 out of 3 of the technical replicates per sample) was applied. *m/z* values were aligned using a 2ppm error window across the technical replicates. Since 4 technical replicates were run, the combination of the 3 replicates giving the lowest median peak intensity relative standard deviation (RSD) were used. The median peak intensity RSD was calculated as the standard deviation of a peak divided by the mean peak intensity multiplied by 100. A 2ppm peak alignment across all samples was then carried out, followed by removal of background features. Background features were removed if they appeared in the blank samples and if they were at least 10-fold more intense within the samples.

2.2.11 Metabolite Identification Package (MI-Pack)

Peak annotation was carried out using the metabolite identification package (MI-Pack) (301). The ToxCast library of 4,462 chemicals was used as a suspect screening database, provided by the EPA as part of the ENTACT project. This was employed for matching, screening through all 4,462 chemicals for a match. A 1ppm error matching window was used, and peak pattern matching was set to yes to allow for isotope pattern searching. For positive ion-mode, $[M+H]^+$,

$[M+Na]^+$, and $[M+K]^+$ adducts were searched for, and ^{13}C, ^{34}S, ^{37}Cl and ^{41}K isotopes patterns were searched for. For negative ion-mode, $[M-H]^-$, $[M+Na-2H]^-$, $[M+K-2H]^-$and $[M+Hac-H]^-$ were searched for and the same list of isotopes as for positive ion-mode was used.

2.2.12 ToxCast Library

The US-EPA provided the Toxcast library that contained 4,462 chemicals (details about ToxCast library in section 3.1). Included in this list were 1,940 chemicals contained in the standards and spiked into the samples, making this database a suspect screening reference list. The database contained molecular formula, monoisotopic mass, chemical name, chemical ID (in the form of DTXSIDs, which indicate the substance identifier in the much larger EPA's DSSTox database), InChI, SMILES, and MS-ready molecular formula. This database was used for matching.

2.2.13 Calculation of True and False Positive Rates

To calculate the TPR, the number of correctly named compounds (each unique DTXSID identifier was considered as a correctly named compound) for each standard mixture was divided by the number of compounds known to be in each standard mixture and then multiplied by 100. Individual ion forms (where one metabolite was detected as multiple adducts for example) were not taken into consideration, only the number of unique compound IDs was considered. Isomers were present in some of the standard mixtures. Where a single *m/z* was annotated as being 2 (or more isomers) correctly or incorrectly, these annotations were all counted as being correct and incorrect and they contributed to the true and false positive calculation. False positive rates were calculated as the "number of incorrectly named compounds/the number of compounds known to be in each mixture x 100".

2.3 Assessing the Performance of U(H)PLC-MS During the Non-targeted Analysis of Xenobiotics in Different Matrices as Part of the US-EPA's Non-Targeted Analysis Collaborative Trial (ENTACT)

2.3.1 Preparation of Stock Standard Mixtures and Fortified and Unfortified Reference Materials

Standard mixtures and fortified and unfortified reference materials were selected, prepared, extracted, and shipped to participating laboratories by the US-EPA or Evotec, a company that manages the US-EPA's ToxCast library (362), as described in sections 2.2.1.

2.3.2 Preparation of Mobile Phases for C_{18} Aqueous Reversed Phase (aqRP C_{18}) and HILIC Analysis

For the aqRP C_{18} mobile phases, 2L of 100% methanol, and 2L of 100% water, both with 0.1% formic acid, were prepared, shaken to mix, and the solutions sonicated with the sonicator set to degas for 10 minutes. These were used for the aqueous reversed phase (aqRP C_{18}) positive and negative ion-modes.

The HILIC mobile phases were prepared as follows. To prepare a stock solution of ammonium formate 6.3056g of ammonium formate was weighed, added to 500mL of HPLC-grade water and shaken to dissolve the ammonium formate and make a 200mM stock solution. For mobile phase A, a solution of 95:5 acetonitrile: 10mM aqueous ammonium formate with ~0.1% formic acid was prepared. For mobile phase B, a solution of 50:50 acetonitrile: water with 10mM ammonium formic and ~0.1% formic acid was prepared. Both mobile phases were shaken to mix, and sonicated with the sonicator set to degas for 10 min. These mobile phases were used for the HILIC positive ion-mode assays.

To prepare a stock solution of ammonium acetate, 7.7083g of ammonium acetate was weighed and added to500mL of HPLC-grade water to make a 200mM stock solution of aqueous ammonium acetate. Preparation of mobile phases here was identical to HILIC positive ion mode mobile phases but using ammonium acetate instead, to give 95:5 acetonitrile: 10mM aqueous ammonium acetate and ~0.1% acetic acid for mobile phase A, and 50:50 acetonitrile: water with 10mM ammonium acetate and ~0.1% acetic acid for mobile phase B. These were used for the HILIC negative ion-mode assays.

2.3.3 Preparation of Resuspension Solvents for C_{18} Aqueous Reversed Phase (aqRP C_{18}) and HILIC Analysis

For the aqRP C_{18}, dried samples were resuspended (see details on resuspension volumes in 2.3.4 and 2.3.5) in 1:1 methanol: water (v/v). For the HILIC, dried samples were resuspended were resuspended (see details on resuspension volumes in 2.3.4 and 2.3.5) in 1.5:1.5:1 acetonitrile: methanol: water (v/v/v).

2.3.4 Preparation of ToxCast Standard Mixtures for U(H)PLC-MS Analysis

Ten standard mixtures and a dimethyl sulfoxide (DMSO) solvent blank prepared at the same time as the standard mixtures were placed on ice and allowed to thaw for ~1 h, or until completely thawed. They were then each vortex-mixed for 1 min. Aliquots of up to 10 µL of each standard mixture and DMSO blank were then transferred to 1.5mL Eppendorf tubes and diluted in solvent in the range of 200-400µL. Aliquot volumes relied on remaining sample volume after the research performed in 2.2 and are shown in Table 18. Each of the 11 aliquots were dried in a vacuum concentrator (Thermo Scientific Savant) for ~ 3 h. If after this time solvent was still observed in the Eppendorf tubes, standard mixtures were dried for an additional 1 h. Samples were then stored in a -80°C freezer until the day of U(H)PLC-MS analysis.

On the day of analysis, samples were placed on ice and resuspended in the appropriate resuspension solvent for each assay (see sections 2.3.3 for resuspension solvents). Each standard in the ten standard mixtures was prepared to a final concentration of ~1µM, and resuspension volumes to achieve this are shown in Table 18. Resuspended standard mixtures and blank were vortex-mixed for 1 min and then centrifuged (Microcentrifuge Biofuge Primo R, Thermo Scientific) at 4°C for 20min at 20,000g. 50µL aliquots of each standard were transferred into U(H)PLC vials. Each standard mixture and blank had 2 LC vials prepared, one for MS1 only data collection and one for only MS/MS data collection. Samples from each of these two vials would be analysed twice, once in positive once in negative ion-modes for each assay.

Standard	aQRP_Aliquot Volume (µL)	aQRP_Aliquot Resuspension Volume (µL)	HILIC_Aliquot Volume (µL)	HILIC_Aliquot Resuspension Volume (µL)
Blank	10.0	400	10.0	400
1	10.0	400	10.0	400
2	10.0	400	10.0	400
3	10.0	400	8.0	320
4	5.0	200	5.0	200
5	10.0	400	10.0	400
6	10.0	400	7.0	280
7	10.0	400	10.0	400
8	5.0	200	5.0	200
9	5.0	200	5.0	200
10	10.0	400	10.0	400

Table 18: Table showing the aliquot and resuspension volumes of ten standard mixtures and a blank prepared at the same time as the standard mixtures. Volumes are shown for both the aqueous reversed phase and HILIC assays alongside resuspension volumes to create ~1µM concentrations per chemical in each standard mixture.

2.3.5 Preparation of Fortified and Unfortified Standard Reference Materials (SRMs) for U(H)PLC-MS Analysis

2.3.5.1 Serum

Two serum samples extracted in acetonitrile (see 2.2.1), one fortified and one unfortified, and a corresponding extraction blank prepared at the same time as the serum samples, were placed on ice and allowed to thaw for ~1h, or until completely thawed. Samples were then vortex-mixed for 1 min.

To prepare samples for the HILIC positive and negative ion-mode assays, 10µL aliquots were transferred to 1.5mL Eppendorf tubes. 65µL of acetonitrile was added, then 75µL of methanol, then 50µL of water (dilution factor of 20, concentration of 2.5pg/µL for a 100g/mol chemical, and a total volume of 200µL). All samples were then vortex-mixed for 1 min each and centrifuged (Microcentrifuge Biofuge Primo R, Thermo Scientific) at 4°C for 20min with 20,000g. Four aliquots of each sample, and 2 aliquots of the blank, were transferred (40µL aliquots of each sample) into LC vials and placed in a LC rack and into the LC autosampler ready for analysis.

To prepare samples for the aqueous C_{18} reversed phase assays, 10µL aliquots of each sample were transferred to 1.5mL Eppendorf tubes. Samples were then dried in a vacuum

concentrator (Thermo Scientific Savant) for ~3 h, or until samples were completely dry. The samples and blank were then removed from the vacuum concentrator, placed on ice, and 200µL of 50:50 methanol: water (v/v) added to create a 20-fold dilution and concentration of 2.5pg/µL for a 100g/mol chemical of the samples and blank. Samples were then centrifuged (Microcentrifuge Biofuge Primo R, Thermo Scientific) at 4°C for 20min with 20,000g. Four 40µL aliquots of each sample, and 2 aliquots of the blank, were then transferred to LC vials, and samples placed on an LC rack and into the LC autosampler ready for analysis.

2.3.5.2 House-dust

Two house-dust samples extracted in methanol (see 2.2.1), one fortified and one unfortified, and a corresponding extraction blank prepared at the same time as the house-dust samples, were placed on ice, and allowed to thaw for ~1h, or until completely thawed. Samples were then vortex-mixed for 1 min. 10µL of each sample were aspirated and added to separate, appropriately labelled 1.5mL Eppendorf tubes.

To prepare samples for the HILIC positive and negative ion-mode assays, 65µL of methanol was added to each sample and blank, then 75µL of acetonitrile, and then 50µL of water to give a total of 200µL of each sample and blank. The samples, diluted 20-fold, had a concentration of 2.5pg/µL for 100g/mol chemical. The samples and blank were then centrifuged (Microcentrifuge Biofuge Primo R, Thermo Scientific) at 4°C for 20min at 20,000g. Four aliquots of each sample, and 2 aliquots of the blank, were then prepared by aspirating 40µL of each sample into LC vials, and prepared samples were placed on an LC rack and into the LC autosampler ready for analysis.

To prepare samples for the aqueous C_{18} reversed phase positive and negative ion-mode assays, 10µL aliquots of each sample and blank were dried in 1.5mL Eppendorf tubes in a vacuum concentrator (Thermo Scientific Savant) for ~3 h, or until samples were completely dry. Dried samples were dissolved in a resuspension solvent of 50:50 methanol: water (v/v) of volume 200µL to give a 20-fold dilution. The samples had a concentration of 2.5pg/µL for a 100g/mol chemical. Samples were then centrifuged (Microcentrifuge Biofuge Primo R, Thermo Scientific) at 4°C for 20min at 20,000g. Four aliquots of each sample, and two aliquots

of the blank, were then prepared by transferring 40µL of each sample into LC vials. Prepared samples were placed on an LC rack and into the LC autosampler ready for analysis.

2.3.5.3 Wristbands

Two wristband samples extracted in ethyl acetate (see 2.2.1), one fortified and one unfortified, and a corresponding extraction blank prepared at the same time as the wristband samples, were, placed on ice, and allowed to thaw for ~1h, or until completely thawed. Samples were then vortex-mixed for 1 min. 10µL aliquots of each sample were transferred to 1.5mL Eppendorf tubes.

To prepare samples for the HILIC assays, the 10µL aliquots of samples and blank were dried in a vacuum concentrator (Thermo Scientific Savant) for ~3 h, or until completely dry. Once dry, samples were dissolved in 200µL of 1.5:1.5:1 acetonitrile: methanol: water (v/v/v) to give a 20-fold dilution and concentration of 2.5pg/µL for a 100g/mol chemical. Samples were then vortex-mixed for 1 min and centrifuged (Microcentrifuge Biofuge Primo R, Thermo Scientific) at 4°C for 20min at 20,000g. Four aliquots of each sample, and 2 aliquots of the blank, were then prepared by transferring 40µL of each sample into LC vials. Prepared samples were placed on an LC rack and into the LC autosampler ready for analysis.

To prepare samples for the aqRP C_{18} assays, 10µL aliquots of samples and blank were dried in a vacuum concentrator (Thermo Scientific Savant) for ~3 h, or until completely dry. Once dry, samples were dissolved in 200µL of 1:1 methanol: water (v/v) to give a 20-fold dilution and concentration of 2.5pg/µL for a 100g/mol chemical. Samples were then vortex-mixed for 1 min and centrifuged (Microcentrifuge Biofuge Primo R, Thermo Scientific) at 4°C for 20min with 20 000g. Four aliquots of each sample, and 2 aliquots of the blank, were then prepared by transferring 40µL into LC vials. Prepared samples were placed on an LC rack and into the LC autosampler ready for analysis.

2.3.6 Data acquisition

Data were collected on an electrospray Q-Exactive Plus mass spectrometer coupled to a Vanquish U(H)PLC system (Thermo Fisher Scientific, San Jose, USA). Data were collected in MS1 and MS2 modes for each of the four assays applied, HILIC positive and negative ion-mode

assays and aqRP C_{18} positive and negative ion-mode assays. Methods presented are for the MS1 experiments and MS2 data, but results presented in this chapter thesis focus on MS1 data only. MS2 data were collected for each of the 4 assays as follows:

a) An MS1 blank injection was used to create an exclusion list, and dynamic exclusion was applied.
b) MS1 injection of the standard mixtures was used to generate inclusion lists, and dynamic exclusion was applied using a loop count of 3.
c) All MS2 data were collected using stepped collision energies. Three NCEs (%) were applied for this. For the positive ion-mode = 20, 40, 100 and for the negative ion-mode =40, 60, 130.
d) All MS2 data were collected using 3 x DDA windows [100-200, 200-300, 300-900]. Each DDA window was a separate LC injection.

2.3.6.1 U(H)PLC-MS Methods

2.3.6.1.1 HILIC Positive and Negative Ion-modes

The HILIC positive and negative ion-mode LC assays were performed using a Thermo Fisher Scientific Accucore-Amide-HILIC U(H)PLC column (100mm x 2.1 mm, 2.6µm) and a 15 min gradient elution. The column temperature was set to 35°C, and the flow rate was operated at 0.5mL/min. Sample injection volume was 2µL. The HILIC gradient elution method is described in Table 19.

No	RT (min)	Flow Rate (mL/min)	% Solvent A (organic)	% Solvent B (aqueous)	Curve
1	0.0	0.5	99	1	5
2	1.0	0.5	99	1	5
3	3.0	0.5	85	15	5
4	6.0	0.5	50	50	5
5	9.0	0.5	5	95	5
6	10.0	0.5	5	95	5
7	10.5	0.5	99	1	5
8	15.0	0.5	99	1	5

Table 19: The LC gradient method for the HILIC positive and negative ion-mode assays is shown.

The HILIC positive and negative ion-mode MS assays collected MS1 data with a scan range of 70-1050 *m/z.* The method duration was 14 min, the AGC target was 1 x 10^6, the maximum injection time was 100ms, "micro-scans" was set to 1, and the resolution was set to 70,000 (FWHM at *m/z* 200). Ion source settings are described in Table 20.

Ion Source Settings	
Sheath gas flow rate	55
Aux gas flow rate	14
Sweep gas flow rate	4
Spray voltage (kV)	Pos =3.2 Neg =2.7
Capillary temp (°C)	380
S-lens RF level	30
Aux gas heater temp (°C)	440

Table 20: ESI settings are shown for the HILIC positive and negative ion-mode assays.

2.3.6.1.2 Aqueous C_{18} Reversed Phase Positive and Negative Ion-mode

The aqueous reversed phase (aqRP C_{18}) positive and negative ion-mode assays were carried out using a Thermo Fisher Scientific Hypersil GOLD aQ column (100 x 2.1 mm, 1.9 µm) and a 15 min gradient elution. The column temperature was set to 45°C and the flow rate was operated at 0.3mL/min for sample analysis and 0.4mL/min for equilibration. Sample injection volume was 2µL. The aqueous C_{18} reversed phase gradient method is described in Table 21.

Step	RT (min)	Flow Rate (mL/min)	% Solvent A	% Solvent B	Curve
1	0.0	0.3	99	1	5
2	0.5	0.3	99	1	5
3	2.0	0.3	50	50	5
4	9.0	0.3	1	99	5
5	10.0	0.3	1	99	5
6	10.5	0.3	99	1	5
7	15.0	0.3	99	1	5

Table 21: The LC gradient method for the aqueous C_{18} reversed phase positive and negative ion-mode assays is shown.

The aqueous reversed phase positive and negative ion-mode MS assays collected MS1 data with a scan range of 100-1500 *m/z.* The method duration was 15 min, the AGC target was 1 x 10^6, the maximum injection time was 100ms, "micro-scans" was set to 1, and the resolution was set to 70,000 (FWHM at *m/z* 200). Ion source settings are demonstrated in Table 22.

Ion Source Settings	
Sheath gas flow rate	30
Aux gas flow rate	13
Sweep gas flow rate	0
Spray voltage (kV)	Pos =3.2 Neg =2.7
Capillary temp (°C)	350
S-lens RF level	40
Aux gas heater temp (°C)	400

Table 22: ESI settings for the aqueous reversed phase positive and negative ion-modes.

2.3.7 Data Processing

2.3.7.1 XCMS

XCMS data processing in Galaxy-M used a node called "XCMS-Basic-Birmingham-Pipe". This pipeline reads individual spectra in mzML format, groups, and then aligns spectral features across samples using the R package "XCMS" (363). The XCMS pipeline is modular, and thus can be modified to add or remove steps. This gives the user some flexibility in the data processing applied and allows for custom processing of datasets.

Peak picking, grouping and alignment were carried out using XCMS (292). This was run on the Galaxy-M instance (300). Most parameters were used are applied widely in metabolomics. However, the "minfrac" used for these samples was set uncharacteristically low. This is because Phase I of ENTACT experiments involved analysis of samples of unknown composition. Overlap of chemical composition between each standard mixture was unknown, thus the "minfrac" parameter was set to allow a feature to only be present in a single standard mixture or fortified sample. Parameters for XCMS data processing are shown in Table 23.

XCMS Parameter	Value
SNR	5
Sigma	3
Prefilter	5, 1000
Minfrac	0.06
Noise	1000
Fitgauss	FALSE
Integrate	1
mzCenterFun	wMean

Table 23: XCMS processing parameters for HILIC and aqueous C_{18} reversed phase assays are shown.

2.3.8 Birmingham mEtabolite Annotation for Mass Spectrometry (BEAMS)

Data were analysed in the form of a peak intensity matrix. Annotation of U(H)PLC-MS datasets was carried out using the Birmingham mEtabolite Annotation for Mass Spectrometry (BEAMS) software. Features were grouped across samples, peak pattern annotation was carried out by computing adducts, isotopes and neutral losses, and database matching was carried out against the ToxCast database containing 4,462 chemicals. Parameters used for the BEAMS-based annotations are shown in Table 24.

Group Features					
Maximum RT Difference (s)	Grouping Method	Coefficient Threshold	p-value threshold	cpus	Block size
5	Pearson	0.7	0.05	7	5,000
Annotate Peak Patterns					
Adducts	Isotopes	Neutral Losses	Oligomers	Mass Tolerance (ppm)	
Yes (default list) pos=[M+H], [M+Na], [M+K] neg=[M-H], [M+Cl], [M+Na-2H], [M+K-2H], [M+Hac-H]	Yes (default list) pos=^{13}C, ^{34}S, ^{41}K neg=Cl^{37}, ^{13}C, ^{34}S, ^{41}K	Yes (default list) H_2O, CO, NH_3, C_2H_2, C_2H_4, CO_2, C_2H_4O	No	5	
Annotate Compounds/Metabolites					
Reference List	Mass Tolerance (ppm)				
Yes (ToxCast database uploaded as the reference list)	5				

Table 24: Parameters used for BEAMS annotation are shown. Data were grouped, peak patterns were annotated as adducts, isotopes, or neutral losses, and compounds were annotated using the ToxCast database as a reference list.

2.3.9 ToxCast Database

Annotation of ENTACT samples at Phase I was carried out by matching against the ToxCast database (361) provided to all participating laboratories, containing 4,462 chemicals using the BEAMS software and *m/z* data only (see 2.2.12).

2.3.10 Targeted ToxCast Databases for Phase II Annotation

Annotation of chemicals present in ENTACT samples at Phase II was carried out against small databases containing only the chemicals revealed to be in each standard mixture using the BEAMS software and *m/z* data only. Additionally, a retention time (RT) knowledgebase was constructed by Dr Donna O'Neil within the ACGM in a Microsoft Excel format using TraceFinder 4.1 (Thermo Fisher Scientific) by searching for each of 1,940 chemicals in a targeted approach applying the [M+H]+, [M+Na]+, [M+K]+, and [M+NH4]+ adducts in positive ion-mode, and the [M-H]-, and [M+Cl]- adducts in negative ion-mode. An *m/z* precision of 4 decimal places and a mass tolerance of 100ppm was used. These results were then divided into their corresponding sample mixtures such that only compounds found in each standard mixture remained in the knowledgebase for each standard mixture. These knowledgebases were filtered by retaining only annotations that appeared in all three MS1 technical replicates for that assay. Thus, four small knowledgebases were created, containing only chemicals known to be in each of ten standard mixtures, and containing RT information, which was averaged across three MS1 injections. These knowledgebases were used for Phase II annotation. Phase II annotation was carried out using BEAMS and exact parameters detailed for Phase I experiments.

2.4 Optimising U(H)PLC-MS Metabolite Annotation Parameters for Full-Scan Data using the Birmingham mEtabolite Annotation for Mass Spectrometry (BEAMS) Tool

2.4.1 Participant recruitment and data collection

Experimental design, participant recruitment, and sample collection were carried out as part of a collaboration between Professors Warwick Dunn and Wiebke Arlt as described in (364). Blood serum samples were applied in a non-targeted U(H)PLC-MS metabolomics study performed by staff in Phenome Centre Birmingham (Prof. Warwick Dunn, Dr Lukas Najdekr, Dr Andris Jankevics, Dr and Dr Ralf Weber) and the study was led by Prof. Wiebke Arlt, Dr

Alessandro Prete and Dr Vasilis Chortis from the Institute of Metabolism and Systems Research at the University of Birmingham. The methods applied for data collection and raw data processing are identical to those reported in (365). The XCMS outputs for each of the four assays was applied and no filtering of the data after XCMS processing was performed. Metabolite identification to level 1 of the Metabolomics Standards Initiative recommendation (366) was performed using matching to in-house RT (match within +/- 5 seconds) and MS/MS mass spectral libraries (match score using Compound Discoverer v3.1 of greater than 70%) to define a sub-group of metabolites with high confidence metabolite identifications.

2.4.2 Annotation using the Birmingham mEtabolite Annotation of Mass Spectrometry (BEAMS) Tool

Annotation of metabolites detected in four U(H)PLC-MS datasets were performed using the Birmingham mEtabolite Annotation for Mass Spectrometry (BEAMS) software tool. Features were grouped across samples, peak pattern annotation was carried out by computing adducts, isotopes and neutral losses, and database matching was carried out against HMDB (75) for the two HILIC assays, and LipidMaps (367) for the two lipid assays. For each BEAMS annotation step, various parameters were optimised as described in Table 25.

BEAMS ANNOTATION STEP	Default Parameters	Tested Parameters
Maximum RT Diff (s)	5	0.2, 0.5, 1, 2, 5
Pearson Correlation (r, p)	r=0.7, p=0.05	r=0, 0.25, 0.5, 0.7, 0.9 p=0.05, 0.005, 0.0005
Mass error (ppm)	5	0.5, 1, 3, 5, 10, 25
Adducts	pos=$[M+H]^+$, $[M+Na]^+$, $[M+K]^+$ neg=$[M-H]^-$, $[M+Cl]^-$, $[M+Na-2H]^-$, $[M+K-2H]^-$, $[M+Hac-H]^-$	pos=$[M+H]^+$, $[M+Na]^+$, $[M+K]^+$, $[M+NH_4]^+$ neg=$[M-H]^-$, $[M+Cl]^-$, $[M+Na-2H]^-$, $[M+K-2H]^-$, $[M+Hac-H]^-$, $[M+Br]^-$, $[M+NH_4-2H]^-$, $[M+C_2H_4O_2-H]^-$, $[M+C_2HF_3O_2-H]^-$
Isotopes	pos=^{13}C, ^{34}S, ^{41}K neg=^{37}Cl	pos= ^{13}C, ^{33}S, ^{34}S, ^{41}K, ^{15}N, ^{17}O, ^{18}O neg= ^{37}Cl
Neutral Losses	H_2O, CO, NH_3, C_2H_2, C_2H_4, CO_2, C_2H_4O	H_2O, CO, NH_3, C_2H_2, C_2H_4, CO^2, C_2H_4O, C, O, Na, H_2, $CHCl_3$, CH_2O_2, HCOOK, HCOONa, $C_2F_3NaO_2$, NaCl, CH_3OH, H_3O_4P, C_2, CH_3CN, $NaNO_3$, HCl, $C_2H_4O_2$, $C_2HF_3O_2$

Table 25: The group features, annotate peak patterns, and the annotate compounds/metabolites steps applied during BEAMS annotation will be optimised. Parameters to be tested are shown in the table.

2.4.3 Creation of Customised Adduct, Isotope, and Neutral Loss Lists

A newly developed and currently unpublished R-package was created by Dr William Nash in the ACMG to characterise the *m/z* differences within each dataset, annotate these *m/z* differences against a reference list, and use this information to create customised adduct, isotope, and neutral loss lists to be used during BEAMS annotation.

2.4.4 Data Analysis of BEAMS Outputs

For each BEAMS parameter to be optimised, multiple runs were carried out, varying the parameter for optimisation. The parameters were optimised sequentially, starting with maximum RT difference, then correlation analysis, then mass error, and finally adduct, isotope, and neutral loss lists. The parameter found to be optimal at each step was used in optimisations of the subsequent steps, until all parameters had been optimised. This optimisation design assumed that all parameters were independent of one another (Figure 21).

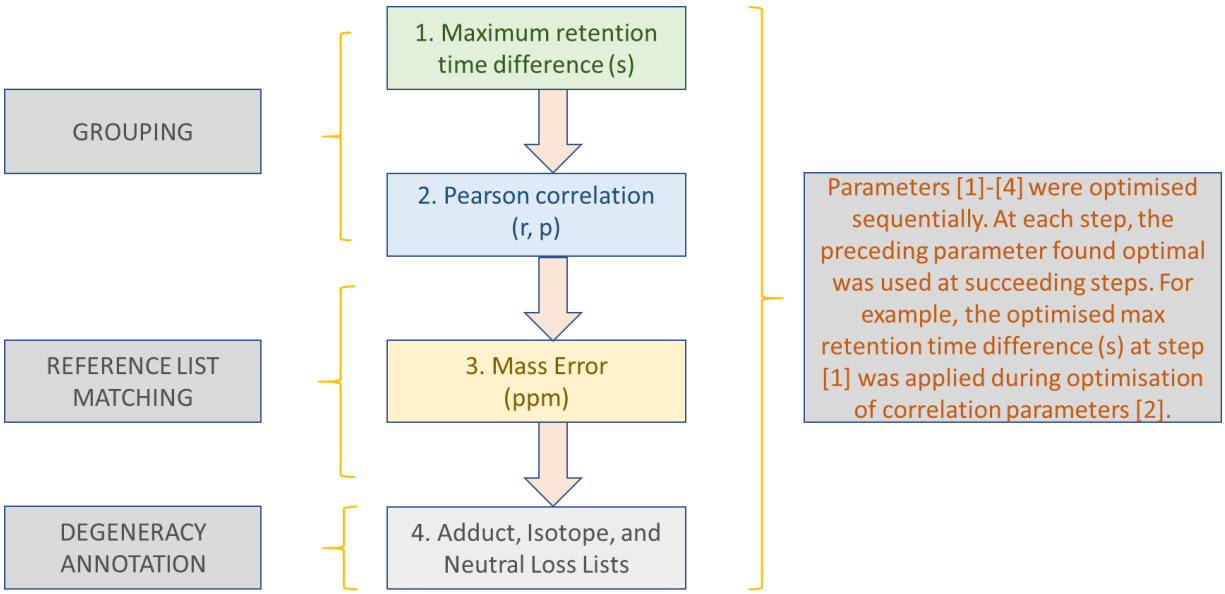


Figure 21: The optimisation of BEAMS parameters is summarised. Each parameter was optimised separately and independently of other parameters.

Data were then analysed in RStudio (368) version 4.1.0. Data counted within groups, for example counting of unique *m/z* in each RT group based on RT similarity, were carried out using the "dplyr" package in R, and all plots were created using the "ggplot2" package. Any data used to create boxplots was calculated using the "ggplot2" package. For all BEAMS outputs, a "group_id" and "sub_group_id" were assigned during annotation as demonstrated in Figure 30 and Figure 31. Only features with these values were used for all calculations and features not assigned a group and subgroup id were filtered out.

The key parameters used to filter the outputs included the name, *m/z*, rt, group_id, sub_group_id, and compound_id. The name field denotes a unique ID number for each unique *m/z* feature within the dataset (irrespective of RT). The *m/z* and rt fields are self-explanatory. The group_id and sub_group_id are important as all data were filtered first by presence of both of these, thus any data analysis carried out composed of only *m/z*-rt pairs that had both a group_id and sub_group_id. When correlation analysis is carried out on *m/z* intensities across all samples within a dataset in BEAMS, all *m/z* features that correlate above the user specified threshold are assigned a group_id. This simply reflects correlation but does not mean these *m/z* features belong to the same metabolite, i.e., assignment of a group_id does not mean features within this group are degenerate features of the same metabolite. Instead, assignment of the same group_id suggests that these *m/z* features are biologically related metabolites. The sub_group_id is assigned when two or more *m/z* features that correlate above the user specified threshold (i.e., *m/z* with the same group_id) have RT differences that fall within a user specified threshold. Not only this, but to be awarded the sub_group_id, these *m/z* features with similar RTs must also have mass differences that have been annotated as either an adduct, isotope, or neutral loss. The sub_group_id therefore represents degenerate features of the same metabolite.

2.4.5 Calculation of True and False Positive Rates

Identifications were made for some metabolites in each of the four U(H)PLC-MS datasets. This was done through MS1 annotation corroborated with MS2 spectral matching and RT matching to in-house libraries. For the HILIC positive ion-mode dataset 52 metabolites were identified, whilst for the HILIC negative ion-mode 27 metabolites were identified. For the lipids positive ion-mode 43 metabolites were identified, whilst for the Lipids negative ion-mode 29 metabolites were identified. These metabolites were to calculate true and false positive rates.

The TPR for each tested parameter was calculated by counting the number of true positives found after BEAMS annotation/the total number of true positives known to be in each dataset and multiplying that by a 100. For example, in the HILIC positive ion-mode dataset, the unique HMDB IDs for the 52 true positive identifications were searched for after each set of parameters were used for BEAMS annotation. The number of hits from this search were then

divided by 52 and this fraction then multiplied by a 100 to give a TPR (%) for this dataset. The false positive rate was calculated by creating a subset of all *m/z*-RT pairs known to be true positives and counting all other assigned HMDB or LM IDs that were incorrect. This number was divided by the number of true positive identifications and multiplied by 100. For example, the in the HILIC positive ion-mode dataset, a subset of all the *m/z*-RT pairs for each of the 52 correct identifications was created. The number of incorrect HMDB IDs assigned to each of these *m/z*-RT pairs was counted, divided by 52 and multiplied by 100.

3 CHAPTER 3: Assessing the Performance of (nano)ESI-DIMS During the Non-targeted Analysis of Xenobiotics in Different Matrices as Part of the US-EPA's Non-Targeted Analysis Collaborative Trial (ENTACT)

Abstract

Non-targeted analyses (NTA) are increasingly applied in exposomics, allowing more comprehensive analysis of the exposome. However, their reproducibility, detection limits, and chemical coverage are not yet known in this field. We participated in the US EPA's Non-Targeted Analysis Collaborative Trial (ENTACT) which aimed to determine if current direct-infusion mass spectrometry (DIMS) metabolomics methods can be used to screen for xenobiotics, determine the chemical coverage, and the false discovery rates of annotation.

Four (nano)ESI-DIMS methods (polar and non-polar positive and negative ion modes) were applied for the analysis of ten xenobiotic mixtures in DMSO, and fortified into serum, house-dust, and wrist-band matrices. Experiments were in two Phases. At Phase I, sample compositions were unknown to the analyst, and data annotation was achieved using a large suspect screening list containing 4,462 chemicals (ToxCast database). At Phase II, sample compositions were revealed, and this information was used to create smaller databases to be used for annotation, containing only the chemicals revealed to be in each standard mixture.

At Phase I, high TPRs (48-74%) were obtained for ten standard mixtures in a clean solvent matrix but FPRs were also high (up to 879%). TPRs decreased in the more complex sample matrices of serum and house-dust, whilst FPRs increased. The wrist-band extracts were not analysed because stable electrospray could not be achieved reproducibly. At Phase II experiments, the use of smaller databases (containing only chemicals known to be in the standard mixtures) increased TPRs and decreased FPRs relative to Phase I results for both the standard mixtures and the fortified serum and house-dust, whilst TPRs still declined with increasing sample complexity.

Overall, the methods could successfully be applied for analysis of xenobiotics, but false positive rates were exceedingly high. Alternative methods may be more suitable, and performance of U(H)PLC-MS methods will be tested in the next research chapter.

3.1 Introduction

Chemicals, or xenobiotics, have become ubiquitous in the environment, with ever-mounting evidence that they are having an adverse effect on both human and environmental health (369–371). As such, scientific tools are required to measure and quantify them as well as their impact on endogenous metabolites in different species. The gold standard for measuring xenobiotics in human and environmental samples is targeted analyses (TA) applying chromatography-triple quadrupole mass spectrometry. These assays use authentic chemical standards and where possible isotopically labelled internal standards for absolute quantification and confirmation of the chemical's structural identification. However, they are often limited to relatively few target compounds because chemical standards are often costly or are commercially unavailable, although analyses targeting pesticides have been known to detect numbers greater than 100 (372), even up to 300 (373). Non-targeted analyses (NTA) do not directly use chemical standards for quantification and confirmation of chemical structural identity, but instead are applied for 'screening' of samples for chemicals present in large reference lists while providing relative and not absolute quantification.

The type of analysis applied, whether targeted or non-targeted, depends on how much information is available about the xenobiotics. US Defence Attorney Donald Rumsfeld stated that "there are known knowns; there are things we know we know. We also know there are known unknowns; that is to say we know there are some things we do not know. But there are also unknown unknowns—the ones we don't know we don't know" (374).

In the context of chemical analysis, known knowns are xenobiotics about which information is available prior to analysis, thus assays can be custom developed to target these chemicals or chemical classes and standards can be purchased or synthesized for absolute quantitation and identification. Most pesticides fall into this category (375–377), alongside polyaromatic hydrocarbons (PAHs) (378), cyclosiloxanes, and parabenzoic acids (379), to name only a few classes. TA are most suitable for known knowns where the chemical targets are known and are expected to present in biological samples with a high confidence. Known unknowns are xenobiotics whose existence is known, but whose identities or presence in a sample prior to its analysis is not known with a high confidence. As such, authentic chemical standards cannot be purchased prior to analysis to develop a targeted assay and NTA are therefore more suited to screen for known unknowns. This is not to say known unknowns cannot become known

knowns. Indeed, the end-goal of non-targeted chemical analyses is full structural elucidation of chemicals, so NTA can also be referred to as non-targeted screening assays. Data are collected for all constituents of a sample and "screened" by matching against lists of xenobiotics and reference libraries. These matches, or a subset of them, can then be analysed using TA for absolute quantitation (380). Non-targeted metabolomics and exposomics experiments look for known unknowns. Unknown unknowns, on the other hand, are xenobiotics we do not even know exist currently, i.e., they are not currently listed anywhere. These have been termed new emerging compounds (NECs) (381–383). Since their existence is unknown, unknown unknowns are not present in lists of xenobiotics or reference databases and therefore will not be annotated using this screening strategy. The use of NTA to detect, annotate and then identify known unknown environmental contaminants is rising (384–388). However, the chemical coverage limitations, sensitivity and selectivity, and reproducibility have not been fully characterised (389). The US Environmental Protection Agency (EPA), dedicated to protecting both human and environmental health through regulation and legislation of chemicals and enforcement of chemical control rules, has a major focus on chemical pollutants and their potential adverse impacts. In 2006, the ToxCast project was launched with objectives to use computational chemistry, toxicogenomics (see 1.5) and high throughput screening (HTS) to predict, or forecast, the potential for chemical toxicity through observation of bioactivity induced by these chemicals based on their structures (361). Prioritised chemicals can then undergo full toxicity testing, thus using testing resources more efficiently (361). The chemicals initially selected as part of ToxCast were those for which toxicological information was available, including tumorigens, developmental and reproductive toxicants, neurotoxicants, and immunotoxicants (361). To test these chemicals, chemical standards were required, and each chemical added to the ToxCast list therefore had a chemical standard. Over nearly two decades, ToxCast's list of chemicals, and associated chemical standards, has grown from 310 at the start of the project (mostly pesticides), which was termed Phase I, to 4,200 unique chemicals at Phase III, which commenced in 2014 (362). Along the way, chemicals not suitable to forecast chemical toxicity have been removed from the list. The EPA also noted that a database with informative structure-activity relationship (SAR) (390) and toxicity data did not exist but could offer great utility within academic, government, and regulatory groups researching chemical toxicity (391). To this end, a distributed structure-searchable toxicity public database network (DSSTox) was proposed,

requesting input from any group that had sufficient data to add to this database. Currently, this database has approximately 740,000 chemicals. The database was curated manually at first until it had approximately 24,000 chemicals, and then curation using computational approaches to populate information from other data repositories was applied (392). The EPA therefore has a wealth of useful resources for investigating the limiting factors in the application of NTA for the measurement of xenobiotics.

Accordingly, the EPA organised a workshop in 2015 called "Advancing NTA of xenobiotic chemicals in environmental and biological media" (393). Its purpose was to bring together experts in NTA to present and discuss methods for data collection, analysis, interpretation, and sharing (360). Discussions from this workshop contributed to the genesis of a global ring trial involving over 20 academic and industry research groups to "extensively characterise the performance of NTA when measuring xenobiotics, to establish performance and benchmarks for NTA, to develop reporting standards, and increase the data available in current chemical reference libraries" (360).

The University of Birmingham's Metabolomics Laboratory participated in the ring trial, called the EPA's non-targeted analysis collaborative trial (ENTACT) (360). This study comprised of ten mixtures of chemical standards selected from the ToxCast chemicals list. These ten standard mixtures were prepared to contain between 95-365 chemicals each, with a total of 1,940 chemicals selected. Additionally, reference material extracts, representing three different matrices relevant to exposomics, were also prepared and spiked with one of the ten ToxCast mixtures (360). The content of these mixtures was not known by the participating laboratories on sample delivery and data collection, but the data were released later. Samples were prepared in DMSO at a concentration of ~0.05 mM per chemical. Two extracts of standard reference material (SRM) serum, house-dust, and silicone wristbands were prepared. One of each sample type was fortified with one of the unrevealed 10 mixtures of ToxCast chemicals. It was not known to participating laboratories which of two samples (henceforth referred to as serum 1 and 2, house-dust 1 and 2 and wristband 1 and 2) were fortified and with which of the ten standard mixtures. Participating laboratories performed an initial blinded analysis of these study samples (Phase I). The composition of each sample was revealed (Phase II) to participating laboratories only after initial analyses had been completed and all required documentation had been shared with the EPA project team. Computational or analytical

sample re-analysis could be carried out to improve Phase I results upon sample revelation if required. As such, results are presented for Phase I and II of ENTACT in this chapter and in chapter 4 (section 4) (358).

To aid data processing, participating laboratories were given full access to the DSSTox database, containing 740 000 chemicals (392), and the ToxCast library with 4,462 chemicals (362).

The overall aim of the research presented in this chapter was to evaluate the performance of (nano)ESI-DIMS non-targeted metabolomics workflows when analysing xenobiotics (ToxCast chemical mixtures) in increasingly more complex sample matrices relevant to exposomics. In chapter 4 we will describe a similar process when applying U(H)PLC-MS instead of (nano)ESI-DIMS. Method performance was assessed for both Phase I and Phase II of the experiments, and assessments were carried out on each sample matrix individually, considering the standards in solvent as a clean matrix, and then house-dust, serum, and wrist-band extracts as matrices of varying complexity relevant to exposomic studies. To assess method performance, the following specific research questions were considered for Phase I and then for Phase II:

In each of four different and complementary (nano)ESI-DIMS assays using the putative annotation of chemicals based on *m/z* only in each sample (standard mixture, serum, house-dust, and wrist-band extract) I and matched against the 4,462 ToxCast chemicals reference list at Phase I and smaller reference lists at Phase II:

1. Of the putatively annotated chemicals, how many were correct (true positives (TP) or true positive rate (TPR))?
2. Of the putatively annotated chemicals, how many were incorrect (false positives (FP) or false positive rate (FPR))?
3. Can (nano)ESI-DIMS methods be used to determine which of each pair of reference material extracts (serum 1 and 2, house-dust 1 and 2 and wristband 1 and 2), was fortified with one of the ten ToxCast mixtures?
4. Can (nano)ESI-DIMS methods be used to determine with which of the ten ToxCast mixtures each sample was fortified with?

5. For each of the four complementary assays, comparing TP and FP for simple vs increasingly more complex sample matrices (i.e., standards in DMSO vs serum, house-dust then wrist-band extracts), is there a relationship between the number of TP and sample complexity?
6. Comparing TP and FP of Phase I and Phase II results (very large *vs*. very small databases), is there a relationship between the number of TP and the size of the databases searched?

3.2 Results and Discussion

3.2.1 Dilution Studies

Dilution studies were carried out for the ENTACT standard mixtures, and the fortified and unfortified serum, house-dust, and wrist-band extracts. These were carried out to determine the optimal dilution at which the highest number of peaks would be detected for each sample type, and the dilution at which the lowest median peak intensity RSDs would be achieved. The number of peaks detected and the median peak intensity RSD in (nano)ESI-DIMS analyses can be influenced by the degree of sample-specific matrix effects, hence the dilution studies were a measure of the dilution at which matrix effects were minimal for each sample type. Dilution studies were carried in polar and non-polar positive ion-modes only since more *m/z* features are detected in positive ion-mode compared to negative ion-mode which benefits from lower background levels since generally less analytes ionise in this mode. Consequently, sample-specific matrix effects can more easily be evaluated in positive ion-mode. As these data were not part of the major results, data are not shown, and results are summarised below.

- For the standard mixtures, a 100-fold dilution was found to be optimal for both the polar and non-polar positive ion-modes.
- For serum 1 and 2, a 10-fold dilution was found to be optimal for both the polar and non-polar positive ion-modes. Moreover, it was suspected that serum 2 had been spiked with an unknown standard mixture since decreasing numbers of *m/z* features were observed as dilution increased for this sample but not for serum 1.
- For house-dust 1 and 2, a 100-fold dilution was found to be optimal for both the polar and non-polar positive ion-modes. Dilution patterns for house-dust 1 and 2 did not give any indication of which of these samples had been fortified.
- For the wrist-band extracts (1 and 2), stable spray in the (nano)ESI source could not be achieved, with at least 50% of infused samples failing to run to completion with a stable spray throughout the run for each attempted dilution series. This could have been due to a variety of reasons, including sample precipitation and sample degradation releasing gas, both of which can disrupt stable electrospray. Wristband extracts were therefore excluded from further (nano)ESI-DIMS analysis, as

optimisations of sample preparation methods were required and are recommended but were beyond the scope of this thesis.

Each of the dilutions found to be optimal for each sample type was used to prepare samples for (nano)ESI-DIMS analyses using polar positive and negative ion-mode assays, and non-polar positive and negative ion-mode assays, with optimal dilutions extrapolated and applied for analysis of negative ion-mode data.

The standard mixtures, fortified and unfortified serum and fortified and unfortified house-dust samples were analysed at 100-fold, 10-fold, and 100-fold dilutions, respectively, using the polar positive and negative ion-mode assays, and the non-polar positive and negative ion-mode (nano)ESI-DIMS assays. Data for these analyses are presented below.

3.2.2 Database Matching

Database matching was carried out for the standard mixtures and fortified and unfortified serum and house-dust samples which were analysed using the polar positive and negative, and non-polar positive and negative assays, respectively. Initial database matching was carried out using a 3ppm mass error database matching tolerance, guided by the manufacturer's instrument specifications for the Orbitrap Elite (mass accuracy of 3ppm using external calibration) (394). This matching tolerance, however, yielded 15,402 annotations for all standards combined, when it was in-fact known at the start of Phase I ENTACT experiments that only 1,940 chemicals were present in the standard mixtures. Testing of ppm error tolerances demonstrated that the larger the mass error matching tolerance, the more annotations reported. Work by Fiehn and collaborators showed that mass accuracy alone, even at < 1ppm mass error was insufficient to yield a single elemental composition during database matching. However, mass accuracy and the use of isotopic abundance patterns could reduce elemental composition candidates by >95% (395). Annotations using MI-Pack (301) included peak pattern matching, so having data with mass accuracies of <1ppm, and thus using a <1ppm matching tolerance for database matching, would decrease elemental composition assignment and ultimately the number of annotations greatly. This in turn could decrease false positive annotations. To this end, investigations were carried out to determine the mass error distribution of each dataset, and thereby determine if using a mass error

database matching tolerance as small as 1ppm would be appropriate. Since no known compounds within the samples could be used to determine this, mass error histograms were plotted for all assays, and clusters of mass error regions were taken to reflect areas where most true positive annotations were more likely present. These histograms were then used to guide data mass correction if required. As this study did not form part of the major conclusions, data are not shown here, and results are summarised below.

- Mass error regions likely containing the TP in the polar positive ion-mode were +0.5 to +2.5ppm.
- Mass error regions likely containing the TP in the polar negative ion-mode were +0.4 to +2.0ppm.
- Mass error regions likely containing the TP in the non-polar positive ion-mode were +0.4 to +3.0ppm.
- Mass error regions likely containing the TP in the non-polar negative ion-mode were -0.6 to +1.2ppm.

Using this information, *m/z* data were then corrected by subtracting the median mass error by which histograms were skewed from each recorded *m/z* in the dataset. A median value was calculated from the mass error window containing TP for each assay and ion-mode, and this median mass error was subtracted to give data theoretically clustered around 0ppm. Corrected data were then annotated using a 1ppm error matching tolerance.

3.2.3 Method Performance for Data Collected at PHASE I (BLINDED) of ENTACT

The standard mixtures were analysed using 4 (nano)ESI-DIMS methods (polar positive and negative, and non-polar positive and negative ion-mode assays). Data were processed as described in sections 2.2.10 and 2.2.11, and each dataset was mass corrected and annotated using a 1ppm error matching tolerance. The objective of these analyses was to evaluate how well (nano)ESI-DIMS methods applied in metabolomics perform at detecting xenobiotics in a clean solvent matrix. Method performance for the 4 (nano)ESI-DIMS assays was carried out by calculating the number of TP and FP for the standards in solvent (see 2.2.13 for calculation). These calculations were carried out after sample compositions were revealed (see details in section 3.1), thus method performance was evaluated after all ENTACT data for Phase I and Phase II had been submitted.

Additional to the standard mixtures, fortified and unfortified serum and house-dust samples were also analysed in the same way and using the same assays. The objective of these analyses was to evaluate how well (nano)ESI-DIMS methods applied in metabolomics performed at detecting xenobiotics spiked into complex matrices. Methods performed well if they could determine, through calculation of TPRs and FPRs (see 2.2.13 for calculation), which of two serum and house-dust samples had been spiked and with which of the ten standard mixtures they had been spiked with.

Results for each of these sample types are presented below.

3.2.3.1 Standard Mixtures

TPRs for each of four assays applied for analysis of the standard mixtures are shown in Table 26. Standard 6 had an uncharacteristically low TPR of 2% using the non-polar positive ion-mode method. Reviewing the unprocessed data did not reveal any issues with data collection (no electrospray instability), so it was suspected that the low TPR was because the standard mixture had few non-polar compounds ionising in positive ion mode.

		Polar Positive	Polar Negative	Non-polar Positive	Non-polar Negative
TRUE POSITIVES (%)	Standard 1	49.5	23.2	36.8	29.5
	Standard 2	50.5	28.4	47.4	29.5
	Standard 3	48.4	32.6	38.9	30.5
	Standard 4	42.7	34.6	38.4	35.7
	Standard 5	43.2	38.9	42.1	35.8
	Standard 6	40.3	21.6	1.9	25.2
	Standard 7	34.7	26.3	36.8	30.5
	Standard 8	55.6	25.8	47.9	35.9
	Standard 9	51.2	39.2	47.4	42.5
	Standard 10	28.5	32.9	24.7	19.5

Table 26: Summary of the TPRs for 10 standard mixtures analysed using 4 (nano)ESI-DIMS assays. TPRs were calculated as "the number of correctly named compounds/the number of compounds known to be in each mixture x 100". These ranged from 28-56, 21-39, 2-48 and 20-42% for the polar positive, polar negative, non-polar positive, and non-polar negative assays, respectively.

FPRs for each of four assays applied for analysis of the standard mixtures are shown in Table 27. As with the TPRs, the FPR for standard mixture 6 was relatively much lower than the other

standard mixtures, especially for the non-polar positive ion mode (4%). This was unexpected, and further literature review on the ENTACT experimental design found that standard mixture 6 had been designed to be of low purity (396). As such, the low true and false positive annotations achieved were likely due to ion suppression from contaminants within this low purity mixture.

		Polar Positive	Polar Negative	Non-polar Positive	Non-polar Negative
FALSE POSITIVES (%)	Standard 1	369.5	412.6	175.8	302.1
	Standard 2	370.5	485.3	200.0	240.0
	Standard 3	396.8	627.4	177.9	363.2
	Standard 4	204.3	344.9	104.9	247.0
	Standard 5	274.7	657.9	116.8	350.5
	Standard 6	151.0	143.0	4.1	141.6
	Standard 7	314.7	525.3	150.5	246.3
	Standard 8	131.2	132.6	96.4	121.6
	Standard 9	138.6	180.3	98.1	149.9
	Standard 10	115.3	227.7	70.7	86.6

Table 27: Summary of the FPRs for 10 standard mixtures analysed using four (nano)ESI-DIMS assays. FPRs were calculated as "the number of incorrectly named compounds/the number of compounds now known to be in each mixture x 100". These ranged from ranged from 115-400, 180-660, 4-200 and 87-360% for polar positive, polar negative, non-polar positive, and non-polar negative assays, respectively.

More FP annotations were observed in the negative *vs.* positive ion-mode assays for some standard mixtures. This was because generally, more peaks were detected in the negative ion-mode assays for these data, as shown in Table 29. Data processing can also impact higher background ions observed if blank subtraction parameters are too conservative. Possibly, some peaks were detected in both the blank and the standard mixtures but were not 10-fold more intense within the standards and were therefore not filtered out.

These results demonstrate that these methods can be applied for the analysis of xenobiotics since there were some analytes detected. However, there are concerns regarding the high FPRs. These data were annotated using a reference list with 4,462 chemicals only. Databases commonly applied in NTA are much larger. For example, HMDB has ~ 114,000 compounds (74,75). This means application of (nano)ESI-DIMS methods can be impeded by FP annotations; for example, there is no method available to differentiate isomers currently

though ion mobility coupled to DIMS assays may be one appropriate tool to investigate in the future. Application of (nano)ESI-DIMS for the analysis of xenobiotics therefore requires highly accurate mass spectral data so a narrow database matching tolerance can be used at the annotation step. The Orbitrap Elite offers mass accuracy of 3ppm with external calibration. For ten relatively simple standard mixtures (with between 95-365 chemicals each) in a clean matrix, there were 15,402 annotations using a 3ppm matching tolerance. This was reduced significantly to ~5,000 when a 1ppm mass error matching tolerance was applied. Few mass spectrometers can give data with mass accuracies as low as 1ppm, and even those offering these mass accuracies (e.g., Orbitrap Fusion Lumos and Orbitrap Eclipse) often have increasing mass errors for larger compounds. Thus, a data correction strategy as demonstrated here would vastly improve applicability of (nano)ESI-DIMS for the analysis of xenobiotics.

3.2.3.2 Fortified and Unfortified Serum

TPRs for all assays applied to analyse serum 1 and serum 2 were calculated for each standard mixture and are shown in Table 28.

		Polar Positive		Polar Negative		Non-polar Positive		Non-polar Negative	
SERUM TRUE POSITIVES (%)	Standard 1	5.3	7.4	7.4	2.1	2.1	3.2	4.2	10.5
	Standard 2	7.4	10.5	7.4	2.1	1.1	1.1	3.2	5.3
	Standard 3	12.6	12.6	9.5	3.2	3.2	3.2	5.3	12.6
	Standard 4	5.9	21.1	4.3	2.7	1.6	13.0	3.2	6.5
	Standard 5	6.3	34.7	3.2	3.2	3.2	24.2	3.2	8.4
	Standard 6	6.0	8.2	6.3	3.0	2.7	4.7	3.3	7.1
	Standard 7	7.4	12.6	6.3	2.1	3.2	8.4	3.2	12.6
	Standard 8	9.0	7.9	7.1	1.4	2.7	3.6	4.1	9.6
	Standard 9	9.0	11.8	4.9	2.5	3.6	9.3	3.3	7.7
	Standard 10	5.2	4.7	3.8	0.8	1.9	2.7	1.9	6.3
		Serum 1	Serum 2	Serum 1	Serum 2	Serum 1	Serum 2	Serum 1	Serum 2

Table 28: Summary of the TPRs for a fortified and unfortified serum sample analysed using 4 (nano)ESI-DIMS assays. TPRs were calculated as "the number of correctly named compounds/the number of compounds now known to be in each mixture x 100". TPRs for serum 1 were 5-12%, 3-9%, 1-3%, and 2-5% for the polar positive, polar negative, non-polar positive and non-polar negative ion-mode assays, respectively. In comparison, the TPRs for serum 2 were 5-34%, 1-3%, 1-24%, and 5-12% for the polar positive, polar negative, non-polar positive, and non-polar negative ion-mode assays, respectively.

It was not immediately clear from these TPR distributions which of these two samples had been fortified. To investigate further, the total number of reproducibly detected peaks for each assay and sample were considered. Peaks remaining after SNR, replicate, and blank filtering for the polar and non-polar positive ion-mode assays remained approximately the same for both serum 1 and 2 (Table 29).

	Polar Positive	Polar Negative	Non-polar Positive	Non-polar Negative
Standard 1	1963	1660	1344	1402
Standard 2	1972	2609	1389	2580
Standard 3	2132	2202	1450	2677
Standard 4	2519	2932	2163	4085
Standard 5	1777	2480	1309	2598
Standard 6	3036	2357	2833	4301
Standard 7	1966	2289	859	2546
Standard 8	3106	2477	3742	6139
Standard 9	2745	2512	3579	5705
Standard 10	2575	2303	2533	3219
Serum1	3025	2400	4050	5554
Serum2	3054	1835	4106	1304
House-dust1	2965	3968	5254	6213
House-dust2	3156	3989	5249	6140

Table 29: The number of features detected in each assay for the standard mixtures, fortified and unfortified serum and house-dust samples after SNR, replicate, and blank filtering. For the polar and non-polar positive assays, the number of peaks detected does not vary much between serum 1 and serum 2. However, the polar and non-polar negative assays, a decrease is observed in the number of peaks detected in serum 2. The cells highlighted are the largest peak counts observed for all sample types and assays. The biggest peak counts are observed in the non-polar negative assay for standard 8, standard 9, serum 1, house-dust 1 and house-dust 2. The non-polar positive assay also has large peak counts for house-dust 1 and 2. From these observations, it can be concluded that house-dust is a more complex sample type than serum, having a bigger number of peaks detected in the non-polar assays particularly.

However, for the polar and non-polar negative ion-mode assays, the number of reproducibly detected peaks decreased for serum 2. This decrease in peak count could either be attributed to spiked chemicals reducing the detection of matrix components and resulting in a reduction in the number of peaks detected or indicate that the sample with less peaks (serum 2) had not been fortified. The largest number of chemicals that could be spiked in was 365 (standard mixtures had between 95-365 chemicals each), thus it was likely that 365 high concentration chemicals or less could result in the reduction of matrix components through ion suppression and thus lead to less peaks being reproducibly detected rather than increase peak counts by

~600 and 4 000 for the polar and non-polar negative ion-mode assays, respectively. Additionally, the dilution studies for serum 2 using the non-polar positive ion-mode suggested that serum 2 had been spiked as no dilution behaviour was observed for serum 1 but was observed for serum 2. However, there was no definitive way to determine which mixture had been spiked. The TPR calculations were used to determine which sample had been spiked and with which standard mixture simply by taking the sample and standard mixture that had the highest TPR.

Serum 1 was therefore concluded to not be spiked with an unknown standard mixture, consistently having TPRs no higher than 12% for all assays. Serum 2 analysis with the polar and non-polar positive assays yielded TPRs highest for standard 5 (34% for polar positive and 24% for non-polar positive) (Figure 22). This suggested that standard 5 had been spiked into serum 2. N.B. that the next highest TPRs were observed for standard 4. This can be explained by the overlap in compounds present in both standard 4 and 5. Standard 5 was in-fact a subset of standard 4, with all 95 chemicals present in standard 5 being present in standard 4, which had 185 chemicals. This explained the multiple correct annotations observed and resulting in standard 4 having the second highest TP annotations.

The TPRs in the negative ion-mode assays were much lower. The polar negative ion-mode TPR was still highest, albeit at only 3%, for standard 5, further supporting the conclusion that this mixture was spiked into serum 2. The non-polar negative TP, however, did not support this conclusion, with the highest TPRs being for standard 3 and 7 at only 13%. As observed with the standard mixtures (section 3.2.3.1), the negative ion-mode assays gave lower TPRs. Indeed, the TPR for standard 5 was lowest for the non-polar negative assay when analysing the standard mixtures in a much cleaner solvent matrix. This can explain why the non-polar negative assay could not clearly determine which standard mixture was spiked into serum 2. Despite these results, the other three assays were able to suggest which standard mixture had been spiked and into which of the 2 serum samples.

Figure 22: Radar plot showing TPRs for four (nano)ESI-DIMS assays for both serum 1 and 2. The orange and green traces labelled serum 2 show clearly that serum 2 had the highest TPR for standard 5 being spiked in when the polar and non-polar positive assays were applied therefore it was concluded that serum 2 had been spiked with standard 5. N.B. the other colours represent TPRs for each of the other standard mixtures applying the four assays.

FPRs for serum 1 and 2 are shown in Table 30. The FPRs were highest for serum 1 for 3 out of 4 assays (serum 2 FPRs were highest for serum 2) which is consistent with the conclusion that serum 1 had not been spiked. It is unclear why the FPRs were highest in serum 2 for the non-polar negative assay as the same modifier was used for the solvents. It could, however, be due to accidental cross contamination of the sample wells. This was not investigated further as this was beyond the objective of the research.

		Polar Positive		Polar Negative		Non-polar Positive		Non-polar Negative	
SERUM FALSE POSITIVES (%)	Standard 1	629.5	502.1	513.7	181.1	253.7	258.9	215.8	515.8
	Standard 2	627.4	498.9	513.7	181.1	254.7	261.1	216.8	521.1
	Standard 3	622.1	496.8	511.6	180.0	252.6	258.9	214.7	513.7
	Standard 4	320.0	240.5	263.2	91.4	129.7	121.6	109.7	263.8
	Standard 5	628.4	474.7	517.9	180.0	252.6	237.9	216.8	517.9
	Standard 6	159.2	124.4	129.3	44.7	63.8	63.6	54.0	129.9
	Standard 7	627.4	496.8	514.7	181.1	252.6	253.7	216.8	513.7
	Standard 8	156.2	124.7	128.5	46.3	63.8	64.7	53.2	127.4
	Standard 9	156.2	120.8	130.7	45.2	63.0	58.9	54.0	129.3
	Standard 10	160.0	127.9	131.8	46.8	64.7	65.5	55.3	130.7
		Serum 1	Serum 2	Serum 1	Serum 2	Serum 1	Serum 2	Serum 1	Serum 2

Table 30: Summary of the FPRs for fortified and unfortified serum samples analysed using four (nano)ESI-DIMS assays. FPRs were calculated as "the number of incorrectly named compounds/the number of compounds now known to be in each mixture x 100". FPRs for serum 1 were 160-630, 128-518, 63-254, and 54-215% for the polar positive, polar negative, non-polar positive and non-polar negative assays, respectively. FPRs for serum 2, ranging from 120-502, 45-181, 64-261 and 127-521%, respectively, were slightly lower than for serum 1, with the lowest being for the polar negative assay which had less features detected overall.

The (nano)ESI-DIMS methods applied for analysis of fortified and unfortified serum samples suggested which of the two serum samples had been fortified and with which standard mixture. However, conclusions were not made with a high level of confidence, and in a real-life study such data would need confirmation using TA. The TPRs for each assay were relatively low, reflecting the complexity of analysis of xenobiotics in the serum matrix, particularly at low concentrations as is expected in a real-life scenario. The polar and non-polar positive ion-mode assays gave higher TPRs than the negative ion-mode assays, but this is likely an indication of more xenobiotics ionisable in positive ion-mode than negative.

The FPRs for the fortified serum sample were much higher than those observed in the clean solvent matrix of standard 5. This is undoubtedly due to matrix effects and further demonstrates the challenge of analysing xenobiotics in complex matrices such as serum.

Relative to the standard mixtures; data correction, a narrow mass error tolerance for database matching, and the use of isotopic abundance patterns were insufficient in reducing the FPR since generally more peaks were detected in the serum samples compared to the standard mixtures (Table 29).

Analysis of xenobiotics in serum matrices must therefore be carried out using an alternative approach such as U(H)PLC-MS for higher confidence annotation results when multiple types of data can be used including *m/z of* the intact chemical but also retention time (RT) and MS/MS data which can be matched to reference libraries/databases constructed with data acquired for authentic chemical standards. Indeed, these samples were cleaned prior to analysis using solid phase extraction (SPE), so even such strategies are insufficient at reducing false positive annotations.

3.2.3.3 Fortified and Unfortified House-dust

TPRs for all assays applied to analyse house-dust 1 and 2 were calculated for each standard mixture and are shown Table 31.

		Polar Positive		Polar Negative		Non-polar Positive		Non-polar Negative	
HOUSE-DUST TRUE POSITIVES (%)	Standard 1	9.5	13.7	18.9	16.8	22.1	8.4	10.5	10.5
	Standard 2	6.3	7.4	17.9	14.7	21.1	5.3	7.4	10.5
	Standard 3	11.6	15.8	24.2	23.2	26.3	11.6	13.7	16.8
	Standard 4	9.7	14.6	9.2	16.2	13.5	8.6	9.7	14.1
	Standard 5	9.5	17.9	9.5	18.9	14.7	11.6	11.6	18.9
	Standard 6	11.5	16.7	12.9	14.8	15.9	7.4	8.2	11.0
	Standard 7	9.5	13.7	11.6	13.7	14.7	8.4	5.3	9.5
	Standard 8	13.2	14.0	17.0	18.4	19.2	5.8	11.8	12.1
	Standard 9	11.2	20.5	11.8	20.0	15.3	11.8	9.6	15.6
	Standard 10	6.0	6.6	8.2	8.2	9.3	3.6	6.8	8.8
		House-dust 1	House-dust 2	House-dust 1	House-dust 2	House-dust 1	House-dust 2	House-dust 1	House-dust 2

Table 31: Summary of the TPRs for the fortified and unfortified house-dust sample analysed using 4 (nano)ESI-DIMS assays. TPRs were calculated as "the number of correctly named compounds/the number of compounds now known to be in each mixture x 100". For house-dust 1, these were 6-13, 8-24, 9-26, and 5-14% for the polar positive, polar negative, non-polar positive and non-polar negative assays, respectively. In comparison, the TPRs for house-dust 2 were 7-21, 8-23, 4-12, and 9-14% for the polar positive, polar negative, non-polar positive, and non-polar negative assays, respectively.

It was not immediately clear which of the 2 house-dust samples, a more complex sample matrix than serum based generally on the number of peaks reproducibly detected

Table 29), had been spiked and with which standard mixture. Considering the TPR alone did not suggest which standard mixture had been spiked and into which of the 2 house-dust samples. The highest TPR was for standard 3 being spiked into house-dust 1 (Figure 23). This conclusion was therefore submitted to ENTACT as part of Phase I results.

FPRs for house-dust 1 and 2 are shown in Table 32. These were lower for house-dust 1. This was also likely due to the higher number of peaks detected in this sample type explaining why it was challenging to confidently discern which standard mixture had been spiked into the house-dust samples.

		Polar Positive		Polar Negative		Non-polar Positive		Non-polar Negative	
HOUSE-DUST FALSE POSITIVES (%)	Standard 1	781.1	829.5	1685.3	1705.3	2017.9	335.8	825.3	941.1
	Standard 2	784.2	835.8	1686.3	1707.4	2018.9	338.9	828.4	941.1
	Standard 3	778.9	827.4	1680.0	1698.9	2013.7	332.6	822.1	934.7
	Standard 4	396.2	418.4	865.9	868.1	1034.1	168.1	419.5	474.6
	Standard 5	781.1	825.3	1694.7	1703.2	2025.3	332.6	824.2	932.6
	Standard 6	194.2	202.7	430.7	433.4	515.1	82.2	209.3	236.7
	Standard 7	781.1	829.5	1692.6	1708.4	2025.3	335.8	830.5	942.1
	Standard 8	192.6	205.5	426.6	429.9	511.8	83.8	205.8	235.6
	Standard 9	194.5	198.9	431.8	428.2	515.6	77.8	207.9	232.1
	Standard 10	199.7	212.9	435.3	440.0	521.6	86.0	210.7	238.9
		House-dust 1	House-dust 2	House-dust 1	House-dust 2	House-dust 1	House-dust 2	House-dust 1	House-dust 2

Table 32: Summary of the FPRs for a fortified and unfortified house-dust sample analysed using 4 (nano)ESI-DIMS assays. FPRs were calculated as "the number of incorrectly named compounds/the number of compounds now known to be in each mixture x 100". FPRs were 192-784, 426-1694, 512-2025, and 206-831% for house-dust 1, and 199-8299, 428-1707, 78-336, 232-942% for house-dust 2, for the polar positive, polar negative, non-polar positive, and non-polar negative assays, respectively.

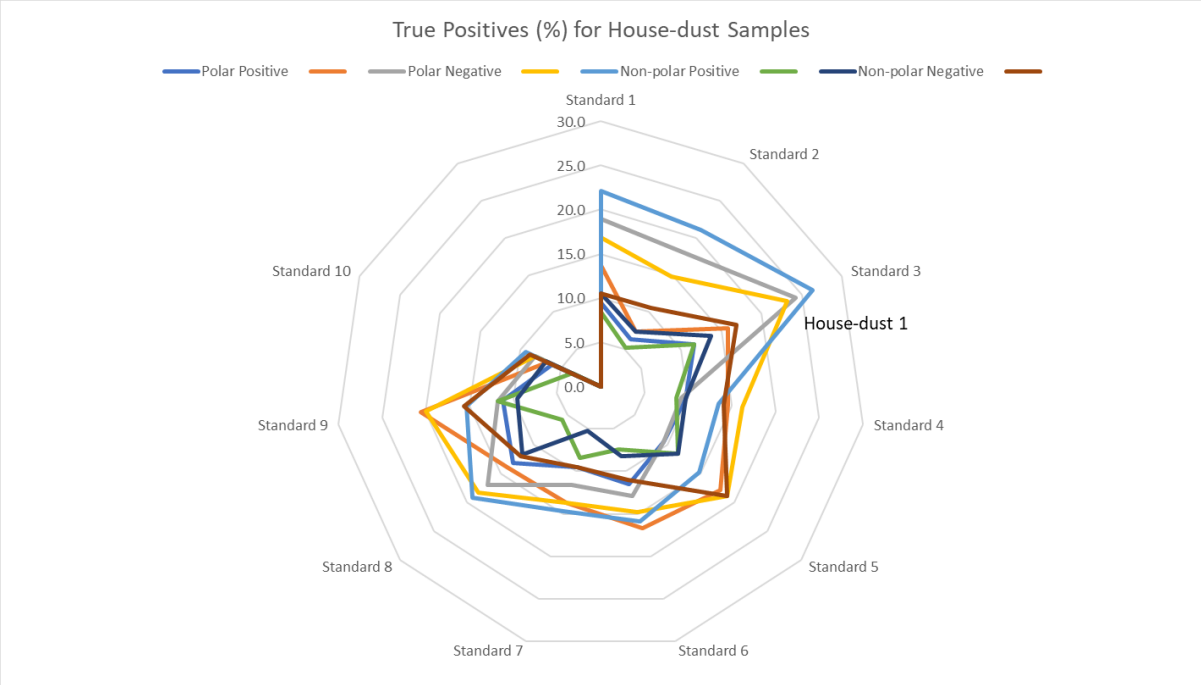


Figure 23: Radar plot showing TPRs for 4 (nano)ESI-DIMS assays for both house-dust 1 and 2. The blue trace suggests that standard 3 may have been spiked into house-dust 1, but based on the distribution of TP, no single candidate mixture can be arrived at. N.B. the other colours represent TPRs for each of the other standard mixtures and assays.

The (nano)ESI-DIMS methods could not successfully be applied for the analysis of xenobiotics in a house-dust matrix. It could not unambiguously be determined which house-dust sample had been spiked and with which of the standard mixtures. Analysis of xenobiotics in house-dust is not recommended using (nano)ESI-DIMS methods as results have very low TPRs, and very high FPRs. A separation technique such as U(H)PLC-MS is recommended. However, (nano)ESI-DIMS methods could be applied in combination with other analytical strategies, providing improved sensitivity and a fast-screening approach which other techniques cannot always offer.

3.2.4 Phase I (BLINDED): Summary

All four assay annotations were combined for each sample type, and the results are summarised below. The cleanest sample matrix (standard mixtures in solvent) had TPRs ranging from 48-74% (Table 33). The clean matrix of the standards resulted in fewer peaks detected relative to the serum and house-dust samples for some assays (

Table **29**), making the annotation step more accurate with less neighbouring peaks for each analyte. Two standard mixtures (6 and 7) had TPRs below 50%. The rest had TPRs above 60%. FPRs for the standards ranged from 254-879% (Table 34). These were concerning for the analysis of xenobiotics as they reduce confidence in annotations gained from applying these analyses. Single *m/z* features can have multiple annotations with no way to discern correct from incorrect annotations, or at least further filter the annotations.

TPRs for serum 2 ranged from 16-43% (Table 33), with the highest TPR corresponding to serum 2 being fortified with standard mixture 5. FPRs for serum 2, however, ranged from 238-984% (Table 34), with the FPR for fortification with standard mixture 5 being 956%. These results reiterated concerns about the applicability of (nano)ESI-DIMS methods for the analysis of xenobiotics in the serum matrix as they yielded such high FP annotations.

The highest TPR for the house-dust corresponded to house-dust 2 being fortified with either standard mixture 5 or 9 (both with TPRs of 38%), with the TPRs for house-dust 2 ranging from 23-38% (Table 33). However, it could not be concluded with certainty which of these 2 standard mixtures had been spiked into house-dust 2 (Table 33). FPRs for house-dust 2 ranged from 409-1694%, with FPRs for standard mixture 5 and 9 being spiked into house-dust 2 at 1679 and 409%, respectively (Table 34). The FPRs offered no further clarification on which of the 2 standard mixtures had been fortified.

Overall, the more complex the sample matrix, the lower the TPRs and the higher the FPRs were. Sample complexity can be indicated by the number of features detected. For these sample matrices, the solvent matrix had the smallest number of peaks reproducibly detected, followed by serum and then house-dust. (nano)ESI-DIMS data acquisition methods and non-targeted data processing workflows employing a relatively large reference list (4,462 chemicals) could successfully be used for the blind analysis of xenobiotics in a clean solvent matrix but not confidently for xenobiotics in a serum and house-dust matrix, where false positive annotations were worryingly high.

TRUE POSITIVES (%)	Standard 1	56.8	13.7	18.9	28.4	28.4
	Standard 2	57.9	15.8	15.8	24.2	23.2
	Standard 3	63.2	21.1	24.2	28.4	31.6
	Standard 4	61.1	11.9	28.1	20.0	30.3
	Standard 5	66.3	11.6	43.2	22.1	37.9
	Standard 6	47.7	13.4	17.3	22.2	28.2
	Standard 7	48.4	13.7	23.2	17.9	26.3
	Standard 8	70.7	17.5	17.8	25.5	27.9
	Standard 9	73.7	14.2	22.2	22.5	38.4
	Standard 10	72.4	17.8	21.6	27.0	28.1
		Standards	Serum 1	Serum 2	House-dust 1	House-dust 2

Table 33: Summary table showing TPRs, calculated as "the number of correctly named compounds/the number of compounds now known to be in each mixture x 100", for all 4 (nano)ESI-DIMS assays combined. TPRs for standard mixtures, fortified and unfortified serum, and house-dust are shown. The cleanest sample matrix (standard mixtures in solvent) had TPRs ranging from 48-74%, TPRs for serum 2 ranged from 16-43%, and TPRs for house-dust 2 ranging from 23-38%.

FALSE POSITIVES (%)	Standard 1	705.3	1101.1	981.1	1582.1	1688.4
	Standard 2	769.5	1098.9	984.2	1586.3	1693.7
	Standard 3	878.9	1093.7	975.8	1582.1	1685.3
	Standard 4	522.2	560.5	485.4	807.0	851.4
	Standard 5	868.4	1103.2	956.8	1588.4	1678.9
	Standard 6	253.2	276.7	243.0	397.0	418.6
	Standard 7	817.9	1101.1	976.8	1592.6	1690.5
	Standard 8	272.3	272.6	242.5	393.7	418.9
	Standard 9	269.9	275.9	238.1	396.7	408.5
	Standard 10	429.7	554.6	491.9	800.0	853.5
		Standards	Serum 1	Serum 2	House-dust 1	House-dust 2

Table 34: Summary table showing FPRs, calculated as "the number of incorrectly named compounds/the number of compounds now known to be in each mixture x 100", for all 4 (nano)ESI-DIMS assays combined. FPRs for standard mixtures, fortified and unfortified serum, and house-dust are shown. FPRs for the standards ranged from 254-879%, the FPR for serum 2 fortification with standard mixture 5 was 956%, and the FPRs for house-dust 2 ranged from 409-1694%, with FPRs for standard mixture 5 and 9 being spiked into house-dust 2 at 1679 and 409%, respectively.

3.2.5 PHASE II (UNBLINDED): Method Performance

Revelation of sample compositions at Phase II experiments facilitated use of this knowledge to either apply modified analytical and/or computational approaches to the same samples and re-evaluate method performance.

One strategy that could improve the effectiveness of (nano)ESI-DIMS methods is the use of smaller, more targeted databases for chemical annotation. A common reason for false positive annotation in (nano)ESI-DIMS studies is isomers, which cannot be differentiated using *m/z* measurements of the intact chemical alone. The presence of isomers in a database will always result in a single *m/z* being multiply annotated as the many isomers corresponding to this *m/z*, thereby increasing false positive annotation. The ToxCast database contains 4,462 chemicals of which 1,570 chemicals could be matched by empirical formula to another chemical in the database and were therefore isomers. Moreover, *m/z* values with small differences (<5ppm) between them can be annotated as one another during database matching or form different ion forms that fall within a small enough mass error tolerance that they are annotated incorrectly.

Therefore, during Phase II experiments, the focus was on reducing FP annotations for the standards, serum, and house-dust samples by reducing the search space through creating small databases containing only the chemicals revealed to be in each standard mixture. Data for all sample types were not re-acquired, only the annotation step was repeated using the 10 small databases containing between 95-365 chemicals as per the revealed standard mixture compositions. Database matching was then only carried out by matching each standard mixture (and samples evaluated to determine which standard mixture had been spiked in carried out at Phase I experiments) against the corresponding database containing chemicals revealed to be in the standard mixture. TPRs and FPRs were then calculated as carried out at Phase I experiments (described in 3.2.3).

3.2.5.1 Standard Mixtures

TPRs for the four assays are shown in a Table 35. Using these smaller databases and comparing to Phase I results, the lowest TPRs increased for the polar positive, non-polar positive, and non-polar negative ion-mode assays (from 28 to 37%, 2 to 26% and 20 to 21%, respectively), whilst they decreased slightly for the polar negative assay (from 21 to 19%). The

highest TPRs for all assays also increased from 56 to 65%, 39-48%, 48 to 57%, and 42 to 50% for the polar positive, polar negative, non-polar positive and non-polar negative on mode assays, respectively. More importantly, the FPRs decreased dramatically.

		Polar Positive	Polar Negative	Non-polar Positive	Non-polar Negative
TRUE POSITIVES (%)	Standard 1	65.3	32.6	47.4	36.8
	Standard 2	63.2	34.7	56.8	35.8
	Standard 3	61.1	38.9	46.3	40.0
	Standard 4	50.8	43.2	44.3	43.2
	Standard 5	50.5	48.4	47.4	43.2
	Standard 6	42.5	22.7	33.2	27.1
	Standard 7	36.8	29.5	38.9	31.6
	Standard 8	59.5	28.2	51.5	38.9
	Standard 9	56.4	45.5	51.0	47.9
	Standard 10	38.4	19.5	25.8	21.4

Table 35: Summary of the TPRs for 10 standard mixtures analysed using 4 (nano)ESI-DIMS assays. TPRs were calculated as "the number of correctly named compounds/the number of compounds now known to be in each mixture x 100". These ranged from 37-65%, 19-48%, 26-57% and 21-50% for the polar positive, polar negative, non-polar positive, and non-polar negative ion-mode assays, respectively.

FPRs for all four assays annotated using smaller databases are shown in Table 36, and are much lower than FPRs achieved at Phase I, which ranged from 115-400%, 180-660%, 4-200% and 87-360% for the polar positive, polar negative, non-polar positive, and non-polar negative assays, respectively.

		Polar Positive	Polar Negative	Non-polar Positive	Non-polar Negative
FALSE POSITIVES (%)	Standard 1	34.7	15.8	29.5	34.7
	Standard 2	41.1	29.5	33.7	23.2
	Standard 3	37.9	40.0	28.4	27.4
	Standard 4	32.4	28.1	23.2	38.9
	Standard 5	43.2	56.8	25.3	53.7
	Standard 6	51.2	21.6	38.4	44.1
	Standard 7	66.3	35.8	57.9	49.5
	Standard 8	30.4	12.6	26.8	27.1
	Standard 9	30.4	34.5	29.0	37.8
	Standard 10	45.2	14.0	14.2	13.4

Table 36: Summary of the FPRs for 10 standard mixtures analysed using 4 (nano)ESI-DIMS assays. FPRs were calculated as "the number of correctly named compounds/the number of compounds now known to be in each mixture x 100". These ranged from 30-66%, 13-57%, 14-58% and 13-54% for the polar positive, polar negative, non-polar positive, and non-polar negative ion-mode assays, respectively.

These results demonstrate the impact of reducing the search space during database matching. By removing isomers and irrelevant compounds from the database as was done creating these targeted databases, the likelihood of incorrect annotation is greatly reduced, whilst the chance of correct annotation is increased so long as the compound is detected and present within the database. However, reducing the search space in this manner is not easily achieved in NTA, where all chemical constituents in the environment or all metabolites within an organism are unknown. As such, databases are often created with a view to making them as comprehensive as achievable, which often means they are large and contain many isomers. Semi-targeted assays (397) attempt to alleviate this issue, mostly detecting all sample constituents, but importantly also focusing on a subset of expected analytes to confidently identify, often through MS2 fragmentation. Such a strategy can increase TPRs of annotation and identification.

3.2.5.2 Fortified and Unfortified Serum

In Phase II of ENTACT, it was revealed which of 2 serum samples was fortified and unfortified, and with which of the ten standard mixtures. TP and FP were still calculated for all standards to evaluate the impact of using smaller databases. Using the ten small, targeted databases created after sample compositions were revealed, TPRs and FPRs were calculated for serum 1 and 2, assuming in sequence that each of ten standard mixtures had been spiked in. The same approach was used for the house-dust samples.

TPRs for serum 1 and 2 are shown in Table 37. For serum 1, smaller minimum and maximum TPRs than at Phase I were achieved (Phase I TPRs ranged from 5-12%, 3-9%, 1-3% and 2-5%). For serum 2, the minimum and maximum TPRs generally increased (Table 37) with the use of the smaller databases., compared to 5-34%, 1-3%, 1-24% and 5-12% for the polar positive, polar negative, non-polar positive, and non-polar negative ion-mode assays, respectively, achieved at Phase I. Indeed, serum 2 was more easily concluded to have been fortified with standard mixture 5, which had the highest TPR for all assays. The largest TPRs increased when using the targeted databases compared to the larger ToxCast database (from 34%, 3%, 24% and 12% compared to 44%, 6%, 29% and 16% for the polar positive, polar negative, non-polar positive and non-polar negative ion-mode assays, respectively). The TPRs for the standard

mixtures that had not been spiked into each serum mixture decreased for both serum 1 and 2.

The FPRs for both serum 1 and 2 decreased significantly and are shown in Table 38. These were both lower than FPRs in Phase I, which were 160-630%, 128-518%, 63-254% and 54-215% for serum 1 and 120-502%, 45-181%, 64-261% and 127-521% for serum 2 for the polar positive, polar negative, non-polar positive and non-polar negative ion-mode assays, respectively. The polar negative assay had no FP annotations for standards 4, 5 and 9, demonstrating how effective smaller databases can be at reducing the number of false annotations. There is a clear relationship between TPRs and FPRs. Looking for TP in samples where there are not any results in much higher FPRs. To demonstrate, serum 1 was known to be unfortified. The FPRs for this sample were observed to be much higher than those of the fortified sample (serum 2).

		Polar Positive		Polar Negative		Non-polar Positive		Non-polar Negative	
SERUM TRUE POSITIVES (%)	Standard 1	4.2	3.2	3.2	0.0	1.1	1.1	1.1	2.1
	Standard 2	3.2	3.2	5.3	0.0	0.0	1.1	1.1	2.1
	Standard 3	6.3	4.2	7.4	0.0	2.1	2.1	3.2	2.1
	Standard 4	2.7	22.7	2.2	3.2	0.5	15.1	2.7	8.6
	Standard 5	3.2	44.2	2.1	6.3	1.1	29.5	2.1	16.8
	Standard 6	1.9	4.7	3.0	0.8	1.1	3.3	1.1	2.2
	Standard 7	4.2	10.5	4.2	2.1	1.1	6.3	2.1	7.4
	Standard 8	7.4	0.8	6.3	0.0	2.5	0.3	3.6	0.5
	Standard 9	5.2	11.5	4.1	1.6	1.9	7.7	2.2	4.4
	Standard 10	3.8	0.8	3.6	0.0	1.4	0.3	1.6	0.5
		Serum 1	Serum 2	Serum 1	Serum 2	Serum 1	Serum 2	Serum 1	Serum 2

Table 37: Summary of the TPRs for a fortified and unfortified serum sample analysed using 4 (nano)ESI-DIMS assays. TPRs were calculated as "the number of correctly named compounds/the number of compounds now known to be in each mixture x 100". For serum 1, TPRs ranged from 3-7%, 3-7%, 0-2%, and 1-3% for the polar positive, polar negative, non-polar positive and non-polar negative ion-mode assays, respectively. For serum 2, TPRs ranged from 1-44%, 0-6%, 0-29% and 1-16% for the polar positive, polar negative, non-polar positive and non-polar negative ion-mode assays, respectively.

		Polar Positive		Polar Negative		Non-polar Positive		Non-polar Negative	
SERUM FALSE POSITIVES (%)	Standard 1	351.6	67.4	330.5	6.3	140.0	45.3	163.2	33.7
	Standard 2	352.6	67.4	328.4	6.3	141.1	45.3	163.2	33.7
	Standard 3	349.5	66.3	326.3	6.3	138.9	44.2	161.1	33.7
	Standard 4	180.0	13.5	169.2	0.0	71.9	8.6	81.6	9.7
	Standard 5	352.6	26.3	331.6	0.0	140.0	16.8	162.1	18.9
	Standard 6	90.7	13.7	83.8	0.8	35.6	8.8	41.6	7.1
	Standard 7	351.6	60.0	329.5	4.2	140.0	40.0	162.1	28.4
	Standard 8	85.2	17.5	80.5	1.6	34.2	11.8	39.2	8.8
	Standard 9	87.4	6.8	82.7	0.0	34.8	4.4	40.5	4.9
	Standard 10	88.8	17.5	83.3	1.6	35.3	11.8	41.1	8.8
		Serum 1	Serum 2	Serum 1	Serum 2	Serum 1	Serum 2	Serum 1	Serum 2

Table 38: Summary of the FPRs for a fortified and unfortified serum sample analysed using 4 (nano)ESI-DIMS assays. FPRs were calculated as "the number of incorrectly named compounds/the number of compounds now known to be in each mixture x 100". Serum 1 had FPRs of 85-353%, 81-332%, 35-141% and 39-163% for the polar positive, polar negative, non-polar positive, and non-polar negative ion-mode assays, respectively. Serum 2 had FPRs of 7-67%, 0-6%, 4-45% and 5-34% for the polar positive, polar negative, non-polar positive, and non-polar negative ion-mode assays, respectively.

Using these smaller databases for annotation of xenobiotics in a complex serum matrix yielded results with lower FPRs and higher TPRs. The (nano)ESI-DIMS methods can therefore be successfully applied for the analysis of xenobiotics in a serum matrix only if smaller reference lists are used to minimise the number of false annotations. These methods were able to correctly determine that serum 2 had been spiked with standard mixture 5.

3.2.5.3 Fortified and Unfortified House-dust

TPRs for house-dust 1 and 2 are shown in Table 39. These were lower than the TPRs obtained by matching against the larger ToxCast database (house-dust 1 TPRs at Phase I ranged from 6-13%, 8-24%, 9-26% and 5-14%, and house-dust 2 were 7-21%, 8-23%, 4-12% and 9-14%, for the polar positive, polar negative, non-polar positive and non-polar negative ion-mode assays, respectively). The highest TPRs corresponded to house-dust 2 being fortified with either standard mixture 5 or 9.

		Polar Positive		Polar Negative		Non-polar Positive		Non-polar Negative	
HOUSE-DUST TRUE POSITIVES (%)	Standard 1	6.3	2.1	10.5	2.1	4.2	3.2	6.3	3.2
	Standard 2	4.2	2.1	11.6	2.1	5.3	3.2	7.4	3.2
	Standard 3	6.3	3.2	12.6	2.1	7.4	4.2	8.4	3.2
	Standard 4	4.3	14.6	5.4	13.5	3.8	8.1	5.9	13.5
	Standard 5	6.3	28.4	5.3	26.3	5.3	15.8	7.4	26.3
	Standard 6	4.9	9.9	4.7	6.8	3.6	4.4	3.6	5.8
	Standard 7	5.3	12.6	6.3	10.5	3.2	8.4	4.2	8.4
	Standard 8	10.1	0.5	13.2	0.5	4.9	0.8	9.6	0.8
	Standard 9	7.4	26.6	9.0	25.2	5.8	14.8	7.9	19.5
	Standard 10	4.9	0.5	5.8	0.5	2.5	0.8	6.3	0.8
		House-dust 1	House-dust 2	House-dust 1	House-dust 2	House-dust 1	House-dust 2	House-dust 1	House-dust 2

Table 39: Summary of the TPRs for a fortified and unfortified house-dust sample analysed using 4 (nano)ESI-DIMS assays. TPRs were calculated as "the number of correctly named compounds/the number of compounds now known to be in each mixture x 100". TPRs for house-dust 1 were 4-7%, 5-13%, 2-7% and 4-10% for the polar positive, polar negative, non-polar positive and non-polar negative ion-mode assays, respectively. TP for house-dust 2 were 1-28%, 1-26%, 1-16% and 1-26% for the polar positive, polar negative, non-polar positive and non-polar negative ion-mode assays, respectively.

FPRs for house-dust 1 and 2 are shown in Table 40. These were lower than the FPRs observed in Phase I, (192-784%, 426-1694%, 512-2025% and 206-831% for house dust 1 and 199-8299%, 428-1707%, 78-336% and 232-942% for house-dust 2), for the polar positive, polar negative, non-polar positive and non-polar negative ion-mode assays, respectively. It could not be confidently concluded from the TPRs of each assay if house-dust 2 was fortified with standard mixture 5 or 9, so these were both submitted as candidates. Standard mixture 5 was a subset of standard mixture 9, explaining the lack of differentiation between these two as likely the same analytes were detected. Regardless, this was an improvement from Phase I results, where it was unclear both which of the 2 house-dust samples had been fortified and with which standard mixture.

		Polar Positive		Polar Negative		Non-polar Positive		Non-polar Negative	
HOUSE-DUST FALSE POSITIVES (%)	Standard 1	427.4	122.1	989.5	216.8	222.1	70.5	552.6	149.5
	Standard 2	429.5	122.1	988.4	216.8	221.1	70.5	551.6	149.5
	Standard 3	427.4	121.1	987.4	216.8	218.9	69.5	550.5	149.5
	Standard 4	218.4	49.2	508.1	98.9	112.4	29.7	281.1	64.9
	Standard 5	427.4	95.8	994.7	192.6	221.1	57.9	551.6	126.3
	Standard 6	107.9	22.5	255.6	50.1	55.3	14.8	141.9	34.0
	Standard 7	428.4	111.6	993.7	208.4	223.2	65.3	554.7	144.2
	Standard 8	102.7	31.8	247.1	56.4	54.0	18.4	135.9	38.9
	Standard 9	105.5	5.8	251.2	31.8	53.2	4.4	137.5	20.3
	Standard 10	107.9	31.8	254.5	56.4	56.4	18.4	139.2	38.9
		House-dust 1	House-dust 2	House-dust 1	House-dust 2	House-dust 1	House-dust 2	House-dust 1	House-dust 2

Table 40: Summary of the FPRs for a fortified and unfortified house-dust sample analysed using 4 (nano)ESI-DIMS assays. FPRs were calculated as "the number of incorrectly named compounds/the number of compounds now known to be in each mixture x 100". FPRs for both house-dust 1 and 2 were lower compared to Phase I experiments, ranging from 103-429%, 247-995%, 53-223% and 136-555% and ranging from 6-122%, 32-217%, 4-71% and 20-149% for house-dust 1 then 2, for the polar positive, polar negative, non-polar positive and non-polar negative assays, respectively.

Using the (nano)ESI-DIMS methods, xenobiotics could successfully be annotated in a house-dust matrix only if smaller reference lists were used to reduce false positive annotation. These methods successfully determined that house-dust 2 had been spiked. Although it could not be determined with certainty if house-dust 2 had been spiked with standard mixture 5 or 9, the results still accurately posited the correct standard mixture spiked in (standard mixture 9) as one of two options.

3.2.6 Phase II (UNBLINDED): Summary

All 4 assay annotations for each of the different sample types were combined, showing a gradual decrease in TPRs with an increase in sample matrix complexity.

The TPRs for all samples are shown in

Table **41** and FPRs are shown in Table 42. For the clean solvent matrix, only 2 standard mixtures (6 and 7) had TPRs below 50%. The other standard mixtures had TPRs greater than 72%. FPRs were significantly lower using the smaller reference lists at Phase II of ENTACT experiments. (nano)ESI-DIMS methods can be applied more confidently for the analysis of xenobiotics in a clean matrix, offering higher TPRs and lower FPRs.

For the serum samples, the highest TPR corresponding to serum 2 being spiked with standard mixture 5. FPRs for serum 2 were also significantly reduced by using the smaller reference lists at Phase II of ENTACT experiments. Indeed, the FPR for standard mixture 5 being spiked into serum 2 was only 2%. When small reference lists are used for analysis of xenobiotics in a serum matrix, higher TPRs are yielded alongside significantly lower FPRs. The conclusion that serum 2 was fortified with standard mixture 5 was correct, thus (nano)ESI-DIMS methods can be applied successfully for the analysis of xenobiotics in this sample type if efforts to filter or reduce false positive annotations, such as the use of smaller databases, are employed. Considered efforts are required to assert how best to achieve these smaller databases in NTA.

For the house-dust samples, the highest TPR corresponded to house-dust 2 being fortified with standard mixture 5. The conclusion that house-dust 2 was fortified with standard mixture 5 was incorrect. House-dust 2 was in-fact fortified with standard mixture 9. It was noted that the second highest TPR for house-dust 2 fortification was with standard mixture 9. Moreover, standard mixture 5 was a subset of standard mixture 9 but containing less analytes in total. This explains why it yielded a higher true positive rate than the correct standard mixture 9.

House-dust samples can therefore still be analysed using (nano)ESI-DIMS methods, but even the use of smaller reference lists can still yield incorrect results. It is important to note that the correct result was a candidate, and thus these methods can be used to screen through large lists of chemicals and offer some potential candidates of interest.

FPRs for house-dust 2 ranged from 2-186%. Interestingly, the lowest FPR corresponded to house-dust 2 being fortified with standard mixture 9. This suggests that in very complex sample matrices, analytes of interest can suppress detection of matrix components, and numbers of false positive annotation can be informative in determining which samples contain real analytes. This was supported by serum 2 data, with a low FPR corresponding correctly to standard mixture 5 being spiked into serum 2 at 2%. Indeed, if a low FPR was considered to guide determination of which of the 2 standard mixtures, 5 or 9, had been spiked into house-dust 2, then the correct answer (standard mixture 9) is arrived at.

Both standard mixtures 5 and 9, which had been spiked into serum 2 and house-dust 2, respectively, had TPRs of 83% in a clean solvent matrix. The TPR for standard 5 decreased when it was spiked into serum 2 to 62%. The TPR for standard 9 decreased when it was spiked into house-dust 2 to 48%. FPRs for each sample type showed a similar trend. For the standards, they ranged from 0-24%, increasing to a range of 1-60% for serum 2, and even further to a range of 2-168% for house-dust 2.

TRUE POSITIVES (%)	Standard 1	75.8	6.3	4.2	15.8	4.2
	Standard 2	72.6	8.4	4.2	16.8	4.2
	Standard 3	77.9	12.6	5.3	15.8	5.3
	Standard 4	78.4	6.5	31.9	11.9	27.6
	Standard 5	83.2	6.3	62.1	14.7	53.7
	Standard 6	52.6	6.0	7.1	9.9	14.8
	Standard 7	54.7	7.4	16.8	10.5	22.1
	Standard 8	78.6	15.3	1.1	20.0	1.1
	Standard 9	83.0	10.1	16.2	16.7	47.7
	Standard 10	80.0	14.6	2.2	21.6	2.2
		Standards	Serum 1	Serum 2	House-dust 1	House-dust 2

Table 41: Summary table showing TPRs, calculated as "the number of correctly named compounds/the number of compounds now known to be in each mixture x 100", for all 4 (nano)ESI-DIMS assays combined. TPRs for standard mixtures, fortified and unfortified serum, and house-dust are shown. TPRs for the standards ranged from 53-83%, TPRs for combined assays for serum 2 ranged from 2-62%, and The TPR for house-dust 2 ranged from 2-54%.

FALSE POSITIVES (%)	Standard 1	1.1	664.2	60.0	948.4	186.3
	Standard 2	0.0	662.1	60.0	947.4	186.3
	Standard 3	1.1	657.9	58.9	948.4	185.3
	Standard 4	1.1	337.8	1.1	483.2	70.3
	Standard 5	2.1	664.2	2.1	949.5	136.8
	Standard 6	20.0	168.5	9.6	241.1	34.8
	Standard 7	24.2	663.2	47.4	953.7	168.4
	Standard 8	0.5	159.2	15.6	231.0	48.5
	Standard 9	2.5	164.4	0.5	234.2	1.9
	Standard 10	0.0	329.7	30.8	473.5	95.7
		Standards	Serum 1	Serum 2	House-dust 1	House-dust 2

Table 42: Summary table showing FPRs, calculated as "the number of incorrectly named compounds/the number of compounds now known to be in each mixture x 100", for all 4 (nano)ESI-DIMS assays combined. FPRs for standard mixtures, fortified and unfortified serum, and house-dust are shown. FPRs for the standard mixtures ranged from 0-24%, FPRs for serum 2 ranged from 1-60%, and FPRs for house-dust 2 ranged from 2-186%.

Using smaller reference lists increased TPRs and decreased FPRs for each sample type. Furthermore, using low FPRs to indicate presence of real analytes in a dataset can be valuable if correspondingly high TPRs for the same sample type exist. (nano)ESI-DIMS assays can successfully be applied for the non-targeted analysis of xenobiotics in a clean matrix using a larger reference list, but analysis of xenobiotics in more chemically complex sample matrices such as serum and house-dust requires the use of smaller reference lists, which greatly improve the TPR whilst decreasing the FPR. Moreover, these methods revealed that lower TPRs are achieved for low purity samples such as standard mixture 6, likely due to ion suppression.

3.3 Conclusions

Ten standard mixtures and fortified and unfortified serum and house-dust samples, were analysed using four polar and non-polar, positive and negative (nano)ESI-DIMS datasets as part of a ring trial called the US-EPA's non-targeted analysis collaborative trial (ENTACT). The samples were each fortified with a different standard mixture. The experiment was carried out in two phases. At Phase I, sample compositions, including which of each pair of samples was fortified and with which standard mixture, were unknown to the analyst, and annotation of xenobiotics was carried out against the ToxCast library with 4,462 chemicals. N.B. the sample constituents were known to be present in this library, alongside other chemicals that

were not present in the samples. At Phase II, sample compositions were revealed, and small databases containing only the chemicals revealed to be in each standard mixture were created, and annotation of the four (nano)ESI-DIMS datasets repeated, this time against these much smaller databases. The objectives of this research were to evaluate how well these methods performed at detecting xenobiotics in increasingly more complex sample matrixes. To assess method performance, TPRs and FPRs for each sample type were calculated for the standard mixtures, fortified and unfortified serum and house-dust samples. Method performance was also evaluated by assessing how effectively (nano)ESI-DIMS methods could determine which of two serum and house-dust samples had been fortified, and with which ENTACT standard mixture.

At Phase I experiments, TPRs for the standard mixtures ranged from 48-74% FPRs ranged from 253-879%. Phase I results for the serum samples successfully pointed to serum 2 being fortified with standard mixture 5 (43% TPR, 956% FPR). Phase I results for the house-dust samples could not successfully suggest which standard mixture had been spiked and into which of the 2 house-dust samples. At Phase II, the standard mixtures had higher TPRs relative to Phase I results, ranging from 53-83%. The FPRs were significantly lower, ranging from 0-24%. Phase II results for the serum correctly suggested standard mixture 5 was spiked into serum 2, with a TPR of 62% and a FPR of only 2%. Indeed, it appeared that when using small reference lists, a low FPR alongside a high TPR can improve accuracy of results. This was demonstrated with the house-dust samples. Phase II results for the house-dust samples incorrectly suggested that standard mixture 5 had been spiked into house-dust 2 when only TP were considered. However, when FP were also considered, then standard mixture 9 could correctly be concluded to have been spiked into house-dust 2. The TPR for house-dust 2 fortification with standard mixture 9 was 48% and the FPR was only 2%. By contrast, the TPR for house-dust 2 fortification with standard mixture 5 was 54% but the FPR was 137%.

These methods can be applied with varying levels of success depending on sample complexity and data processing strategies, specifically the use of large or small reference lists. Phase I results, which used a large reference list containing 4,462 chemicals, demonstrated that chemicals can successfully be detected in a clean matrix with high TPRs, but confidence in results are lowered by antagonistically high FPRs. The ToxCast database contains 4,462 chemicals of which 1,570 chemicals could be matched by empirical formula to another

chemical in the database and were therefore isomers. Moreover, *m/z* values with small differences (<5ppm) between them can be annotated as one another during database matching or form different ion forms that fall within a small enough mass error tolerance that they are annotated incorrectly. Consequently, increasing sample complexity decreased the TPRs achieved and further increased the already high FPRs. Detecting xenobiotics in a serum matrix is possible but use of alternative methods that offer less ambiguous results is recommended when using large reference lists, which often contain isomers, for annotation. Detection of xenobiotics in a house-dust matrix is not possible with a high level of confidence and offers the highest FP annotations. The use of smaller reference lists in Phase II both increased TPRs and decreased FPRs. Moreover, an apparent relationship between TPR and FPR was observed. Given the database searched was not large, a low FPR alongside a high TPR improved accuracy of results, accurately determining the standard mixture fortified into the serum and house-dust matrices. Indeed, this strategy allowed successful screening of xenobiotics in a house-dust matrix. Unfortunately, most NTA have no knowledge of which analytes to expect in their samples, and thus cannot calculate TPR and FPR. However, some strategies for estimating the false discovery rate in metabolite annotation and identification are under development in metabolomics, such as the target-decoy strategy for estimating false discovery (398–402).

Data presented in this chapter demonstrate the advantages and disadvantages of (nano)ESI-DIMS methods, which offer faster analysis times (a maximum of ca. 20 minutes per sample) than the more established hyphenated mass spectrometry methods, which can take over an hour per sample, making it suitable as a high throughput screening (HTS) approach (216). It has, however, been demonstrated that electrospray instability, which can be caused by improper or inadequate sample preparation, can increase acquisition times whilst the C-trap collects sufficient ion counts, and can indeed exclude certain sample types from analysis, as was observed with the wrist-band extracts which could not reproducibly be analysed using (nano)ESI-DIMS methods.

The lack of separation comes at an exceedingly high cost. Isobars and isomers are impossible to distinguish, FPRs are significantly higher and ion suppression is a challenge. This means that all putative results require confirmation using a hyphenated mass spectrometry method (separation using liquid or gas chromatography, capillary electrophoresis, ion mobility or

supercritical fluid chromatography). This requires two separate analyses to be employed, which can increase cost and weakens the advantage of a fast analysis time. More complex sample matrices compound the problem as greater numbers of ions are detected. The use of overlapping SIM windows partially counters this issue by only observing segments of the full mass range at a time, whilst employing the highest possible mass resolution helps to distinguish isotopes and molecules with very similar masses. Isotope abundances have been reported to be a useful companion for low mass error tolerance databases matching, reducing >95% of elemental composition candidates during database matching (403).

Finally, data processing strategies can vastly influence the number of annotation hits against each observed *m/z* peak. Isotopic peak patterns can help to discern real features from the background, and smaller reference lists can decrease the number of FP. Construction of smaller reference lists is highly recommended. However, for NTA, which aim to detect and annotate many thousands of known unknowns, this is a tall order. Construction of smaller reference lists requires comprehensive information about sample compositions, which is seldom available. Great efforts have gone into creating more specific databases. For example, clinical metabolomics studies use databases such as the Human Metabolome Database (74), whilst exposomics studies have access to the BloodExposome database (404). However, both these still contain large numbers of compounds in them (~114 000 and ~66,000 respectively). Small databases used in the Phase II work presented herein were tiny in comparison, ranging from 95-365 compounds in each. The use of databases this small is a real-life study is not feasible, and if information about sample compositions is known in this much detail, then TA are more appropriate. This brings into question just how appropriate (nano)ESI-DIMS analysis of xenobiotics in serum and house-dust matrices is. Indeed, without the use of very small reference lists, these methods cannot be applied successfully for the analysis of xenobiotics, whilst in reality very small reference lists are not feasible to construct. However, these methods can be used for screening against very large databases to provide smaller suspect lists that can be used for more accurate annotation, particularly since it has been shown that smaller databases improve TPRs of annotation.

Analysis of xenobiotics in complex sample matrices is not recommended using only (nano)ESI-DIMS methods. Rather, hyphenated mass spectrometry methods can, in principle, offer much more useful results, with lower FP and higher confidence in results by using additional

information (e.g., RT data) and U(H)PLC-MS is recommended for the analysis of xenobiotics in complex sample matrices. However, use of instruments with higher resolution, such as some newer Orbitrap (Fusion, IDX, and Exploris480) instruments which offer resolutions of up to 1 million, could greatly improve the performance of (nano)ESI-DIMS methods, although isomers would remain a challenge. Indeed, the next chapter will perform the same types of investigations, using the same samples, but using U(H)PLC-MS methods instead, to determine how well these methods perform at analysing xenobiotics in solutions and complex samples.

4 CHAPTER 4: Assessing the Performance of U(H)PLC-MS During the Non-targeted Analysis of Xenobiotics in Different Matrices as Part of the US-EPA's Non-Targeted Analysis Collaborative Trial (ENTACT)

Abstract

It has been shown that nano-electrospray ionisation direct infusion mass spectrometry ((nano)ESI-DIMS) can be applied for the analysis of xenobiotics but yields high false positive rates (FPRs). Ultra-high performance liquid chromatography-mass spectrometry (U(H)PLC-MS) is the most applied analytical method in untargeted metabolomics and offers additional retention time (RT) information that may reduce FPRs. We participated in a ring trial to determine if in-house U(H)PLC-MS metabolomics methods can be used to screen for xenobiotics, with the aim to show that additional RT information can vastly reduce FPRs relative to (nano)ESI-DIMS.

Four U(H)PLC-MS methods (aqueous C_{18} reversed phase and HILIC positive and negative ion-modes) were applied for analysis of ten xenobiotic mixtures in DMSO, and fortified and unfortified serum, house-dust, and wrist-band matrices. Experiments were in two Phases. At Phase I, sample compositions were unrevealed and data annotation used a large suspect screening list containing 4,462 chemicals. At Phase II, sample compositions were revealed, and therefore RT databases created to aid annotation.

At Phase I, standard mixtures yielded TPRs of up to 94% which declined in the more complex matrices, albeit still above 75%, far outperforming the (nano)ESI-DIMS methods. However, the FPRs were high (up to 3,500%), even higher than those achieved for the (nano)ESI-DIMS methods at Phase I (up to ~1,700% for the complex sample matrices).

At Phase II, the coverage of the RT databases created was only ~50%, limited by a lack of reproducible detection of some chemicals. Relative to Phase I experiments, TPRs and FPRs for both the clean and complex matrices decreased when RT databases were used. The U(H)PLC-MS metabolomics methods could be successfully applied for the analysis of xenobiotics. However, relative to the (nano)ESI-DIMS methods, the FPRs were much lower due to the addition of RT information. Such additional data serves as a great advantage and is recommended to be collected and used for NTA where possible.

4.1 Introduction

The challenges of chemical pollution and strategies for measuring environmental chemical pollution have been introduced and discussed (see sections 1.3, 1.3.1, and 3.1). Indeed, non-targeted analyses (NTA) are increasingly applied for measuring xenobiotics in the environment (see sections 1.4 and 3.1). With this increased application has come a realisation that these methods' performances have not been sufficiently evaluated. Indeed, the chemical coverage of these methods is yet uncharacterised, whilst it is also unknown how reproducibly these methods detect chemicals. To fill this gap in knowledge, a ring trail, introduced extensively in 3.1, was designed by the US Environmental Protection Agency (EPA). The University of Birmingham metabolomics team participated in the ring trial, whose objectives were to characterise the performance of NTA when analysing xenobiotics in various matrices. The genesis, experimental design, and objectives of this ring trial, called EPA's non-targeted analysis collaborative trial (ENTACT) have already been introduced in Chapter 3 (section 3). This chapter is a continuation of the ENTACT project. In Chapter 3, (nano)ESI-DIMS methods were applied for the analysis of xenobiotics in matrices of varying levels of complexity. This chapter will be report identical experiments as those in chapter 3 but will instead apply U(H)PLC-MS methods to analyse the same samples.

Specific research goals are detailed below:

In each of four U(H)PLC-MS assays, using the putative annotation of chemicals based on *m/z* only in each standard mixture, serum, house-dust, and wrist-band extract, matched against the 4,462 ToxCast chemicals reference list:

1. Of the putatively annotated chemicals, how many were correct (true positives (TP) or true positive rate (TPR))?
2. Of the putatively annotated chemicals, how many were incorrect (false positives (FP) or false positive rate (FPR))?
3. Can U(H)PLC-MS methods be used to determine which of each pair of reference material extracts (serum 1 and 2, house-dust 1 and 2 and wristband 1 and 2), was spiked with one of the 10 standard mixtures?
4. Can U(H)PLC-MS methods be used to determine with which of the 10 standard mixtures each sample was fortified with?

5. For each of the four U(H)PLC-MS assays, comparing TP and FP for simple vs increasingly more complex sample matrices (i.e., standards in DMSO vs serum, house-dust then wrist-band extracts), is there a relationship between the number of TP annotations and sample complexity?
6. Comparing TP and FP of Phase I and Phase II results (very large vs very small databases, and with and without retention time (RT) information), is there a relationship between the number of TP annotations, the size of the databases searched and the use of RT data?

4.2 Results and Discussion

Ten standard mixtures, each with between 95-365 compounds, were analysed using four U(H)PLC-MS assays (HILIC and aqueous C_{18} reversed phase (aqRP C_{18}) positive and negative ion-modes) as part of ENTACT Phase I experiments. The ten standard mixtures were annotated against the ToxCast reference list containing 4,462 chemicals using MS1 data. Annotation carried out at this Phase I stage did not use RT data as this was the blinded phase and RT data was not available. Method performance was evaluated after sample compositions were revealed at Phase II of ENTACT experiments by calculating TPRs and FPRs for each standard mixture. TPRs and FPRs were calculated as described in 2.2.13, and annotation was carried out as described in 2.3.8.

Two serum, house-dust, and wristband samples, one fortified and one unfortified, henceforth referred to as serum, house-dust, and wristband 1 and 2, were analysed using four U(H)PLC-MS assays (HILIC and aqRP C_{18} positive and negative ion-mode assays) and annotated by matching against the ToxCast library containing 4,462 chemicals. During Phase I experiments, it was unknown which of each pair of samples was fortified and with which of the ten standard mixtures. TPRs and FPRs were calculated, sequentially assuming each of ten standard mixtures had been spiked first into sample 1 and then into sample 2. To determine which sample was fortified, TPRs were considered, assuming that where a real fortification had been made TPRs would be higher than when a fortification had not been made. Retention time information was not used for annotations in Phase I experiments.

TPRs and FPRs were calculated as described in 2.2.13. TPRs and FPRs were calculated only after sample compositions were revealed at Phase II of ENTACT experiments. Method performance was evaluated by considering if the method was able to determine which serum sample was fortified, and with which standard mixture. Results for each sample type are presented below.

4.2.1 PHASE I (Blinded): Method Performance

4.2.1.1 Standard Mixtures

TPRs for the standard mixtures are shown in Table 43. Generally, the aqRP C_{18} assays performed better than the HILIC assays, yielding higher TPRs. Standard mixtures 6 and 7 had very few or no correct annotations for all assays. It was observed that standard 7 was a subset of standard 6. All 95 compounds found in standard 7 were also found in standard 6, which also had 270 other compounds for a total of 365 compounds. To investigate why these 2 standard mixtures had very low or no TP annotations, the number of peaks appearing in 3 out of 3 MS1 technical replicates for each standard mixture were calculated. It was found that standard mixture 6, which had 365 sample constituents always had the highest number of peaks detected, whilst standard mixture 7, which had 95 chemical constituents, always had one of the lowest peak counts detected (Table 44).

		AqRP Positive	AqRP Negative	HILIC Positive	HILIC Negative
TRUE POSITIVES (%)	Standard 1	54.7	17.9	23.2	23.2
	Standard 2	80.0	31.6	38.9	34.7
	Standard 3	70.5	35.8	34.7	25.3
	Standard 4	35.7	21.6	17.3	16.2
	Standard 5	73.7	32.6	40.0	34.7
	Standard 6	1.1	1.1	0.3	0.8
	Standard 7	0.0	0.0	0.0	0.0
	Standard 8	81.1	20.5	41.1	30.1
	Standard 9	61.4	46.6	29.3	22.7
	Standard 10	40.3	20.3	23.6	17.5

Table 43: TPRs for the aqueous C_{18} reversed phase positive and negative ion-modes, and the HILIC positive and negative ion-modes, ranging from 0-81%, 0-47%, 0-41%, and 0-35%, respectively.

	AqRP_Pos	AqRP_Neg	HILIC_POS	HILIC_NEG
Standard Mixture 1	1999	801	518	714
Standard Mixture 2	1806	715	495	660
Standard Mixture 3	2042	1327	475	591
Standard Mixture 4	3031	1732	551	1030
Standard Mixture 5	1884	904	425	746
Standard Mixture 6	5550	2545	945	1266
Standard Mixture 7	2097	844	497	599
Standard Mixture 8	4218	960	821	833
Standard Mixture 9	4863	2741	955	885
Standard Mixture 10	3288	1139	716	667

Table 44: Peak counts are shown for 10 standard mixtures analysed using 4 U(H)PLC-MS assays. Peak counts shown here are those appearing in 3 out of 3 MS1 injections of each standard mixture. Standard mixture 6 has the highest peak counts for all 4 assays, whilst standard mixture 7 has one of the lowest peak counts for all 4 assays. High peak counts are highlighted in peach/pink whilst relatively low peak counts are highlighted in green. It has been reported that these 2 standard mixtures were designed to be more challenging, comprising isomers, isobars, and low molecular weight compounds (405). It is not further investigated why the TPRs were low as this was likely due to lack of selectivity and specificity for these purposefully challenging mixtures.

It was reported in a recent publication that standard mixtures 6 and 7 had been designed to be more challenging (see Figure 24) (405). Standard mixture 6 was a low purity mixture whilst standard mixture 7 contained only isobaric analytes. For standard mixture 6, the high peak count was a result of the low purity of the mixture. As such, the mixture components were likely not detected due to ion suppression, or the few that were detected were then incorrectly annotated due to the impurity of the mixture. For standard mixture 7, the low peak count could be lack of sufficient chromatographic resolution of isobaric analytes. Ion suppression due to lack of chromatographic resolution of isomers and isobars and due to co-eluting contaminants is supported by the fact that TPRs were lowest for these same standard mixtures when the nano(ESI)-DIMS methods were applied (48 and 50% respectively). However, the chemicals were detected. The (nano)ESI-DIMS methods offer increased sensitivity, whilst data were collected in SIM windows to further reduce ion suppression. Detection of chemicals in standard mixtures 6 and 7 using the more sensitive (nano)ESI-DIMS methods strongly suggests that ionisation suppression occurred here, impeding detection when applying the less sensitive U(H)PLC-MS methods. Method optimisation, such as altering the solvent gradient, is required to improve chromatographic resolution.

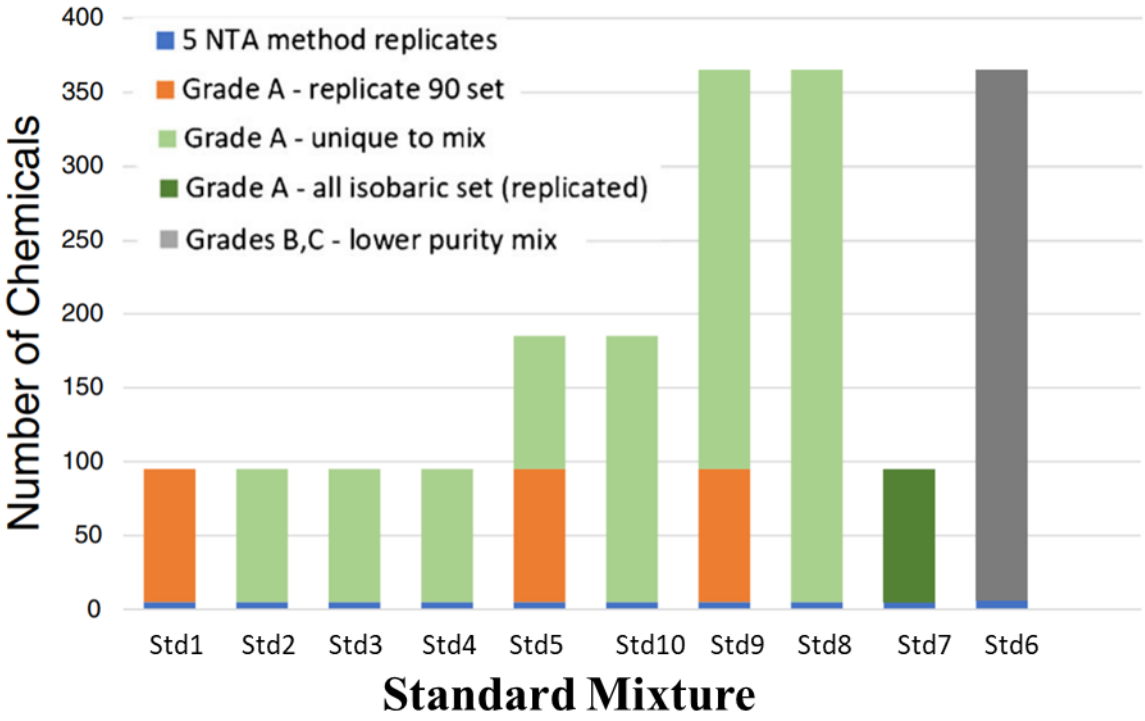


Figure 24: The ten ENTACT standard mixtures are listed. Standard mixture 6 was a mixture of low purity, whilst standard mixture 7 contained all isobaric analytes.

FPRs for the standard mixtures are shown in Table 45. The maximum FPRs observed for the aqRP C_{18} assays were twice as large as those observed for the HILIC assays. Large peak lists can complicate the annotation process. In particular, the more peaks detected in an assay, the greater the importance of selecting appropriate annotation parameters such as RT alignment and mass error tolerance. As such, mis annotation is more likely to occur when peak matrices are larger. Indeed, the aqRP C_{18} positive ion-mode assay yielded four times as many features for annotation as the HILIC positive ion-mode assay, likely because more analytes ionised in this ion-mode (Table 46). The higher number of detected features also partially explains the higher TPRs observed for the aqRP C_{18} (Table 43), simply suggesting that more chemicals within the mixtures were detected using this assay.

The number of features for the aqRP C_{18} assays may seem uncharacteristically large. However, this can be explained. Ten standard mixtures were analysed using each assay, and then processed simultaneously. Although there is some overlap between each standard mixture, there were also multiple standard mixtures that contained unique chemicals. As a result, processing these ten standard mixtures together yielded multiple missing values, and a very long peak list for each assay. Additionally, a SNR of 3 was applied during data processing, as such some noise peak may also have been present.

		AqRP Positive	AqRP Negative	HILIC Positive	HILIC Negative
FALSE POSITIVES (%)	Standard 1	1845.3	820.0	498.9	690.5
	Standard 2	1734.7	697.9	461.1	642.1
	Standard 3	1918.9	1354.7	452.6	589.5
	Standard 4	1475.7	874.6	267.0	498.4
	Standard 5	1826.3	851.6	386.3	693.7
	Standard 6	1411.5	682.2	240.8	327.1
	Standard 7	2081.1	854.7	516.8	611.6
	Standard 8	1010.4	236.2	180.8	192.6
	Standard 9	1186.6	679.2	224.1	207.1
	Standard 10	810.7	286.6	165.5	160.5

Table 45: FPRs for the aqueous C_{18} reversed phase positive and negative ion-modes, and the HILIC positive and negative ion-modes, ranging from 811-2081%, 236-1355%, 166-517%, and 161-691%, respectively.

	AqRP_Pos	AqRP_Neg	HILIC_POS	HILIC_NEG
Standards	80413	51335	20925	19408
Serum	80784	52427	21001	19577
House-dust	81139	52744	21089	19574
Wrist-band	80451	51992	20880	19515

Table 46: The number of features submitted for annotation, after XCMS processing and filtering is shown. The aqueous C_{18} reversed phase positive ion-mode had 4 times as many features as the HILIC assays, whilst the aqueous C_{18} reversed phase negative ion-mode assay had ~twice as many features as the HILIC assays.

4.2.1.2 Fortified and Unfortified Serum

TPRs for serum 1 are shown in Table 47. The highest TPRs for the aqRP C_{18} assays suggested standard mixture 5 was applied for serum fortification. For the HILIC positive ion-mode, the highest TPR was for standard mixture 3, whilst for the HILIC negative ion-mode it was for standard mixture 2 and 5. However, TPRs were generally low, casting doubt on serum 1 being fortified. Moreover, the varying outcomes for the highest TPRs further cast doubt on whether serum 1 had been fortified, since it was known that only one standard mixture had been spiked into the samples.

TPRs for serum 2 are shown in Table 48. The maximum TPRs for each assay were immediately observed to be higher than those for serum 1, suggesting that serum 2 had been fortified. The highest TPRs were observed for standard mixture 5 suggesting that this standard mixture had been applied for fortification of serum 2.

		AqRP Positive	AqRP Negative	HILIC Positive	HILIC Negative
TRUE POSITIVES (%)	Standard 1	9.5	14.7	3.2	9.5
	Standard 2	8.4	10.5	5.3	10.5
	Standard 3	8.4	15.8	8.4	8.4
	Standard 4	3.2	4.9	3.2	4.3
	Standard 5	15.8	17.9	3.2	10.5
	Standard 6	0.8	0.8	0.3	0.3
	Standard 7	0.0	0.0	0.0	0.0
	Standard 8	9.0	7.1	3.3	6.6
	Standard 9	8.5	6.6	2.5	6.3
	Standard 10	3.8	4.7	2.2	3.3

Table 47: TPRs calculated to test if serum 1 had been fortified and with which of ten standard mixtures. TPRs ranged from 0-16%, 0-18%, 0-8%, and 0-11% for the aqueous C_{18} reversed phase positive and negative ion-modes, and the HILIC positive and negative ion-modes, respectively. The maximum TPRs for the aqueous C_{18} reversed phase assays suggested that standard mixture 5 had been spiked into serum 1. However, the HILIC positive ion-mode assay had the highest TPR for standard mixture 3 whilst the HILIC negative ion-mode assay had the highest TPR for standard mixtures 2 and 5. The low TPRs and the varying conclusions from the different assays suggested that serum 1 had not been fortified.

		AqRP Positive	AqRP Negative	HILIC Positive	HILIC Negative
TRUE POSITIVES (%)	Standard 1	20.0	13.7	6.3	11.6
	Standard 2	11.6	7.4	8.4	13.7
	Standard 3	7.4	14.7	8.4	8.4
	Standard 4	3.8	4.3	3.2	5.4
	Standard 5	51.6	25.3	24.2	30.5
	Standard 6	1.1	0.8	0.3	0.8
	Standard 7	0.0	0.0	0.0	0.0
	Standard 8	11.2	5.5	5.5	8.5
	Standard 9	9.9	6.6	3.6	8.2
	Standard 10	6.8	4.4	3.3	4.9

Table 48: TPRs wwre calculated to test if serum 2 had been fortified and with which standard mixture. These ranged from 0-52%, 0-25%, 0-24% and 0-31% for the aqueous C_{18} reversed phase positive and negative ion-modes, and the HILIC positive and negative ion-modes, respectively. Not only were the maximum TPRs for serum 2 immediately noted to be higher than serum 1, but all 4 assays had the highest TPRs for standard mixture 5, suggesting that serum 2 had been fortified with standard mixture 5.

The experimental design for this ring trial stated that only one standard mixture had been spiked into one of the serum samples. Therefore, the TP annotations observed for all 10 standard mixtures highlight that the annotation of isomeric chemicals could be a significant problem. Without chromatographic resolution and resultant RT data to differentiate them, annotation of U(H)PLC-MS data relies solely on annotation applying *m/z* data only, thus isomers cannot be differentiated from one another. Compounding the issue is the fact that many isomers can often not be resolved chromatographically anyway, requiring methods such as different stationary phases to fully resolve them (407). The presence of "TPs" in both serum samples demonstrates the challenges of isomers in NTA. Figure 25 shows detection of diethylstilbestrol dipropionate (DTXSID7047146) found in standard mixture 1 in a clean solvent matrix. This chemical has a chemical formula of $C_{24}H_{28}O_4$ and a monoisotopic mass of 380.1988. A peak with the same RT and mass (detected in its protonated form) is detected in serum 1, and therefore annotated as diethylstilbestrol dipropionate (DTXSID7047146) (Figure 26). However, results suggest that serum 1 was not fortified. This annotation, marked as a TP, is therefore incorrect and demonstrates how isomers cannot be differentiated my *m/z* alone. Chromatographic separation is required, but even when chromatography is applied, resolution of two or more isomers may sometimes not be achieved. Moreover, RT knowledgebases are unavailable or are incomplete because chemical standards for all possible chemicals present may not be available to purchase or the cost makes it prohibitive to buy.

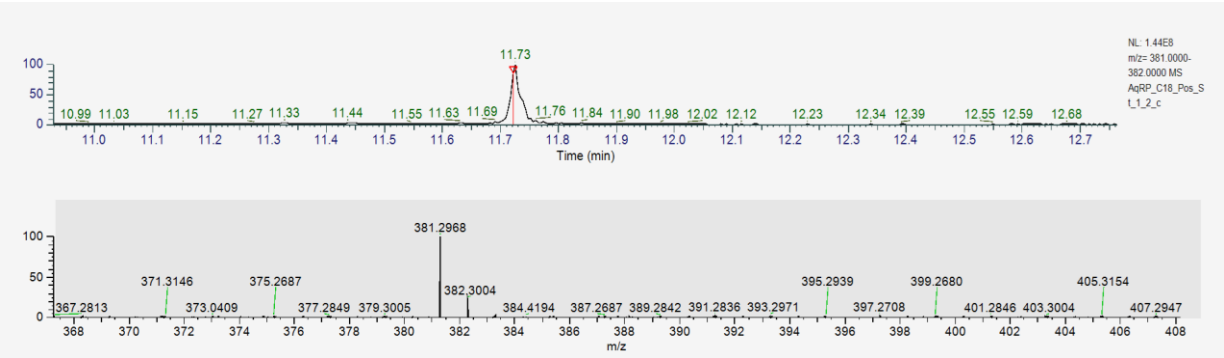
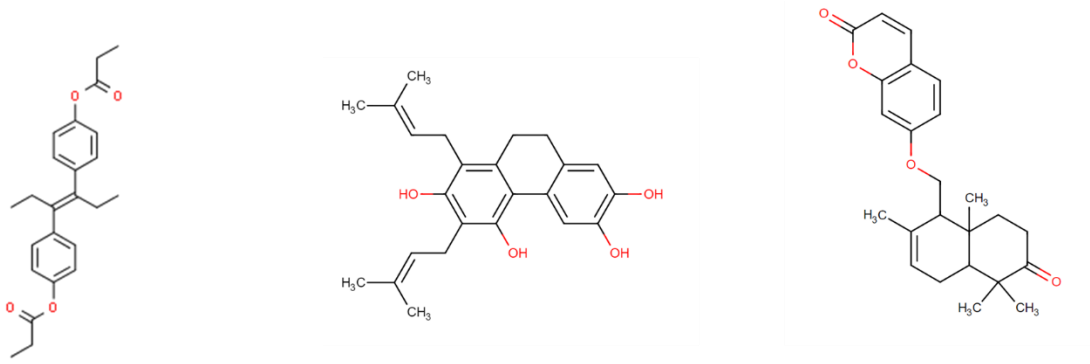


Figure 25: Extracted ion chromatogram (EIC) and mass spectrum for diethylstilbestrol dipropionate (DTXSID7047146) found in standard mixture 1 in a clean solvent matrix. This chemical has a chemical formula of $C_{24}H_{28}O_4$ and a monoisotopic mass of 380.1988.



Diethylstilbestrol dipropionate
DTXSID7047146

Gancaonin U HMDB0037587

Mass=380.1987594
MF=C24H28O4

Conferone HMDB36621

Figure 26: Diethylstilbestrol dipropionate (DTXSID7047146) was annotated in serum 1, a sample suspected to not have been fortified with any chemicals. A search of matching molecular formulae found 9 biological isomers of this chemical in HMDB that could be found in serum. Two are shown as examples here (gancaonin and conferone). This poses a challenge for the analysis of xenobiotics in biological matrices.

FPRs for serum 1 and 2 are shown in Table 49 and Table 50. FPRs for serum 2, which was fortified, were slightly higher than for the unfortified serum 1. Nonetheless, the U(H)PLC-MS assays could successfully be applied for the analysis of xenobiotics in a serum matrix using a reference list of 4,462 chemicals and without using any RT data applied during the annotation process, correctly determining that serum 2 had been fortified with standard mixture 5.

		AqRP Positive	AqRP Negative	HILIC Positive	HILIC Negative
FALSE POSITIVES (%)	Standard 1	734.7	763.2	458.9	587.4
	Standard 2	735.8	767.4	456.8	586.3
	Standard 3	735.8	762.1	453.7	588.4
	Standard 4	378.9	394.6	234.1	302.2
	Standard 5	728.4	760.0	458.9	586.3
	Standard 6	192.9	201.6	120.0	155.1
	Standard 7	744.2	777.9	462.1	596.8
	Standard 8	184.7	195.3	117.0	148.8
	Standard 9	185.2	195.9	117.8	149.0
	Standard 10	189.9	197.8	118.1	152.1

Table 49: FPRs for serum 1 ranged from 185-744%, 196-778%, 117-462%, and 149-588% for the aqueous C_{18} reversed phase positive and negative ion-modes, and the HILIC positive and negative ion-modes, respectively.

		AqRP Positive	AqRP Negative	HILIC Positive	HILIC Negative
FALSE POSITIVES (%)	Standard 1	920.0	769.5	538.9	814.7
	Standard 2	928.4	775.8	536.8	812.6
	Standard 3	932.6	768.4	536.8	817.9
	Standard 4	478.9	397.8	276.8	418.9
	Standard 5	888.4	757.9	521.1	795.8
	Standard 6	243.6	203.0	141.6	214.2
	Standard 7	940.0	783.2	545.3	826.3
	Standard 8	233.4	198.4	136.4	206.6
	Standard 9	234.8	197.3	138.4	206.8
	Standard 10	237.8	199.5	138.6	210.1

Table 50: FPRs for serum 2 ranged from 233-940%, 197-783%, 136-545%, and 207-818% for the aqueous C_{18} reversed phase positive and negative ion-modes, and the HILIC positive and negative ion-modes, respectively. These were slightly higher than those observed for serum 1.

4.2.1.3 Fortified and Unfortified House-dust

TPRs for house-dust 1 are shown in Table 51. The maximum TPRs for the aqRP C_{18} assays were for standard mixture 5. The maximum TPR for the HILIC positive ion-mode was also for standard mixture 5 but was standard mixture 2 for the HILIC negative ion-mode assay. The TPRs were sufficiently high to not immediately discount house-dust 1 as having been fortified. However, the highest TPRs for the four assays did not at all suggest a single standard mixture as the fortifier. It was unclear if house-dust 1 had been fortified and with which standard mixture from these results.

TPRs for house-dust 2 are shown in Table 52. These TPRs were much higher than those observed for house-dust 1, suggesting that house-dust 2 had been fortified in the absence of additional evidence. Moreover, the maximum TPRs for all 4 assays pointed unanimously to house-dust 2 being fortified with standard mixture 5. It was therefore concluded that house-dust 2 had been fortified with standard mixture 5.

		AqRP Positive	AqRP Negative	HILIC Positive	HILIC Negative
TRUE POSITIVES (%)	Standard 1	49.5	46.3	23.2	27.4
	Standard 2	55.8	44.2	24.2	29.5
	Standard 3	51.6	47.4	23.2	24.2
	Standard 4	28.6	23.8	12.4	13.5
	Standard 5	55.8	54.7	26.3	28.4
	Standard 6	1.1	0.8	0.0	0.8
	Standard 7	0.0	0.0	0.0	0.0
	Standard 8	50.1	38.1	20.0	20.0
	Standard 9	39.2	33.2	15.6	19.5
	Standard 10	26.3	20.3	9.9	12.9

Table 51: TPRs for house-dust 1 ranged from 0-56%, 0-55%, 0-26%, and 0-30% for the aqueous C_{18} reversed phase positive and negative ion-modes, and the HILIC positive and negative ion-modes, respectively. The highest TPRs were for standard mixture 5 for the aqueous C_{18} reversed phase positive and negative modes, and for the HILIC positive ion-mode assay. The highest TPR for the HILIC negative ion-mode assay was for standard mixture 2. With multiple candidates suggested for which standard mixture had been spiked into house-dust 1, and with it being known that a single standard mixture was spiked into each sample, it was concluded that house-dust 1 had not been fortified.

		AqRP Positive	AqRP Negative	HILIC Positive	HILIC Negative
TRUE POSITIVES (%)	Standard 1	56.8	60.0	25.3	25.3
	Standard 2	58.9	53.7	22.1	32.6
	Standard 3	49.5	50.5	21.1	27.4
	Standard 4	28.1	24.9	11.4	15.7
	Standard 5	82.1	66.3	42.1	36.8
	Standard 6	1.1	1.1	0.0	0.8
	Standard 7	0.0	0.0	0.0	0.0
	Standard 8	55.3	42.2	21.6	22.2
	Standard 9	60.3	43.6	25.8	23.0
	Standard 10	28.2	24.1	10.1	14.0

Table 52: TPRs for house-dust 2 ranged from 0-82%, 0-66%, 0-42%, and 0-37% for the aqueous C_{18} reversed phase positive and negative ion-modes, and the HILIC positive and negative ion-modes, respectively. These were much higher than TPRs for house-dust 2, thus suggesting that house-dust 2 had been fortified and house-dust 1 had not. The highest TPRs were all for standard mixture 5, thus it was concluded that house-dust 2 had been fortified with standard mixture 5.

FPRs for house-dust 1 and 2 are shown in Table 53 and Table 54. FPRs for the fortified sample were higher than for the unfortified sample. This was observed with the serum samples as well and is likely due to a slight increase in the number of features detected for the fortified samples, but this was not investigated further as the TPR data were of primary interest.

The U(H)PLC-MS assays applied here correctly identified the fortified house-dust 2 sample, but incorrectly concluded that standard mixture 5 was spiked in (house-dust 2 was spiked with standard mixture 9). The incorrect conclusion can be explained by the overlap between standard mixture 5 and 9, where all 95 chemicals in standard mixture 5 were also found in standard mixture 9, which had an additional 270 other chemicals for a total of 365 chemicals. The percentage TPR was therefore higher for standard mixture 5 for overlapping chemicals found in both standard mixture 5 and 9 since the former had fewer compounds present, also suggesting that only the overlapping chemicals were detected in standard mixture 9 and not many of the additional constituents.

		AqRP Positive	AqRP Negative	HILIC Positive	HILIC Negative
FALSE POSITIVES (%)	Standard 1	5725.3	5737.9	2271.6	2710.5
	Standard 2	5718.9	5740.0	2270.5	2708.4
	Standard 3	5723.2	5736.8	2271.6	2713.7
	Standard 4	2936.8	2946.5	1165.9	1392.4
	Standard 5	5718.9	5729.5	2268.4	2709.5
	Standard 6	1501.9	1504.7	597.3	711.8
	Standard 7	5774.7	5784.2	2294.7	2737.9
	Standard 8	1452.9	1467.4	577.3	692.6
	Standard 9	1463.8	1472.3	581.6	693.2
	Standard 10	1476.7	1485.2	587.4	699.7

Table 53: FPRs for house-dust 1 ranged from 1453-5774%, 1467-5784%, 577-2295%, and 693-2738% for the aqueous C_{18} reversed phase positive and negative ion-modes, and the HILIC positive and negative ion-modes, respectively.

		AqRP Positive	AqRP Negative	HILIC Positive	HILIC Negative
FALSE POSITIVES (%)	Standard 1	6709.5	6457.9	2124.2	2935.8
	Standard 2	6707.4	6464.2	2127.4	2928.4
	Standard 3	6716.8	6467.4	2128.4	2933.7
	Standard 4	3446.5	3322.2	1092.4	1504.9
	Standard 5	6684.2	6451.6	2107.4	2924.2
	Standard 6	1760.0	1695.3	559.5	769.9
	Standard 7	6766.3	6517.9	2149.5	2961.1
	Standard 8	1705.8	1654.2	537.8	748.5
	Standard 9	1700.8	1652.9	533.7	747.7
	Standard 10	1732.9	1672.3	549.3	756.7

Table 54: FPRs ranged from 1701-6766%, 1653-6518%, 534-2150%, and 748-2961% for the aqueous C_{18} reversed phase positive and negative ion-modes, and the HILIC positive and negative ion-modes, respectively.

4.2.1.4 Fortified and Unfortified Wristband

TPRs for wristband 1 are shown in Table 55. These TPRs were sufficiently low that it was suspected that wristband 1 had not been fortified. The highest TPRs for both aqRP C_{18} assays were for standard mixture 5. The highest TPRs for the HILIC positive ion-mode assay were for standard mixtures 2 and 5, whilst the highest TPR for the HILIC negative ion-mode assay was for standard mixture 3. This lack of agreement about which standard mixture had been spiked into wristband 1 for each of the four assays led to the conclusion that wristband 1 had not been fortified.

TPRs for wristband 2 are shown in Table 56. The maximum TPRs for wristband 2 were noted to be higher than for wristband 1, suggesting that wristband 2 had been fortified. The highest TPRs for all four assays were for standard mixture 5, thus it was concluded that wristband 2 had been fortified with standard mixture 5. The individual U(H)PLC-MS assays applied here therefore correctly determined which wristband sample had been fortified but incorrectly suggested that standard mixture had 5 had been spiked in, when in-fact it was standard mixture 4.

FPRs for wristbands 1 and 2 are shown in Table 57 and Table 58. As with the other biological samples, it was noted that FPRs were slightly higher for the fortified wristband 2, again likely due to the increase in the number of compounds present in the mixture (matrix and chemicals in the standard mixture).

		AqRP Positive	AqRP Negative	HILIC Positive	HILIC Negative
TRUE POSITIVES (%)	Standard 1	13.7	8.4	9.5	14.7
	Standard 2	12.6	9.5	11.6	15.8
	Standard 3	13.7	8.4	8.4	17.9
	Standard 4	6.5	4.3	6.5	8.6
	Standard 5	16.8	13.7	11.6	12.6
	Standard 6	0.0	0.5	0.0	0.5
	Standard 7	0.0	0.0	0.0	0.0
	Standard 8	12.1	7.4	10.1	7.4
	Standard 9	11.0	5.8	5.8	7.1
	Standard 10	7.1	4.1	3.8	4.9

Table 55: TPRs for wristband 1 ranged from 0-17%, 0-14%, 0-12% and 0-18% for the aqueous C_{18} reversed phase positive and negative ion-modes, and the HILIC positive and negative ion-modes, respectively. The low TPRs alongside the highest TPRs for each assay suggesting different standard mixtures being spiked into wristband 1 led to the conclusion that wristband 1 had not been fortified.

		AqRP Positive	AqRP Negative	HILIC Positive	HILIC Negative
TRUE POSITIVES (%)	Standard 1	29.5	12.6	18.9	25.3
	Standard 2	22.1	8.4	21.1	23.2
	Standard 3	28.4	5.3	14.7	23.2
	Standard 4	29.7	3.2	18.4	14.6
	Standard 5	55.8	22.1	34.7	34.7
	Standard 6	0.5	0.8	0.3	1.1
	Standard 7	0.0	0.0	0.0	0.0
	Standard 8	15.6	6.0	15.1	10.7
	Standard 9	15.9	3.6	8.5	10.4
	Standard 10	9.6	4.7	5.8	6.8

Table 56: TPRs ranged from 0-56%, 0-22%, 0-35% and 0-35% for the aqueous C_{18} reversed phase positive and negative ion-modes, and the HILIC positive and negative ion-modes, respectively. The maximum TPRs were higher than for wristband 1, thus it was concluded that this sample had been fortified. Moreover, the maximum TPRs for all 4 assays pointed to standard mixture 5 being spiked into wristband 2, thus this conclusion was made.

		AqRP Positive	AqRP Negative	HILIC Positive	HILIC Negative
FALSE POSITIVES (%)	Standard 1	871.6	897.9	1201.1	1072.6
	Standard 2	872.6	896.8	1198.9	1071.6
	Standard 3	871.6	897.9	1202.1	1069.5
	Standard 4	448.1	461.1	615.1	549.7
	Standard 5	868.4	892.6	1198.9	1074.7
	Standard 6	230.4	235.3	315.1	282.5
	Standard 7	885.3	906.3	1210.5	1087.4
	Standard 8	218.4	228.5	304.9	275.6
	Standard 9	219.5	230.1	309.3	275.9
	Standard 10	223.3	231.8	311.2	278.1

Table 57: FPRs for wristband 1 ranged from 218-885%, 229-906%, 305-1211%, and 276-1087% for the aqueous C_{18} reversed phase positive and negative ion-modes, and the HILIC positive and negative ion-modes, respectively.

		AqRP Positive	AqRP Negative	HILIC Positive	HILIC Negative
FALSE POSITIVES (%)	Standard 1	1489.5	542.1	1490.5	1365.3
	Standard 2	1496.8	546.3	1488.4	1367.4
	Standard 3	1490.5	549.5	1494.7	1367.4
	Standard 4	750.3	281.6	756.8	699.5
	Standard 5	1463.2	532.6	1474.7	1355.8
	Standard 6	394.8	143.6	392.6	360.8
	Standard 7	1518.9	554.7	1509.5	1390.5
	Standard 8	379.7	138.4	377.8	351.2
	Standard 9	379.5	140.8	384.4	351.5
	Standard 10	385.8	139.7	387.1	355.1

Table 58: FPRs for wristband 2 ranged from 380-1519%, 140-555%, 378-1510%, and 351-1391% for the aqueous C_{18} reversed phase positive and negative ion-modes, and the HILIC positive and negative ion-modes, respectively.

4.2.2 PHASE I: Summary

All four assay annotations were combined for each sample type, and the results are summarised. Standard mixtures had TPRs ranging from 0-94% (Table 59). Standard mixture 6 and 7 had very few or no correct annotations as they were of low purity (standard 6) and comprised of isobars and isomers (standard 7) (405), lowering the minimum TPRs observed. The TPRs for the combined U(H)PLC-MS assays were higher than the combined (nano)ESI-DIMS assays at Phase I of ENTACT experiments, which ranged from 48-74%. Successful analysis of xenobiotics in a clean solvent matrix can be performed using the aqRP C_{18} and HILIC assays and matching against a large reference list of 4,462 compounds without using RT data. However, isomers and isobars are challenging to separate using these assays. Moreover, low purity mixtures cannot be analysed successfully. Where these (isomers and isobars) are suspected to be prevalent in the sample, some method development is advised to ensure chromatographic selection of isomers and isobars (408).

		Standards	Serum 1	Serum 2	House-dust 1	House-dust 2	Wristband 1	Wristband 2
TRUE POSITIVES (%)	Standard 1	69.5	30.5	46.3	77.9	84.2	32.6	54.7
	Standard 2	89.5	26.3	36.8	82.1	84.2	37.9	48.4
	Standard 3	86.3	26.3	30.5	80.0	80.0	35.8	54.7
	Standard 4	88.1	27.6	50.3	75.7	87.0	33.0	76.8
	Standard 5	83.2	33.7	74.7	75.8	93.7	32.6	75.8
	Standard 6	1.1	1.1	1.1	0.5	1.1	0.5	1.1
	Standard 7	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Standard 8	89.9	22.2	23.8	67.9	73.2	26.8	34.2
	Standard 9	92.9	27.4	40.0	72.6	94.2	28.2	44.9
	Standard 10	93.5	21.6	29.2	73.5	77.8	28.1	34.6

Table 59: TPRs for the standard mixtures, ranging from 0-94%. These TPRs are higher than those observed for (nano)ESI-DIMS experiments at Phase I. The aqueous C_{18} reversed phase and HILIC assays can therefore be successfully applied for the analysis of xenobiotics in a clean solvent matrix, outperforming (nano)ESI-DIMS. TPRs for serum 1 and 2 ranged from 0-34% and 0-75%, respectively. TPRs for house-dust 1 and 2 ranged from 0-82% and 0-94%, respectively. Finally, for the combined wristband 1, the TPRs ranged from 0-38% whilst for wristband 2 they ranged from 0-77%.

FPRs for the standard mixtures are shown in Table 60. These were significantly higher than those observed in the (nano)ESI-DIMS combined assays at Phase I of ENTACT experiments, which were 254-879%. This is likely due to the different data acquisition and processing strategies (see 2.2.10 and 2.2.11). Despite the higher FPRs, TPRs for U(H)PLC-MS data still outperform (nano)ESI-DIMS methods, even without the presence of RT information in the database used for annotation (ToxCast).

		Standards	Serum 1	Serum 2	House-dust 1	House-dust 2	Wristband 1	Wristband 2
FALSE POSITIVES (%)	Standard 1	1789.5	1453.7	1553.7	3374.7	3530.5	1747.4	1937.9
	Standard 2	1701.1	1457.9	1563.2	3370.5	3530.5	1742.1	1944.2
	Standard 3	2030.5	1457.9	1569.5	3372.6	3534.7	1744.2	1937.9
	Standard 4	1064.3	734.6	771.4	1697.3	1769.2	881.1	946.5
	Standard 5	1627.4	1450.5	1525.3	3376.8	3521.1	1747.4	1916.8
	Standard 6	680.8	355.9	375.9	815.1	852.3	424.9	455.3
	Standard 7	1663.2	1449.5	1545.3	3367.4	3518.9	1729.5	1925.3
	Standard 8	653.2	364.1	392.6	830.7	867.7	436.4	484.4
	Standard 9	681.6	358.9	376.4	826.0	846.6	435.1	473.7
	Standard 10	1202.2	740.5	792.4	1699.5	1778.4	885.9	988.6

Table 60: FPRs for the standard mixtures, ranging from 650-2031%. For the combined U(H)PLC-MS assays, the FPRs for house-dust 1 ranged from 815-3377%, whilst for house-dust 2 they ranged from 852-3535%. The FPRs for the combined assays for wristband 1 ranged from 425-1747%, whilst for wristband 2 they ranged from 455-1944%. Finally, FPRs for serum 1 and 2 were 356-1458% and 376-1570%, respectively.

TPRs for the serum samples are shown in Table 59. The maximum TPR was for standard mixture 5, thus it was concluded that serum 2 had been fortified with standard mixture 5. The U(H)PLC-MS workflows applied could successfully detect and annotate xenobiotics. The U(H)PLC-MS assays performed better than the (nano)ESI-DIMS methods, whose TPRs for serum 2 ranged from 16-43%. The highest TPR for the (nano)ESI-DIMS was 43%, whilst the U(H)PLC-MS assays yielded an excellent 75% TPR.

FPRs for serum 1 and 2 are shown in Table 60. Despite the high FPRs, the U(H)PLC-MS methods were still able to correctly determine which sample had been fortified and with which standard mixture. The FPRs for the fortified serum 2 for the (nano)ESI-DIMS methods ranged from 238-984%. These were lower than those observed in the U(H)PLC-MS assays. In this complex sample matrix, the lower FPRs observed for the (nano)ESI-DIMS methods are testament to the stringent data processing applied to these data and highlight an important factor. Data processing parameters can impact results drastically, and data processing parameters such as mass error tolerance must be carefully considered for each dataset regardless of the inherent advantages of the analytical technique. Most NTA apply software default settings. It is recommended that each dataset or sample type's processing parameters are optimised (409–411).

TPRs house-dust 1 and 2 are shown in Table 59. Looking at TPR data alone did not provide convincing evidence of which of the two house-dust samples had been fortified. In the absence of further evidence, the sample with the highest TPRs was concluded to have been fortified (house-dust 2 fortified with either standard mixture 5 or 9). The combined U(H)PLC-MS methods were able to successfully suggest the correct standard mixture as one of two candidates that had been spiked into the house-dust. The reason for the lack of differentiation between these 2 standard mixtures was because all 95 chemicals found in standard mixture 5 were also found in standard mixture 9, which had an additional 270 other compounds for a total of 365. It can therefore be concluded that the U(H)PLC-MS assays could successfully be applied for the analysis of xenobiotics in a complex house-dust matrix using a large reference list of 4,462 chemicals and without the use of RT information. The TPRs for the fortified house-dust 2 using the (nano)ESI-DIMS assays ranged from 23-38% (section 3.2.4, Table 33). The U(H)PLC-MS assays yielded much higher TPRs compared to the (nano)ESI-DIMS assays, despite the stringent processing parameters applied to the latter. This highlights that very

complex sample matrices such as house-dust, which predominantly contains fungal and bacterial organic matter, and human skin cells (412) will pose challenges even when maximally strict processing parameters are applied. Pre-separation of analytes prior to detection by MS affords huge advantages in such a scenario, and it is recommended that analysis of xenobiotics in a house-dust matrix should be carried out using a hyphenated technique.

FPRs for house-dust 1 and 2 are shown in Table 60. Once again, despite this very high FPR, TPRs were exceptionally high, even without the use of RT data during annotation. The FPRs for house-dust 2 using the (nano)ESI-DIMS assays were also once again lower than the U(H)PLC-MS assays, ranging from 409-1694%. This highlighted the advantages of strict data processing parameters, but also revealed a limitation. As matrix complexity increases, even with strict processing parameters and a reduced false positive annotation rate true positive annotation is still low.

TPRs for wristband 1and 2 are shown in Table 59. The highest TPR for wristband 2 was for standard mixture 4, therefore it was concluded that standard mixture 4 had been spiked into wristband 2. The U(H)PLC-MS assays applied can successfully be used for analysis of xenobiotics in a silicone wristband extract matrix using a large reference list of 4,462 chemicals and without RT information in the database. Results here highlight an interesting phenomenon. Individual assay's TPRs did not all successfully reveal the correct standard mixture spiked in but combining the four assays did. This shows the importance of using as many different stationary phases and ion-modes when feasible, or indeed as many analytical approaches as possible to maximise chemical coverage (413–415).

FPRs wristband 1 and 2 are shown in Table 60. Despite these high FPRs, a correct determination of which standard mixture had been spiked into wristband 2 could still be made.

4.2.3 PHASE II (Unblinded): Method Performance

4.2.3.1 Standard Mixtures

Revelation of sample compositions at Phase II experiments facilitated use of this knowledge to either apply modified analytical and/or computational approaches to the same samples and re-evaluate method performance.

The datasets collected at Phase I experiments were re-annotated against small, custom-built RT databases containing only the chemicals revealed to be in each standard mixture. Therefore, four databases were created for each assay (described in 2.3.10). Any compounds that did not have at least one adduct appearing in 3 out of 3 MS1 injections were removed from the RT knowledgebases, thus some chemicals were missing from each knowledgebase depending on whether they were detected reproducibly within each assay. The coverage of each RT database is demonstrated in Table 61.

Since the database coverage of the assays was low, particularly for the HILIC assays, the overlap in the chemicals detected in each assay was investigated and results shown in Figure 27.

	Unique ENTACT Chemicals in RT Database	Total Number of ENTACT Chemicals in 10 Stds	Database Coverage (%)
AqRP_Pos	1116	1940	57.5
AqRP_Neg	722	1940	37.2
HILIC_POS	898	1940	46.3
HILIC_NEG	611	1940	31.5

Table 61: The number of correct unique chemical IDs (DTXSID IDs) where RT information was added to a database for each assay is shown. A percentage of chemical coverage is calculated in relation to the total number of chemicals revealed to be in all ten standard mixtures. Chemical overlap between standard mixtures was not removed for these calculations.

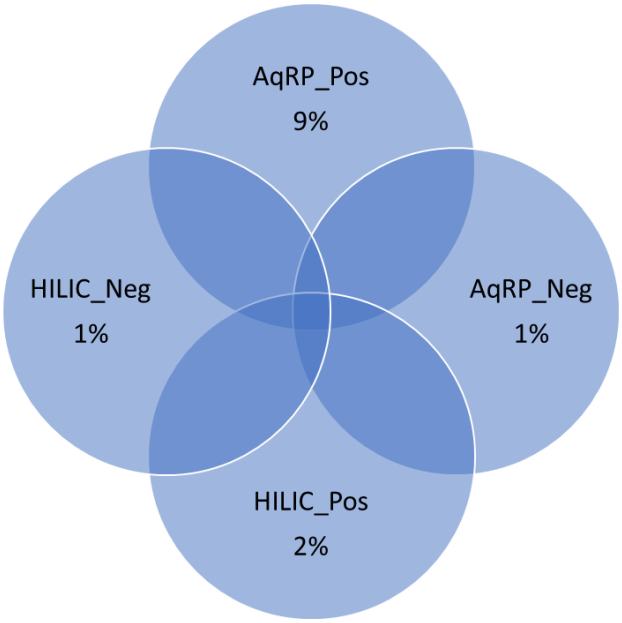


Figure 27: Venn diagram showing the overlap in detected ToxCast chemicals in ten standard mixtures. A combined total of 1,235 unique chemicals out of the 1,940 chemicals revealed to be in the standard mixtures were detected by the four assays and used for creating the retention time databases. Of these 1,235 chemicals, 29% were able to be detected by all four assays. The aqRP C_{18} positive ion-mode assay exclusively detected 9%, whilst the aqRP C_{18} negative ion-mode assay exclusively detected only 1%. The HILIC positive ion-mode assay exclusively detected 2% of the 1,235 chemicals in the RT databases, whilst the HILIC negative ion-mode dataset exclusively detected 1%. This demonstrates how complementary these two methods are, and the importance of applying each assay to detect analytes specific to each one.

The percentage of chemicals exclusively detected in the HILIC assays was much lower than expected. This was investigated by mapping the relationship between RT and *m/z*. It was observed that all chemicals in the mass range 100-600 for the HILIC positive ion-mode assay and 100-450 for the HILIC negative ion-mode dataset eluted with a RT <2 minutes. As such, these chemicals were in the void volume, where no interaction between the analyte and SP occurred and none of these chemicals benefitted from chromatographic separation, meaning they also were subject to ion suppression having all co-eluted early in the chromatographic run (Figure 28). By contrast, for both the aqueous C_{18} reversed phase assays, a gradual increase in RT was observed with increasing *m/z,* meaning analytes enjoyed the benefit of chromatographic separation (Figure 29).

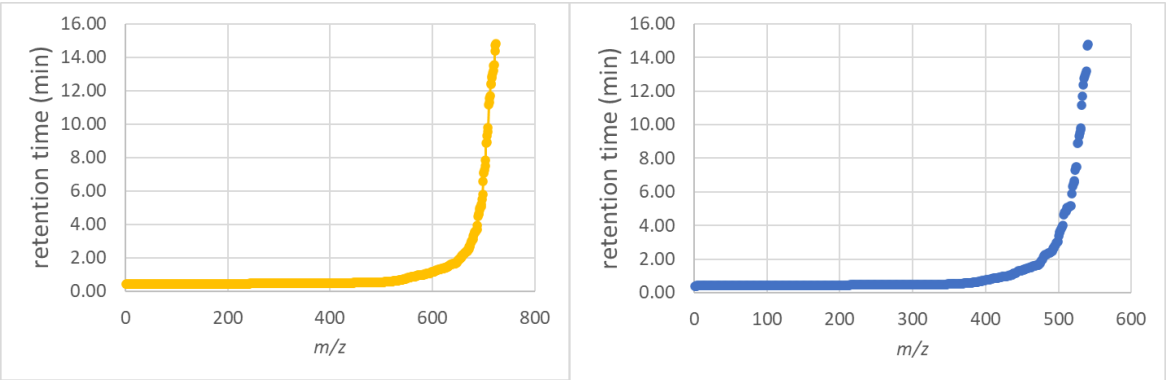


Figure 28: A mapping of the relation between *m/z* and RT for the HILIC positive and negative ion-mode assays revealed that chemicals in the mass ranges of 100-600 and 100-450, respectively, eluted by 2 min. As such, these chemicals were in the void volume and thus coeluted together and therefore were subjected to ion suppression.

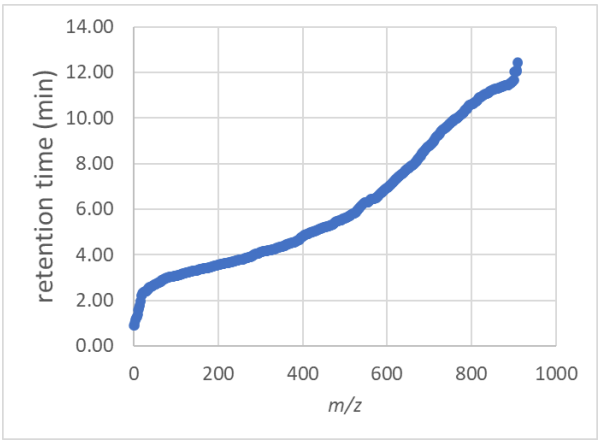


Figure 29: A mapping of the relationship between m/z and RT for both aqueous C_{18} reversed phase assays revealed a steady increase in RT with an increase in m/z, thus the majority of chemicals experienced some retention.

TPRs for the standard mixtures are shown in Table 62. These were much lower than those observed at Phase I experiments, which ranged from 0-81%, 0-47%, 0-41%, and 0-35% for the aqRP C_{18} positive and negative ion-mode assays, and the HILIC positive and negative ion-mode assays, respectively.

		AqRP Positive	AqRP Negative	HILIC Positive	HILIC Negative
TRUE POSITIVES (%)	Standard 1	16.8	7.4	13.7	12.6
	Standard 2	46.3	22.1	33.7	28.4
	Standard 3	48.4	6.3	27.4	6.3
	Standard 4	19.5	10.3	16.8	11.9
	Standard 5	53.7	8.4	38.9	30.5
	Standard 6	0.3	0.0	0.3	0.5
	Standard 7	0.0	0.0	0.0	0.0
	Standard 8	51.0	4.9	32.3	19.7
	Standard 9	38.1	22.5	21.1	12.9
	Standard 10	19.7	4.7	18.9	12.6

Table 62: TPRs for the aqueous C_{18} reversed phase positive and negative ion-modes, and the HILIC positive and negative ion-modes, ranging from 0-54%, 0-23%, 0-39% and 0-31%, respectively.

This is due to chemicals whose RT information was not recorded and were thus missing from the small databases. The use of RT information during database matching increases confidence that a correct annotation has been reported as two complementary types of data have been applied, particularly where a RT knowledgebase has been created using authentic chemical standards (416). However, the lower TPRs observed because of the missing information emphatically demonstrate one of the major challenges in creating and implementing such knowledgebases. They are often incomplete, reducing the number of annotations made, and ultimately, useful information can be lost (417). Care must therefore be taken in how results are interpreted when such a strategy is employed, and a two-stage process should be applied where one stage performs annotations using MS1 and RT data and another stage uses MS1 data only.

FPRs for the standard mixtures are shown in Table 63. These were significantly lower than those observed for the Phase I experiments, which ranged from 811-2081%, 236-1355%, 166-517%, and 161-691% for the aqRP C_{18} positive and negative ion-mode assays, and the HILIC positive and negative ion-mode assays, respectively. The benefits of using small databases containing RT data are illustrated, with FPRs decreasing by factors of 10-20.

		AqRP Positive	AqRP Negative	HILIC Positive	HILIC Negative
FALSE POSITIVES (%)	Standard 1	67.4	26.3	78.9	70.5
	Standard 2	95.8	33.7	98.9	60.0
	Standard 3	137.9	15.8	92.6	67.4
	Standard 4	193.5	64.9	93.0	84.3
	Standard 5	235.8	17.9	96.8	97.9
	Standard 6	178.6	29.9	75.9	63.8
	Standard 7	177.9	18.9	116.8	83.2
	Standard 8	69.0	1.6	26.3	14.5
	Standard 9	163.6	51.5	52.6	35.1
	Standard 10	42.2	7.9	25.2	20.8

Table 63: FPRs for the aqueous C_{18} reversed phase positive and negative ion-modes, and the HILIC positive and negative ion-modes, ranging from 42-236%, 2-65%, 25-117%, and 15-98%, respectively.

The U(H)PLC-MS assays used here can therefore be applied successfully for the analysis of xenobiotics in a clean solvent matrix using RT databases containing only the chemicals known to be in the standard mixtures. However, creation of smaller RT databases has its caveats. TP annotation is possibly reduced by the inherent incompleteness of these databases, limited by the unavailability or cost of authentic chemical standards, and further by the not fully characterised compositions of the exposome. Antagonistically, FPRs decrease sharply with the use of these targeted databases, improving confidence in the annotations that are made, albeit at a lower TPR, and importantly dramatically filtering away the number of false annotation candidates presented for each peak.

4.2.3.2 Fortified and Unfortified Serum

TPRs for serum 1 are shown in Table 64. These low TPRs immediately and correctly suggested that serum 1 had not been fortified. Moreover, they were lower than those observed in the Phase I experiments, which ranged from 0-16%, 0-18%, 0-8%, and 0-11%, respectively. This shows the value of using RT information during database matching. The reduction in these apparent TPs is a result of any incorrect and chromatographically resolved isomers not being matched to the database because of incorrect RTs. However, this lower TPR also reflects the chemicals that are missed out during the creation of the targeted databases. A smaller search space yields less hits, regardless of whether those annotations are correct or not.

		AqRP Positive	AqRP Negative	HILIC Positive	HILIC Negative
TRUE POSITIVES (%)	Standard 1	0.0	1.1	0.0	1.1
	Standard 2	2.1	0.0	3.2	1.1
	Standard 3	1.1	1.1	1.1	1.1
	Standard 4	0.0	0.5	0.5	1.6
	Standard 5	2.1	4.2	1.1	4.2
	Standard 6	0.0	0.0	0.0	0.0
	Standard 7	0.0	0.0	0.0	0.0
	Standard 8	0.8	0.3	0.5	0.5
	Standard 9	2.2	0.8	0.0	0.8
	Standard 10	0.3	1.1	0.5	0.5

Table 64: TPRs for serum 1, ranging from 0-2.2%, 0-4.2%, 0-3.2% and 0-4.2% for the aqueous C_{18} reversed phase positive and negative ion-modes, and the HILIC positive and negative ion-modes, respectively.

TPRs for serum 2 are shown in Table 65. These were higher rates than for serum 1, and the highest TPR was for standard mixture 5. It was therefore correctly concluded that serum 2 had been fortified with standard mixture 5. The TPRs for serum 2 at Phase II of ENTACT experiments were lower than those at Phase I, which ranged from 0-52%, 0-25%, 0-24% and 0-31% for the aqueous C_{18} reversed phase positive and negative ion-mode assays, and the HILIC positive and negative ion-mode assays, respectively. This is once again an observed and expected impact of creating RT databases, which will undoubtedly be less complete than reference lists or databases without RT data.

		AqRP Positive	AqRP Negative	HILIC Positive	HILIC Negative
TRUE POSITIVES (%)	Standard 1	6.3	1.1	2.1	2.1
	Standard 2	4.2	0.0	3.2	3.2
	Standard 3	1.1	1.1	1.1	1.1
	Standard 4	0.0	0.5	1.6	1.6
	Standard 5	34.7	15.8	22.1	23.2
	Standard 6	0.0	0.0	0.0	0.5
	Standard 7	0.0	0.0	0.0	0.0
	Standard 8	2.5	0.5	1.9	1.4
	Standard 9	2.2	1.1	0.3	1.4
	Standard 10	1.1	1.4	1.4	0.5

Table 65: TPRs for serum 2, ranging from 0-35%, 0-16%, 0-22%, and 0-23% for the aqueous C_{18} reversed phase positive and negative ion-modes, and the HILIC positive and negative ion-modes, respectively.

FPRs for serum 1 and 2 are shown in Table 66 and Table 67. These FPRs were lower than those observed at Phase I experiments. For serum 1, these ranged from 185-744%, 196-778%, 117-462%, and 149-588%, and for serum 2, FPRs ranged from 233-940%, 197-783%, 136-545%, and 207-818%, for the aqueous C_{18} reversed phase positive and negative ion-modes, and the HILIC positive and negative ion-modes, respectively. As with the standard mixtures, a sharp decrease in the FPR is observed, illustrating once again the benefits of using smaller RT databases.

		AqRP Positive	AqRP Negative	HILIC Positive	HILIC Negative
FALSE POSITIVES (%)	Standard 1	21.1	21.1	14.7	21.1
	Standard 2	18.9	22.1	11.6	21.1
	Standard 3	20.0	21.1	13.7	21.1
	Standard 4	10.8	10.8	7.0	9.7
	Standard 5	18.9	17.9	13.7	17.9
	Standard 6	5.5	5.8	3.8	5.8
	Standard 7	21.1	22.1	14.7	22.1
	Standard 8	4.7	5.5	3.3	5.2
	Standard 9	3.3	4.9	3.8	4.9
	Standard 10	5.2	4.7	3.3	5.2

Table 66: FPRs for serum 1, ranging from 3-21%, 5-22%, 3-15%, and 5-22% for the aqueous C_{18} reversed phase positive and negative ion-modes, and the HILIC positive and negative ion-modes, respectively.

		AqRP Positive	AqRP Negative	HILIC Positive	HILIC Negative
FALSE POSITIVES (%)	Standard 1	115.8	42.1	71.6	56.8
	Standard 2	117.9	43.2	70.5	55.8
	Standard 3	121.1	42.1	72.6	57.9
	Standard 4	62.7	21.6	36.2	28.6
	Standard 5	87.4	27.4	51.6	35.8
	Standard 6	31.8	11.2	19.2	14.8
	Standard 7	122.1	43.2	73.7	58.9
	Standard 8	29.3	10.7	17.3	14.0
	Standard 9	29.6	10.1	18.9	14.0
	Standard 10	30.7	9.9	17.8	14.8

Table 67: FPRs for serum 2, ranging from 31-122%, 10-43%, 17-74%, and 14-59% for the aqueous C_{18} reversed phase positive and negative ion-modes, and the HILIC positive and negative ion-modes, respectively.

4.2.3.3 Fortified and Unfortified House-dust

TPRs for serum 1 are shown in Table 68. These low TPRs correctly suggested that house-dust 1 had not been fortified. Indeed, these were much lower than the TPRs for the Phase I experiments, which ranged from 0-56%, 0-55%, 0-26%, and 0-30%, respectively. This demonstrated the annotation filtering advantages of using RT databases.

		AqRP Positive	AqRP Negative	HILIC Positive	HILIC Negative
TRUE POSITIVES (%)	Standard 1	4.2	4.2	3.2	7.4
	Standard 2	5.3	5.3	6.3	10.5
	Standard 3	9.5	5.3	5.3	4.2
	Standard 4	3.8	2.7	3.2	5.9
	Standard 5	10.5	9.5	5.3	9.5
	Standard 6	0.0	0.0	0.0	0.5
	Standard 7	0.0	0.0	0.0	0.0
	Standard 8	5.2	2.7	4.7	3.3
	Standard 9	5.8	3.8	3.0	3.3
	Standard 10	5.2	1.9	2.7	1.6

Table 68: For house-dust 1, TPRs ranged from 0-11%, 0-10%, 0-6%, and 0-11% for the aqueous C_{18} reversed phase positive and negative ion-modes, and the HILIC positive and negative ion-modes, respectively.

For house-dust 2, TPRs for house-dust 2 are shown in Table 69. These higher TPRs suggested that house-dust 2 had been fortified with standard mixture 5, for which the highest TPRs were obtained. The determination that house-dust 2 had been fortified was correct, but this processing approach failed to correctly determine that standard mixture 9 had been spiked into the sample. This can be explained by the overlap between standard mixtures 5 and 9, with 95 chemicals found in standard mixture 5 also being present in standard mixture 9, which had an additional 270 other compounds for a total of 365 compounds. Consequently, the same chemicals detected in standard mixture 5 were also detected in standard mixture 9, but the lower TPR of standard mixture 9 was due to the higher total number of chemicals within the mixture.

		AqRP Positive	AqRP Negative	HILIC Positive	HILIC Negative
TRUE POSITIVES (%)	Standard 1	10.5	6.3	5.3	8.4
	Standard 2	7.4	7.4	5.3	12.6
	Standard 3	11.6	6.3	6.3	4.2
	Standard 4	3.8	3.2	3.2	6.5
	Standard 5	48.4	28.4	22.1	13.7
	Standard 6	0.0	0.0	0.0	0.5
	Standard 7	0.0	0.0	0.0	0.0
	Standard 8	8.2	4.1	6.8	3.6
	Standard 9	33.2	15.9	14.2	7.7
	Standard 10	6.6	3.3	3.8	3.0

Table 69: For house-dust 2, TPRs ranged from 0-48%, 0-28%, 0-22%, 0-14% for the aqueous C_{18} reversed phase positive and negative ion-modes, and the HILIC positive and negative ion-modes, respectively.

For house-dust 1, FPRs for house-dust 1 are shown in Table 70. These were much lower than those observed in the Phase I experiments, which ranged from 1453-5774%, 1467-5784%, 577-2295%, and 693-2738%, respectively.

		AqRP Positive	AqRP Negative	HILIC Positive	HILIC Negative
FALSE POSITIVES (%)	Standard 1	162.1	101.1	94.7	123.2
	Standard 2	161.1	100.0	91.6	120.0
	Standard 3	156.8	100.0	92.6	126.3
	Standard 4	81.6	51.4	47.0	61.1
	Standard 5	155.8	95.8	92.6	121.1
	Standard 6	43.3	27.4	25.5	33.4
	Standard 7	166.3	105.3	97.9	130.5
	Standard 8	38.1	24.7	20.8	30.7
	Standard 9	37.5	23.6	22.5	30.7
	Standard 10	38.1	25.5	22.7	32.3

Table 70: For house-dust 1, FPRs ranged from 38-166%, 24-105%, 21-98%, 31-131% for the aqueous C_{18} reversed phase positive and negative ion-modes, and the HILIC positive and negative ion-modes, respectively.

FPRs for house-dust 2 are shown in Table 71. These were also lower than those observed in the Phase I experiments, which ranged from 1701-6766%, 1653-6518%, 534-2150%, and 748-

2961%, respectively. For both house-dust samples, the reduction in FP annotation rates were by factors of ~10 when small RT databases are used for annotation.

		AqRP Positive	AqRP Negative	HILIC Positive	HILIC Negative
FALSE POSITIVES (%)	Standard 1	507.4	214.7	176.8	158.9
	Standard 2	510.5	213.7	176.8	154.7
	Standard 3	506.3	214.7	175.8	163.2
	Standard 4	262.2	110.3	90.3	79.5
	Standard 5	469.5	192.6	160.0	153.7
	Standard 6	134.8	57.5	47.4	43.0
	Standard 7	517.9	221.1	182.1	167.4
	Standard 8	126.6	53.4	40.5	40.0
	Standard 9	101.6	41.6	33.2	35.9
	Standard 10	128.2	54.2	43.6	40.5

Table 71: For house-dust 2, FPRs ranged from 102-518%, 42-221%, 33-182%, 36-167% for the aqueous C_{18} reversed phase positive and negative ion-modes, and the HILIC positive and negative ion-modes, respectively.

4.2.3.4 Fortified and Unfortified Wristband

For wristband 1, TPRs for wristband 1 are shown in Table 72. These very low TPRs immediately suggested that wristband 1 had not been fortified and were much lower than those observed in Phase I of the study, which ranged from 0-17%, 0-14%, 0-12% and 0-18%, respectively. This once again demonstrates the impact of RT databases during annotation. However, it is important to note that the reduced number of annotations also reflects on the missing chemicals in each of the four databases.

		AqRP Positive	AqRP Negative	HILIC Positive	HILIC Negative
TRUE POSITIVES (%)	Standard 1	4.2	0.0	3.2	6.3
	Standard 2	2.1	0.0	5.3	5.3
	Standard 3	2.1	0.0	2.1	5.3
	Standard 4	1.1	0.5	2.2	4.3
	Standard 5	3.2	3.2	5.3	6.3
	Standard 6	0.0	0.0	0.0	0.5
	Standard 7	0.0	0.0	0.0	0.0
	Standard 8	2.7	1.1	3.6	1.4
	Standard 9	1.6	1.1	1.4	2.2
	Standard 10	1.1	0.8	1.6	1.6

Table 72: For wristband 1, TPRs ranged from 0-4%, 0-3%, 0-5%, and 0-6% for the aqueous C_{18} reversed phase positive and negative ion-modes, and the HILIC positive and negative ion-modes, respectively.

TPRs for wristband 2 are shown in Table 73. These were higher than wristband 1 and suggested that wristband 2 had been fortified. The highest TPR was for standard mixture 5, therefore it was concluded that wristband 2 had been fortified with standard mixture 5. This conclusion was incorrect as wristband 2 had in-fact been fortified with standard mixture 4. Standard mixture 5 was found to also be a subset of standard mixture 4, with all 95 chemicals found in standard mixture 5 also found in standard mixture 4, which had an additional 90 other compounds for a total of 185 compounds. This explains the high TPRs for this standard mixture when in-fact standard mixture 4 had been spiked into wristband 2. TPRs for wristband 2 in the Phase II experiments were much lower than those obtained in the Phase I experiments, which ranged from 0-56%, 0-22%, 0-35% and 0-35%, respectively. This reflects on the incompleteness of the small databases as has already been discussed.

		AqRP Positive	AqRP Negative	HILIC Positive	HILIC Negative
TRUE POSITIVES (%)	Standard 1	8.4	3.2	7.4	8.4
	Standard 2	4.2	1.1	12.6	10.5
	Standard 3	11.6	1.1	7.4	7.4
	Standard 4	18.4	2.7	13.5	8.6
	Standard 5	34.7	13.7	26.3	23.2
	Standard 6	0.0	0.0	0.3	0.5
	Standard 7	0.0	0.0	0.0	0.0
	Standard 8	3.8	1.6	6.0	2.5
	Standard 9	1.6	1.1	2.5	3.0
	Standard 10	1.4	0.3	2.7	1.9

Table 73: For wristband 2, TPRs ranged from 0-35%, 0-14%, 0-26%, and 0-23% for the aqueous C_{18} reversed phase positive and negative ion-modes, and the HILIC positive and negative ion-modes, respectively.

For wristband 1, FPRs for wristband 1 and 2 are shown in Table 74 and Table 75, showing a similar reduction in FPR as the serum and house-dust samples.

		AqRP Positive	AqRP Negative	HILIC Positive	HILIC Negative
FALSE POSITIVES (%)	Standard 1	37.9	20.0	62.1	77.9
	Standard 2	40.0	20.0	60.0	78.9
	Standard 3	40.0	20.0	63.2	78.9
	Standard 4	20.5	9.7	31.4	38.9
	Standard 5	38.9	16.8	60.0	77.9
	Standard 6	11.0	5.2	17.0	21.4
	Standard 7	42.1	20.0	65.3	84.2
	Standard 8	8.2	4.1	13.4	20.5
	Standard 9	9.3	4.1	15.6	19.7
	Standard 10	9.9	4.4	15.3	20.3

Table 74: For wristband 1, FPRs ranged from 8-42%, 4-20%, 13-65%, and 20-79% for the aqueous C_{18} reversed phase positive and negative ion-modes, and the HILIC positive and negative ion-modes, respectively.

		AqRP Positive	AqRP Negative	HILIC Positive	HILIC Negative
FALSE POSITIVES (%)	Standard 1	242.1	46.3	198.9	147.4
	Standard 2	246.3	48.4	193.7	145.3
	Standard 3	238.9	48.4	198.9	148.4
	Standard 4	110.3	22.7	92.4	71.4
	Standard 5	215.8	35.8	180.0	132.6
	Standard 6	65.2	12.9	53.4	40.0
	Standard 7	250.5	49.5	206.3	155.8
	Standard 8	61.4	11.2	47.7	38.1
	Standard 9	63.6	11.8	51.2	37.5
	Standard 10	63.8	12.6	51.0	38.6

Table 75: For wristband 2, FPRs ranged from 61-251%, 11-50%, 48-206%, 39-156% for the aqueous C_{18} reversed phase positive and negative ion-modes, and the HILIC positive and negative ion-modes, respectively.

FPRs for Phase II experiments were significantly lower than those observed for Phase I experiments, where a large reference list of 4,462 chemicals without RT data were screened and results here demonstrated the extent to which FPR can be reduced.

4.2.4 Phase II: Summary

All four assay annotations were combined for each sample type, and TPR and FPR are shown in Table 76 and Table 77, respectively.

	Standards	Serum 1	Serum 2	House-dust 1	House-dust 2	Wristband 1	Wristband 2
Standard 1	31.6	3.2	11.6	17.9	25.3	12.6	22.1
Standard 2	70.5	8.4	12.6	26.3	30.5	13.7	26.3
Standard 3	60.0	5.3	6.3	23.2	25.3	9.5	24.2
Standard 4	66.5	7.6	31.9	25.4	46.5	13.5	57.8
Standard 5	65.3	10.5	55.8	25.3	66.3	13.7	55.8
Standard 6	1.1	1.1	1.1	0.5	1.1	0.5	1.1
Standard 7	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Standard 8	66.8	2.7	5.2	12.9	17.3	7.4	11.2
Standard 9	71.0	6.0	18.4	17.0	62.5	9.0	21.4
Standard 10	68.1	5.4	9.2	19.5	24.3	10.3	13.0

Table 76: Combined U(H)PLC-MS assays yielded TPRs of 0-74%, 0-11%, 0-56%, 0-30%, 0-66%, 7-21%, 0-58% for the standard mixtures, serum, house-dust, and wrsitband samples.

		Standards	Serum 1	Serum 2	House-dust 1	House-dust 2	Wristband 1	Wristband 2
FALSE POSITIVES (%)	Standard 1	121.1	51.6	124.2	212.6	424.2	115.8	249.5
	Standard 2	123.2	46.3	123.2	204.2	418.9	114.7	245.3
	Standard 3	129.5	49.5	129.5	207.4	424.2	118.9	247.4
	Standard 4	71.4	20.5	37.8	93.0	184.3	52.4	81.6
	Standard 5	102.1	44.2	80.0	205.3	383.2	114.7	215.8
	Standard 6	23.3	7.7	19.7	29.9	74.2	17.5	32.1
	Standard 7	93.7	50.5	116.8	207.4	396.8	107.4	235.8
	Standard 8	29.6	11.5	30.1	47.1	99.7	26.0	59.5
	Standard 9	41.9	8.2	17.0	43.0	54.5	24.4	49.3
	Standard 10	70.8	22.7	60.5	98.9	206.5	55.7	126.5

Table 77: Combined U(H)PLC-MS assays yielded FPRs of 23-130%, 8-52%, 17-130%, 30-213%, 55-424%, 18-119%, 32-250%

Standard mixture 6 and 7 had very few or no correct annotations (already discussed in section 4.2.2). The combined U(H)PLC-MS assays annotated against the RT databases had lower TPRs than the (nano)ESI-DIMS combined assays, whose TPRs for the standards ranged from 53-83%. Only two standard mixtures (6 and 7) had TPRs below 50%. The other standard mixtures had TPRs greater than 72%. This reflects the completeness of the small databases created during Phase II of the (nano)ESI-DIMS studies, which contained all chemicals present in the standard mixtures. By contrast, databases created for annotation of the U(H)PLC-MS assays were incomplete. This was because the data collected for the standard mixtures in a clean solvent matrix were used to construct RT databases. To accurately record RTs, only compounds for whom a peak was detected in all three MS1 injections were retained, thus resulting in incomplete databases where peaks were not detected in all 3 MS1 replicate injections. Compounds detected in a only one or two injections were not investigated as only reproducibly detected compounds were desired in the databases, although this could have improved TPRs if a higher number of compounds were present in the database. Indeed, this missing information from the small databases is reflected when these U(H)PLC-MS Phase II results are compared to Phase I results, where despite the larger database used for matching the TPRs yielded were higher (0-94%).

FPRs for the standard mixtures were much lower than those observed during the Phase I experiments, which ranged from 650-2031%; the reduction in the FPR is observed at factors >20. The use of small RT databases reduced FP annotation greatly. FPRs for the (nano)ESI-DIMS Phase II experiments ranged from 0-24%. These were lower than those observed for the combined U(H)PLC-MS assays, but the (nano)ESI-DIMS data were more strictly filtered

using mass error < 1ppm, thereby removing some additional false positive annotations through the filtering process.

The low TPRs for serum 1 correctly suggested that this sample had not been fortified. Serum 2, which had higher TPRs, was concluded to have been fortified. The highest TPR was for standard mixture 5, therefore it was correctly concluded that serum 2 had been fortified with standard mixture 5. The combined U(H)PLC-MS assays performed worse than the combined (nano)ESI-DIMS assays during Phase II of the ENTACT experiments, which yielded higher TPRs of 2-62% for the fortified serum 2 sample. Standard mixture 5 had a TPR of 62% in the (nano)ESI-DIMS combined assays, ~6% lower than the U(H)PLC-MS combined assays. As already discussed, the (nano)ESI-DIMS data correction and strict matching criteria may have influenced this. Moreover, the RT databases created for the U(H)PLC-MS assays had missing chemicals as had already been discussed, thus lowering the TPR.

FPRs for serum 1 and 2were singnificantly lower than those observed during the Phase I experiments, which ranged from 356-1458% and 376-1570% for serum 1 and serum 2, respectively, highlighting the main advantage of using smaller RT databases so to reduce the FP annotation rate. The FPR observed in the (nano)ESI-DIMS Phase II analyses of serum yielded FPRs of 1-60% for the fortified serum 2 sample. The FPRs for the U(H)PLC-MS combined assays were ~ twice as high, illustrating the impact of strict data processing parameters applied to the DIMS data.

House-dust 2 had much higher TPRs than house-dust 1, therefore it was concluded that house-dust 2 had been fortified and house-dust 1 had not been fortified. The highest TPR for house-dust 2 suggested that standard mixture 5 had been spiked in. The combined U(H)PLC-MS assays were able to correctly determine that house-dust 2 had been fortified, but incorrectly suggested which standard mixture had been spiked in. However, this can be explained by the overlap between standard mixture 5 and 9, with all 95 chemicals found in standard mixture 5 also being present in standard mixture 9, which had 270 other compounds present for a total of 365. Indeed, the TPR for standard mixture 9 in the house-dust 2 matrix was 63% (Table 76). It can therefore be concluded that xenobiotics can be successfully analysed using the U(H)PLC-MS assays applied here, and matching against small RTdatabases.

For the combined (nano)ESI-DIMS Phase II experiments, house-dust 2 was also incorrectly concluded to have been spiked with standard mixture 5. The TPRs for house-dust 2 using DIMS ranged from 2-54%, with standard mixture 5 having a 54% TPR. This is also excusable due to the overlap between standard mixtures 5 and 9 as described above, and indeed the TPR for standard mixture 9 being spiked into house-dust was 48%, the 2nd highest TPR in the range. The combined U(H)PLC-MS assays performed better than the (nano)ESI-DIMS combined assays, yielding a higher TPR for both standard mixture 5 (66%) and standard mixture 9 (63%), compared to the 54% and 48% for standard mixtures 5 and 9 for the combined (nano)ESI-DIMS assays. This highlights an important result. In the relatively less complex sample matrices of solvent and serum, (nano)ESI-DIMS methods matching against small reference lists containing only the chemicals revealed to be in each standard mixture and were able to outperform U(H)PLC-MS assays by yielding higher TPRs. However, for the analysis of the more complex house-dust samples, the use of small RT databases containing only chemicals revealed to be in each standard mixture yielded better results for the U(H)PLC-MS combined assays, with higher TPRs obtained compared to the (nano)ESI-DIMS assays.

FPRs for house-dust 1 and 2 during the Phase II experiments were constrained by the use of small RT databases during annotation, compared to FPRs during Phase I (815-3377% and 852-3535% for house-dust 1 and house-dust 2, respectively). The lower FPRs observed in the Phase II U(H)PLC-MS assays were, however, still larger than those observed for the (nano)ESI-DIMS experiments, which ranged from 2-186% for the fortified house-dust 2., demonstrating the impact of strict data processing and how effective this can be at filtering out non-biological data. The U(H)PLC-MS assays applied for these analyses therefore could successfully be used for the analysis of xenobiotics in a complex house-dust matrix when annotation is carried out using small RT databases. The advantages of using these RT databases are demonstrated best in the house-dust matrix, reducing false positive annotation by factors of >4 and up to ~10.

The higher TPRs observed for wristband 2 correctly suggested that it had been fortified. The highest TPR was for fortification with standard mixture 4, and it was correctly concluded that wristband 2 had been spiked with standard mixture 4. The TPRs observed during Phase II expeirments, using small retention databases for annotation, were lower than those observed during the Phase I experiments, which ranged from 0-38% and 0-77% for wristband 1 and

wristband 2, respectively. This was due to the missing chemicals for which RT information was not available during database creation.

FPRs for the wristbands were significantly lower than those observed during the Phase I experiments, which ranged from 425-1747% and 455-1944% for wristband 1 and wristband 2, respectively. This demonstrated the impact of using RT databases on FP annotations, reducing FPRs by factors of ~10-20. U(H)PLC-MS assays could therefore successfully be applied for the analysis of xenobiotics in a silicone wristband extract matrix, correctly determining which sample had been fortified and with which standard mixture.

4.3 Conclusions

The University of Birmingham's metabolomics group participated in a global ring trial created by the US-EPA called ENTACT, which has been extensively described in sections 3.1 and 3.3. Work presented in this chapter is identical to work in chapter 3, except that U(H)PLC-MS assays were applied instead. The objectives of this ring trial were to evaluate the performance of commonly applied NTA for the analysis of xenobiotics in sample matrices of varying complexities. Four U(H)PLC-MS assays (aqueous C_{18} reversed phase and HILIC, positive and negative ion-mode assays) were applied for the analysis of ten standard mixtures, and fortified and unfortified serum, house-dust, and wristband extracts. Method performance for Phase I experiments involved calculating the TPRs and FPRs yielded when data were annotated against a large reference list with 4,462 chemicals, but without the use of RT data. Additionally, the fortified and unfortified samples were also to be annotated against this reference list without using RT matching, and method effectiveness was evaluated by whether the fortified sample and the standard mixture it had been fortified with could be correctly identified. Method performance at Phase II involved the creation of small RT databases from the standards analysed during Phase I. These databases contained only the chemicals known to be present in the standard mixtures and were detected by each assay (thus four databases were created for each of four assays). Moreover, for a RT to be retained in the database, at least one ion form of the compound required detection in all three MS1 injections of the standard mixture. Annotation of the ten standard mixtures and the fortified and unfortified serum, house-dust, and wristband extracts was then carried out by matching against these RT databases.

Phase I results demonstrated that the four U(H)PLC-MS methods can be applied for the successful analysis of xenobiotics in increasingly complex sample matrices, yielding the highest TPRs for the standard mixtures in a clean solvent matrix, and correctly identifying which of the two samples of the same type had been fortified. These methods correctly identified which standard mixture had been spiked into the serum and wristband matrices but were unable to correctly identify which of two standard mixtures had been spiked into the house-dust matrix. However, the ambiguity of which standard mixture had been spiked into the house-dust matrix was due to the overlap in the chemicals present in the two standard mixtures, 5 and 9, with the former being a full subset of the latter. This experimental design was crucial as it allowed accurate evaluation of method specificity. Therefore, the analysis of xenobiotics in a house-dust matrix was successful, yielding a TPR of >90%. For all standard mixtures and the fortified samples, FPRs at Phase I experiments were high, but this did not impact the high TPRs observed. The U(H)PLC-MS results at Phase I were better than those for the (nano)ESI-DIMS results at Phase I, which had lower TPRs for the standard mixtures and the fortified samples. However, these methods were still able to correctly determine which samples had been fortified, and with which standard mixture. This shows the utility of high throughput screening methods such as (nano)ESI-DIMS, which can be used in combination with other analytical techniques to screen through very large reference lists such as KEGG (418) and ChemSpider (419) to shortlist suspect chemicals. These suspect lists can then be used for annotation of data collected using other techniques that do not suffer the same disadvantages as (nano)ESI-DIMS such as ion suppression, and the initial use of (nano)ESI-DIMS serves to provide a reduced search space.

During annotation of U(H)PLC-MS Phase I data, RT matching was not carried out. Data were therefore annotated based on MS1 *m/z* data alone. A limited comparison could therefore be carried out between the U(H)PLC-MS and (nano)ESI-DIMS Phase I results, both matching against the same reference list of 4,462 chemicals. This comparison was limited by the different MS instruments and settings used to collect the data. The (nano)ESI-DIMS methods were collected on Thermo Fisher Scientific's LTQ-Orbitrap Elite mass spectrometer (394), whilst the U(H)PLC-MS data were collected on Thermo Fisher Scientific's Q-Exactive Plus mass spectrometer (420). This was because the ring trial aimed to evaluate existing metabolomics methods applied globally, and these were the instrument configurations in the UoB

metabolomics groups. During data collection, different mass resolutions were applied. The (nano)ESI-DIMS data were collected at a mass resolution of 240,000 (FWHM at *m/z* 400) whilst the U(H)PLC-MS data were collected at a mass resolution of 70,000 (FWHM at *m/z 200*). Lower mass resolution during data collection can impair the detection of isotopes, which have been shown to aid reduction in false positive annotation by >95% (403). However, the most significant differences likely occurred at the data processing stage. Due to the lack of separation, the (nano)ESI-DIMS data underwent extensive data filtering prior to annotation and were crudely mass corrected. Database matching for these data therefore used a very narrow matching tolerance of just 1ppm, whilst the U(H)PLC-MS data used a database matching tolerance of 5ppm and did not undergo any post-acquisition mass correction. It is therefore inappropriate to extensively compare these two types of analysis. However, some observations can still be made, having clearly stated the limitations of comparing the data in this way. The U(H)PLC-MS TPRs for the standard mixtures and the fortified samples were higher than the (nano)ESI-DIMS TPRs for the same samples, even though the U(H)PLC-MS results had much higher FPRs. This demonstrates a characteristic of U(H)PLC-MS data called anisotropy (292), which reduces data complexity of each analyte and in turn can increase TPRs of annotation. The lower FPRs observed in the (nano)ESI-DIMS methods demonstrated a very important consideration. Despite the inherent pros or cons of an analytical technique, data processing parameters can greatly influence the results. The (nano)ESI-DIMS data underwent a crude mass correction, and database matching was carried out using the strictest possible matching criteria (mass error of <1ppm). This yielded noticeably lower FPRs compared to the U(H)PLC-MS assays, for which similar corrections and strict database matching criteria were not applied. Application of stricter parameters could have improved this result, but methods were applied as they were in routine metabolomics experiments as per the ENTACT experimental design.

It is interesting to note that the (nano)ESI-DIMS methods yielded a TPR of ~48% for standard mixtures 6 and 7, where the U(H)PLC-MS methods had TPRs of ~1 and 0% for standard mixtures 6 and 7, respectively. This reinforces the value of the (nano)ESI-DIMS high-throughput screening approach, demonstrating that in combination with other analytical methods, it serves as a vital tool in screening through large reference lists and providing putative annotations, approximately half of which are correct. The major challenge lies in

filtering these results and determining which are correct and which are not. This challenge is compounded by complex matrices, increasing the number of annotations requiring filtering. Finally, standard mixtures 6 and 7 bring forward a growing concern as they reflect the reality of environmental samples, which often do contain isomers and isobars and can be of low complexity. Although sample clean up methods are applied and recommended in some research groups (421), the requirement for specificity impedes the non-targeted nature of exposomics. Here, more evaluations are required to fully characterise the coverage and recovery rates of clean up methods such as solid-phase extraction (see (422).

During annotation of Phase II U(H)PLC-MS data, annotation was carried out by matching against small RT databases. These databases had missing chemicals depending on which ion forms of the standard mixture constituents were detected in all three of the MS1 injections for that standard mixture, but this was done to ensure only reproducibly detected chemicals were added to the database. Phase II results demonstrated that the four U(H)PLC-MS assays could successfully be applied for the successful analysis of xenobiotics in increasingly complex sample matrices. However, relative to Phase I results, TPRs were reduced for all sample types because of the incomplete RT databases. This highlights a common issue in the creation of RT databases for NTA. Since RT matching requires that authentic chemical standards be analysed on the same analytical system and applying the same assay as the sample will be analysed, purchasing standards for all suspected chemicals is a challenge, with standards often unavailable or are costly to purchase all chemical standards. Sharing of RT databases or creation of open-source RT databases is also challenged by the varying RTs which analytes are observed at due to system set-up and assay, although research is emerging that proposes creation of retention indexes as described here (423).These standardise retention times across different instruments, allowing more universal databases to be constructed. The real advantage of using RT databases is observed in the huge reduction in FP annotations of the data during the Phase II experiments. This drastic fall in FP annotation can make data interpretation much simpler and more accurate, whilst also facilitating easier validation of results. For example, purchasing ten chemical standards to confirm the identity of a single peak of interest is much more feasible than purchasing fifty chemical standards.

Results derived from the Phase II studies, and which applied U(H)PLC-MS assays demonstrate the need for larger RT databases that can be shared across the exposomics and metabolomics

communities. Although this is currently challenged by the requirement that authentic chemical standards be analysed on the same analytical system as the samples will be, much headway has been made to find alternative ways to create such databases. Over the last few years, many groups in metabolomics have investigated efficient ways to predict RTs. For example, Creek and co-authors illustrated a RT prediction-model for HILIC assays in metabolomics studies, finding that these predicted RTs reduced FP annotation by 40% (424). It is, however, difficult to accurately predict RTs due to RT drift, but this can be improved by analysis of standards to aid prediction with each NTA analysis (424). Indeed, many metabolomics tools have appeared in the last five years for exactly this purpose (425–428). A similar effort is required for analysis of xenobiotics, particularly in complex sample matrices such as house-dust. This has begun for the creation of chemical databases (429), but not for the implementation of RT prediction for these chemicals.

The US-EPA's ENTACT study demonstrated that both (nano)ESI-DIMS and U(H)PLC-MS assays can be used to measure xenobiotics in various and increasingly complex sample matrices, but highlighted the challenges of applying each technique, particularly in the complex house-dust matrix. The (nano)ESI-DIMS assays are challenged by a lack of separation before *m/z* analysis, therefore ion suppression and/or enhancement can be observed. As a result of the lack of separation, isomers and isobars cannot be differentiated. Moreover, (nano)ESI-DIMS data are very large and complex, and the lack of a third dimension of data (away from *m/z* and response) can make removal of background ions more challenging and increases FPRs during the annotation process. However, TP annotation in these datasets is ~50% in a clean solvent matrix, and slightly lower in more complex sample matrices. This makes this technique useful for screening through very large reference lists. However, the use of U(H)PLC-MS circumvents the issues of ion suppression and/or enhancement, and the third dimension of data in the RT domain reduces data complexity and makes removal of background peaks more efficient. This reduced data complexity increases TP annotation rates, but without the use of RT matching the FPRs are very high. Creation of RT databases is difficult and is the next step in improving applicability of NTA in exposomics.

Finally, the comparison between the (nano)ESI-DIMS and U(H)PLC-MS methods highlighted the value of selecting data processing parameters carefully during database matching. During Phase I experiments, FP annotations were lower for the (nano)ESI-DIMS methods. These data

were very strictly filtered and appropriate database matching criteria selected, reducing FP annotation even in the absence of a separation strategy. The next chapter presented in this thesis will do the same for U(H)PLC-MS data, investigating which parameters affect annotation the most.

5 CHAPTER 5: Optimising U(H)PLC-MS Metabolite Annotation Parameters for Full-Scan Data using the Birmingham mEtabolite Annotation for Mass Spectrometry (BEAMS) Tool

Abstract

Data processing parameters such as intensity correlation coefficients (R) and annotation parameters such as adduct, and isotope reference lists can play a crucial role in reducing false positive rates (FPRs) by ensuring that different ion forms of the same analyte (degenerate features) are grouped together. There is, however, currently no standardisation on optimal parameters. To this end, the Birmingham mEtabolite Annotation for Mass Spectrometry (BEAMS), a software tool dedicated to annotation of metabolomics U(H)PLC-MS data, was used to characterise the impact of annotation parameters, and changing the length of various annotation reference lists in relation to FPRs.

Correlation coefficient thresholds of 0. 0.25, 0.7, 0.7, and 0.9, and p-value thresholds of 0.05, 0.005, and 0.0005 were tested, whilst mass error tolerances of 0.5, 1.0, 3.0, 5.0, 10, and 25ppm were also tested on metabolomics data collected for serum samples. Finally, varying lengths of adducts, isotopes, and neutral loss lists used to annotate degenerate features were tested. When correlation analysis and retention time (RT) similarity were applied to group and annotate degenerate features, and then match against metabolite databases using BEAMS, the TPRs and FPRs were most affected by the correlation thresholds, mass error tolerances, and adduct, isotope, and neutral loss lists used to annotate degenerate features. Generally, for both the polar and the non-polar assays, stricter correlation thresholds decreased both the TPRs and FPRs, but the optimal threshold was 1. The strictest mass error tolerances reduced both TPRs and FPRs, with the maximum TPRs observed between mass error tolerances of 3-10ppm. An optimal mass error tolerance of 5ppm was selected. Finally, longer adducts, isotope, and neutral loss lists increased both TPRs and FPRs of degenerate features for the positive ion-mode datasets, whilst they reduced TPRs in the negative ion-mode datasets whilst still increasing FPRs.

Shorter, more conservative reference lists can decrease FPRs, and it is recommended that anyone processing NTA data investigate degeneracies within their datasets and group them accordingly.

5.1 Introduction

There are relatively few investigations into the impact of the grouping parameters (correlation analysis of different ion forms of the same analyte, and reference lists for annotation of these ion forms) applied during annotation in metabolomics, even though such studies offer insight into how to maximise this step. As such, this chapter will focus on just that, investigating the impact of varying grouping, isotope peak pattern matching, and database matching criteria. Not only will this work recommend optimised parameters for application within the Analytical and Clinical Metabolomics Group (ACMG) at the University of Birmingham, but this work is also relevant to anyone within the metabolomics community applying unsupervised clustering annotation tools that perform retention time (RT) similarity matching, correlation analysis, and subsequent annotation of degenerate features like RAMClust (332), MSClust (327), and PUTMEDID (334).

The Analytical and Clinical Metabolomics Group (ACMG) at the University of Birmingham developed an open-source metabolite annotation tool called the Birmingham mEtabolite Annotation for Mass Spectrometry (BEAMS) tool, available at (https://more.bham.ac.uk/BEAMS/contact/). This is a more recent version of the tool PUTMEDID (334), which has had success in many metabolomics studies (430–438). The BEAMS tool similarly operates in multiple serial steps, divided into grouping features of the same metabolite, annotating peak patterns, annotating molecular formulae, and annotating compounds/metabolites. The feature grouping involves RT similarity matching and pairwise correlation analysis between two features, both governed by parameters including maximum RT difference, correlation threshold, correlation method (Pearson or Spearman Rank) and correlation p-value threshold. The maximum RT difference is a filtering parameter used to filter correlating features that are considered to belong to different metabolites, i.e., all correlating *m/z* features within a user specified RT difference tolerance and greater than or equal to the user defined correlation threshold are considered as different features of the same metabolite. The correlation threshold is user defined and is applied to filter the correlation coefficients (r) of two *m/z* features, which measures the strength of the relationship between the two *m/z* feature's analytical responses. An r value of 1 reflects a quintessential synchronicity in intensity increase between two features, whilst a score of -1 reflects an inverse relationship. The grouping method is a selection between two statistical

tests used to calculate correlations, Pearson's correlation coefficient (439), and Spearman's rank correlation coefficient (440), whilst the p-value threshold filters out correlating features below a user defined p-value threshold. The grouping method can impact annotation accuracy as well as the number of true positives (TP or true positive rate (TPR)), false positives (FP or false positive rate (FPR)), and false negatives (FN or false negative rate (FNR)). If the grouping criteria are too strict, then features belonging to a single metabolite can be divided into separate metabolites. Conversely, too liberal grouping criteria can result in forced grouping of features that do not belong together, thereby interpreting two separate metabolites as a single metabolite and its degenerate features. This phenomenon is exacerbated by the fact that correlation coefficients between related *m/z* features are not always high, as discussed briefly by Kachman and co-authors (322). In such a scenario, RT similarity is important, and must be carefully considered.

Once grouping is complete, annotation of peak patterns is carried out by calculating *m/z* differences within each group and matching these to entries in reference lists of adducts, isotopes and neutral losses. Matching of *m/z* differences between features to the reference lists is restrained by a user defined mass error tolerance. Since the annotation of adducts, isotopes, and neutral losses is limited by the reference lists to match against, this step can also be greatly influenced by which degeneracies are searched for, and adduct, isotope, and neutral loss lists must be carefully considered. One strategy is to use chemical knowledge to create very large and exhaustive reference lists, but it has not been extensively investigated how this will impact on results, especially the rates of true and FP annotations.

The next step is annotation of molecular formulae, which involves matching against a reference list of molecular formulae. This step can be toggled on or off. Skipping this step means molecular formulae annotations are not made, and database matching (next step) is carried out by *m/z*, whilst including this step annotates molecular formulae and these are then used during database matching against a reference list (next step). Finally, database matching against a reference list is carried out. Here, the different ion forms within each created group are annotated within a user defined mass error tolerance. This too can influence the accuracy of annotation and the number of false and TPs. Annotations carried out using strict matching criteria can be affected by the mass accuracy of the data collected, and this must be considered, albeit difficult to determine in NTA (441). On the other hand,

annotations using wide matching criteria will undoubtedly increase FP annotation. A summary of the BEAMS annotation workflow is shown in Figure 30 and a summary of the grouping is shown in Figure 31.

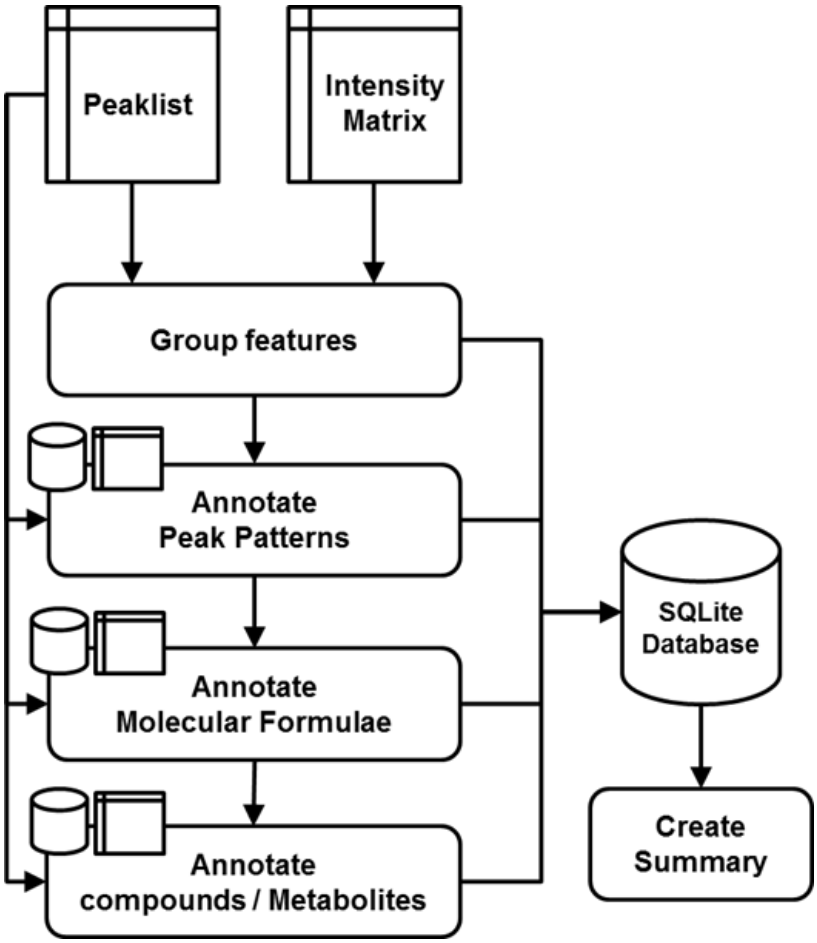


Figure 30: The BEAMS annotation workflow is summarised here. It involves grouping features though calculating correlations between features across all samples and RT similarity matching, annotating isotopic peak patterns through calculating *m/z* differences within each RT group of correlating features and matching against a reference list of isotopes, annotating molecular formulae by matching against a reference list of molecular formulae, and annotating metabolites by matching against a metabolite database or reference list.

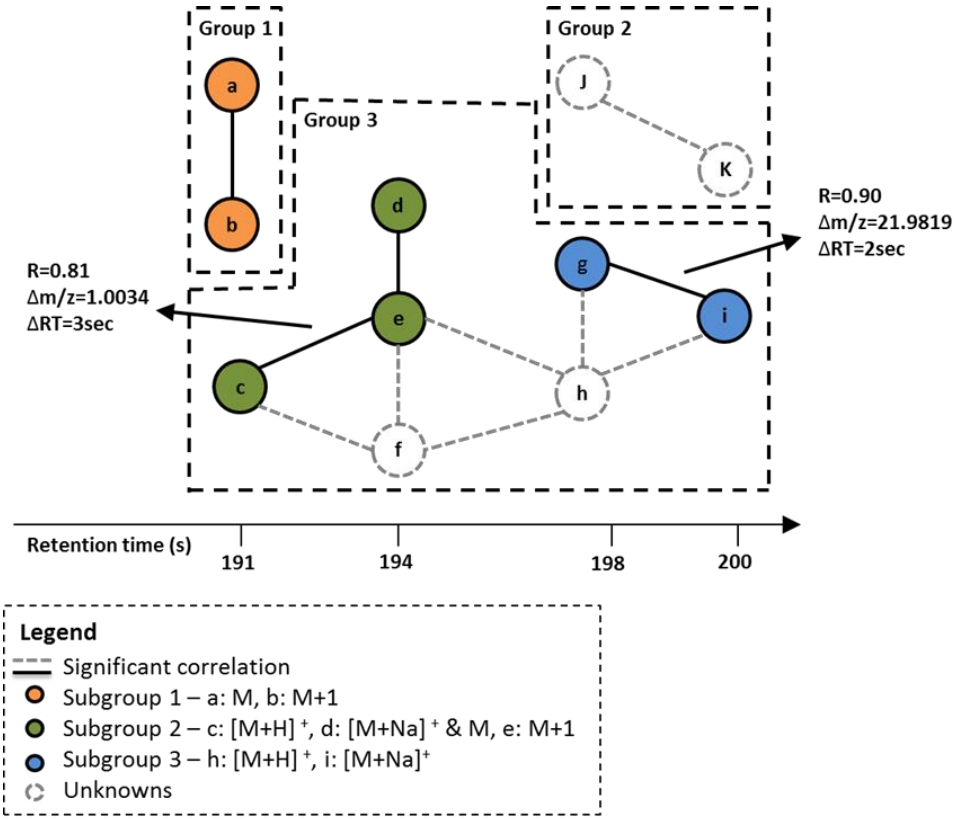


Figure 31: A summary of how BEAMS executes the grouping is shown. Correlation analysis is carried out on *m/z* feature intensities across all samples in the dataset. Based on the correlation analysis results and RT similarity, *m/z* features are placed into "groups," whilst annotated adducts, isotopes, and neutral losses are assigned a "sub-group."

The impact of the annotation parameters on FP annotations has already been demonstrated in the first two chapters of this thesis. Analysis of the same standard mixtures using an (nano)ESI-DIMS method suffering from a lack of analyte separation prior to MS detection still managed to yield lower FP annotations than the gold standard U(H)PLC-MS approach. This strongly demonstrated the impact of using stricter database matching mass error tolerances, and just how effectively this one single parameter can be at reducing FP annotations. The objectives of this research chapter were therefore to investigate the impact of the various annotation parameters applied in the software tool BEAMS (and in other software), from the feature grouping to the database matching. This optimisation was carried out using four U(H)PLC-MS methods; HILIC positive ion-mode, HILIC negative ion-mode, lipids positive ion-mode and lipids negative ion-mode applied to serum samples. The use of these four assays enabled characterisation of the impact of annotation parameters on both polar and non-polar metabolites. The optimised parameters were recommended for use within the ACGM.

5.2 Results and Discussion

5.2.1 Maximum Retention Time Difference

Four U(H)PLC-MS methods were applied for the analysis of human blood serum. These data were processed using XCMS (292), and the resultant peak matrices used to test the impact of varying the maximum RT difference (maxRT) setting during the BEAMS annotation workflow. These results were then used to determine an optimal maxRT recommended to all users of BEAMS and other software using RT similarity and correlation analysis to group related metabolites, or indeed any type of analyte in NTA. MaxRTs of 0.2s, 0.5s, 1.0s, 2.0s, and 5.0s were investigated by setting all other parameters as default (section 2.3.8) and varying only the maxRT parameter in the sequence. Data analysis of the results considered the following:

1. How many unique *m/z*-RT pairs had one or more annotations reported (i.e., metabolite ID (HMDB identification number) or lipid ID (LM identification number)?
2. What was the distribution of the number of annotations for each unique *m/z*-RT pair?
3. How many unique *m/z* were grouped according to the maxRT parameter to form a single RT group based on RT similarity?
4. What was the total number of unique *m/z*-RT pairs grouped?
5. How many unique metabolite groups were reported?
6. For each unique metabolite group, how many annotations were there?

Results for questions 1-7 are reported and discussed in the sub-sections below.

5.2.1.1 Unique *m/z*-retention time pairs

The number of unique *m/z*-RT pairs was calculated for each of five maximum retention difference windows (0.2s, 0.5s, 1s, 2s, and 5s), and for four U(H)PLC-MS datasets separately. This calculation informed as to whether the maxRT parameter affected the number of *m/z*-RT pairs reported within each dataset.

For the HILIC positive and negative ion-modes, and the Lipids positive and negative ion-modes, the number of unique *m/z*-RT pairs was unaffected by varying maxRTs (0.2s, 0.5s, 1s, 2s, and 5s), yielding 8819, 6605, 5611, and 6558 unique *m/z*-RT pairs, respectively.

5.2.1.2 Unique *m/z*-retention time pairs with at least one or more annotations

The number of unique *m/z*-RT pairs with at least one metabolite annotation was calculated for each of 5 maximum retention difference windows (0.2s, 0.5s, 1s, 2s, and 5s). This calculation informed on whether the maxRT parameter affected the number of *m/z*-RT pairs annotated within each dataset.

For all four assays, the number of unique *m/z*-RT pairs with at least one metabolite annotation was not greatly impacted by the maxRT applied (Figure 32 and Figure 33). However, the 0.2s maxRT parameter yielded slightly less unique *m/z*-RT pairs with at least one annotation. This was more noticeable in the HILIC negative and Lipids positive ion-mode datasets, where the 0.2s maxRT parameter yielded ~10% less annotations than the other parameters. This suggests that a 0.2s RT window is too small a value to apply when attempting to maximise the number of annotated MS1 features using RT similarity matching, and points to dataset specific responses to such filtering. The optimised parameters here are therefore suitable for serum samples analysed using similar methods and instruments to those applied here, but these results still have community wide relevance. With most data processing workflows applied using default parameters for ease of use, results here demonstrate the importance of carefully considering which parameters to use when RT similarity is carried out. These results serve as a soft word of caution on using excessively strict RT similarity parameters. However, they do also show that between 0.5-5s, the number of unique *m/z*-RT pairs annotated is largely unchanged, thus there is a wide array of similarity thresholds that could be applied without a reduction in the rate of annotation.

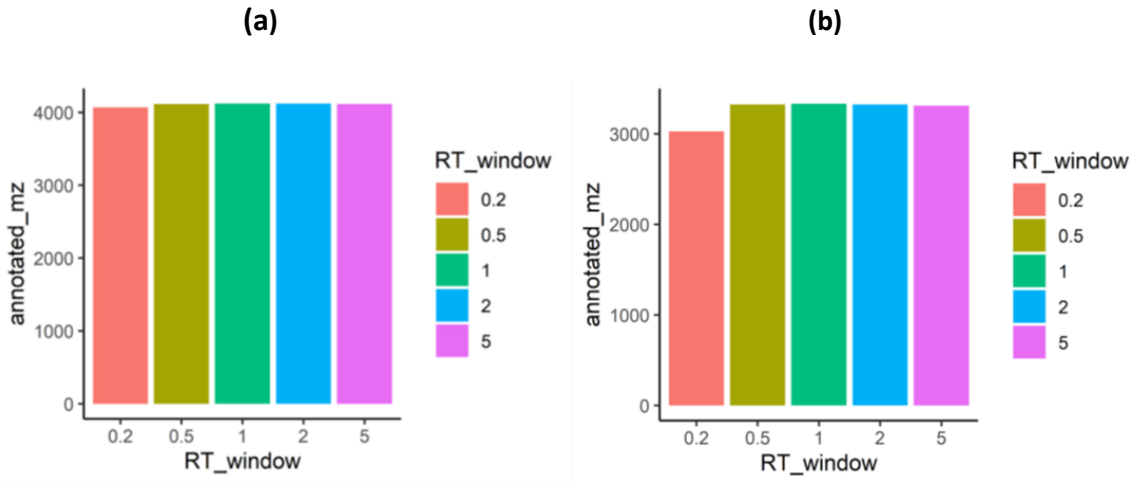


Figure 32: For the HILIC positive ion-mode dataset on the left-hand side (a), ~4100 of 8819 (47%) unique *m/z*-RT pairs had at least one annotation. The changing maxRT parameters (0.2s, 0.5s, 1s, 2s, and 5s) did not impact this result much. For the HILIC negative ion-mode dataset on the right-hand side (b), ~3300 of 6605 (50%) unique *m/z*-RT pairs had at least one annotation. The changing maxRT parameters (0.2s, 0.5s, 1s, 2s, and 5s) did not impact this result much either.

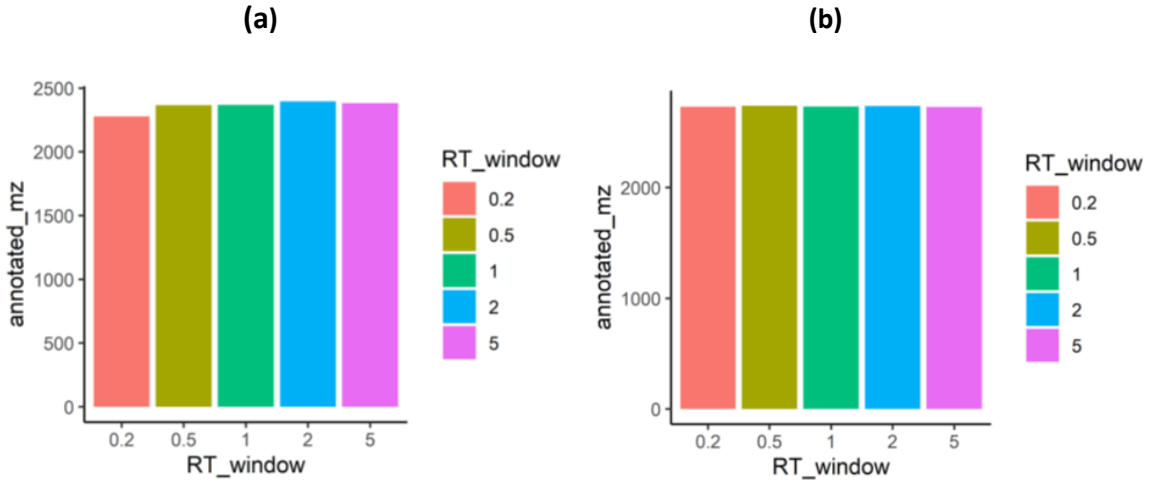


Figure 33: For the Lipids positive ion-mode dataset on the left-hand side (a), ~2300 of 5611 (41%) unique *m/z* had at least 1 annotation. The changing maxRT parameters (0.2s, 0.5s, 1s, 2s, and 5s) did not impact this result. For the Lipids negative ion-mode dataset on the right-hand side (b), ~2700 of 6558 (41%) unique *m/z* had at least one annotation. The changing maxRT parameters (0.2s, 0.5s, 1s, 2s, and 5s) did not impact this result.

5.2.1.3 Distribution of the number of annotations for each unique *m/z*-retention time pair

The distribution of the number of unique annotations for each *m/z*-RT pair was calculated for each of five maximum retention difference windows (0.2s, 0.5s, 1s, 2s, and 5s) for each of the four U(H)PLC-MS datasets. This calculation informed us on whether the maxRT parameter affected the number of annotations assigned to each unique *m/z*-RT pair feature and therefore the number of TP and FP annotations.

The median number of unique annotations for each *m/z*-RT pair, although different between the four assays, was unaffected by changing the maxRT parameters (Figure 34 and Figure 35). Moreover, interquartile ranges were also unaffected. As such, any maxRT parameter could be applied without impacting the unique annotations for each *m/z*-RT pair. The median number of unique annotations for each unique *m/z*-RT pair was higher for both Lipid assays than the HILIC. This is likely due to the isomeric nature of lipid structures (442), resulting in a higher number of isomers being assigned to a single *m/z*-RT pair.

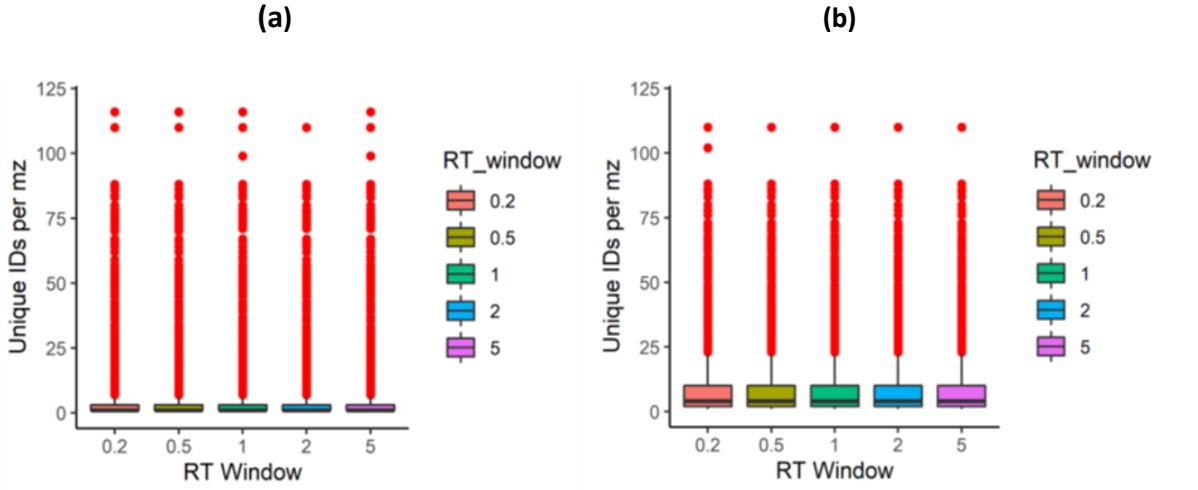


Figure 34: For the HILIC positive ion-mode dataset on the left-hand side (a), the median number of unique annotations per unique *m/z*-RT pair was 3, whilst the interquartile range was from 1-6. The outliers had distributions as high as 117. For each of the tested maxRT parameters (0.2s, 0.5s, 1s, 2s, and 5s), the number of unique annotations assigned to each unique *m*/z-RT pair feature did not change, thus the maxRT parameter did not affect the number of unique annotations assigned to each unique *m/z*-RT pair. For the HILIC negative ion-mode dataset on the right-hand side (b), the median number of unique annotations per unique *m/z*-RT pair was 4, whilst the interquartile range was from 2-10. The outliers had distributions as high as 117. For each of the tested maxRT parameters (0.2s, 0.5s, 1s, 2s, and 5s), the number of unique annotations assigned to each unique *m/z*-RT pair feature did not change, thus the maxRT parameter did not affect the number of unique annotations assigned to each unique *m/z*-RT pair.

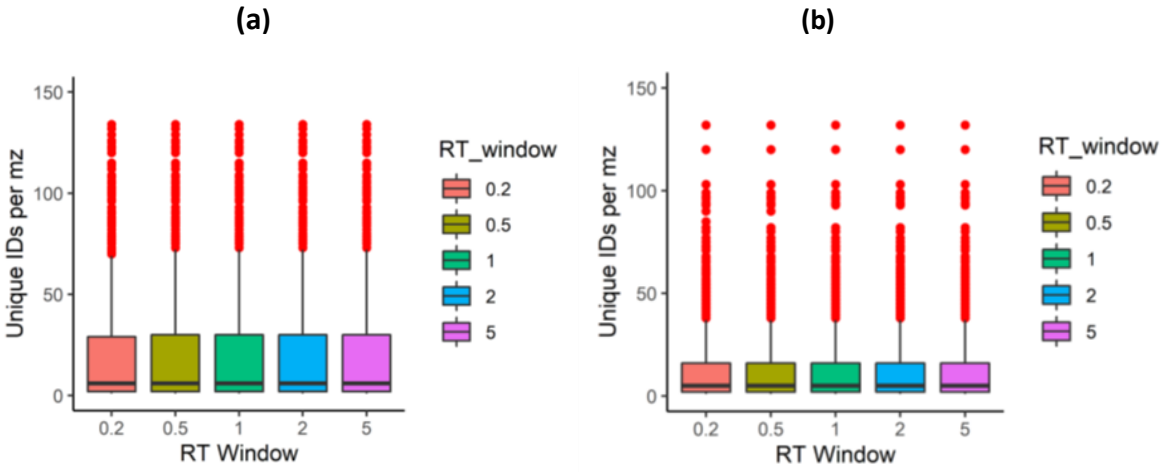


Figure 35: For the Lipids positive ion-mode dataset on the left-hand side (a), the median number of unique annotations per unique *m/z* was 6, whilst the interquartile range was from 2-28. The outliers had distributions as high as 130. For each of the tested maxRT parameters (0.2s, 0.5s, 1s, 2s, and 5s), the number of unique annotations assigned to each unique *m/z* feature did not change, thus the maxRT parameter did not affect the number of unique annotations assigned to each unique *m/z*. For the Lipids negative ion-mode dataset on the right-hand side (b), the median number of unique annotations per unique *m/z* was five, whilst the interquartile range was from 2-16. The outliers had distributions as high as 130. For each of the tested maxRT parameters (0.2s, 0.5s, 1s, 2s, and 5s), the number of unique annotations assigned to each unique *m/z* feature did not change, thus the maxRT parameter did not affect the number of unique annotations assigned to each unique *m/z*.

5.2.1.4 Unique *m/z* features grouped according to retention time similarity

The number of unique *m/z* features grouped together in each RT group was calculated for each of 5 maximum retention difference windows (0.2s, 0.5s, 1s, 2s, and 5s) for each of four U(H)PLC-MS datasets. This calculation informed on the impact of the maxRT parameter on the number of RT groups formed and the number of unique *m/z* within each group that likely represent degenerate feature of the same metabolite.

All four assays had different median numbers of *m/z* features grouped but varying the maxRT parameter did not affect this. Moreover, for most assays, the maxRT parameter also did not impact the outlying number of *m/z* features grouped, with only the HILIC positive ion-mode dataset showing an increase in *m/z* features grouped with an increase in the maxRT parameter (Figure 36, Figure 37, Figure 38, Figure 39). These results can seem a little confusing. Creating smaller RT groups by using smaller maxRTs logically should result in less *m/z* per group. However, since this step was used to filter correlating *m/z*, it makes much

more sense, and indeed reveals the importance of correlation analysis for determining related metabolites. With the number of *m/z* within each group remaining constant for 3 out of the 4 assays (HILIC negative ion-mode, and Lipids positive and negative ion-modes), the results suggested that *m/z* intensity correlation analysis alone was effective at determining related metabolites in U(H)PLC-MS data using BEAMS workflows, since the number of *m/z* features grouped according to similar RTs is largely unaffected by the width of the RT window.

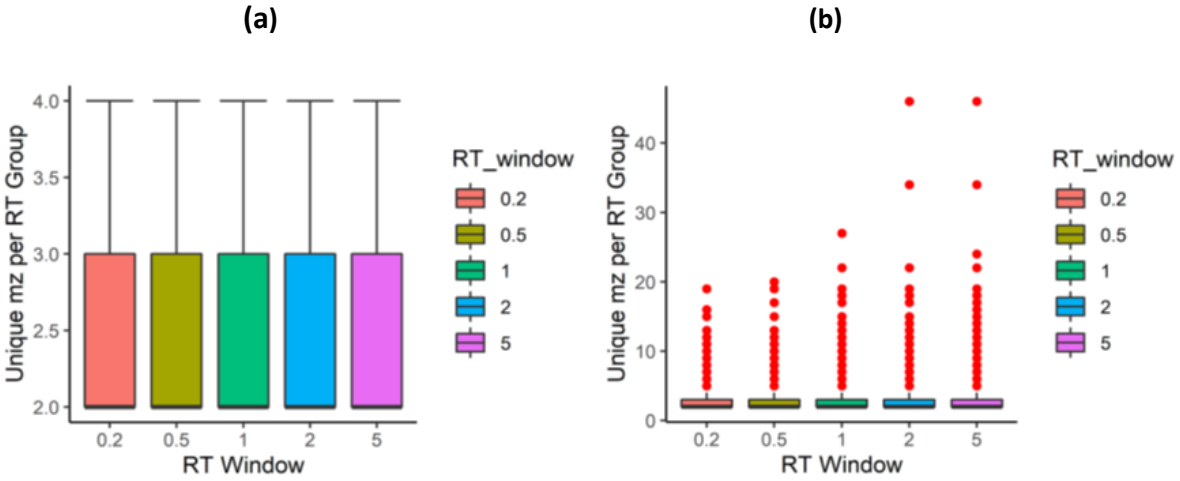


Figure 36: For the HILIC positive ion-mode dataset, the median number of unique *m/z* features per RT group was two, and the interquartile range was from 2-3. The outliers had values ranging between 20 and 40, and a general increase in the number of unique *m/z* per RT group was observed for the outlying data. However, for the rest of the data, no change in the number unique *m/z* per RT group was observed when the multiple values of the maxRT parameter were tested (0.2s, 0.5s, 1s, 2s, and 5s).

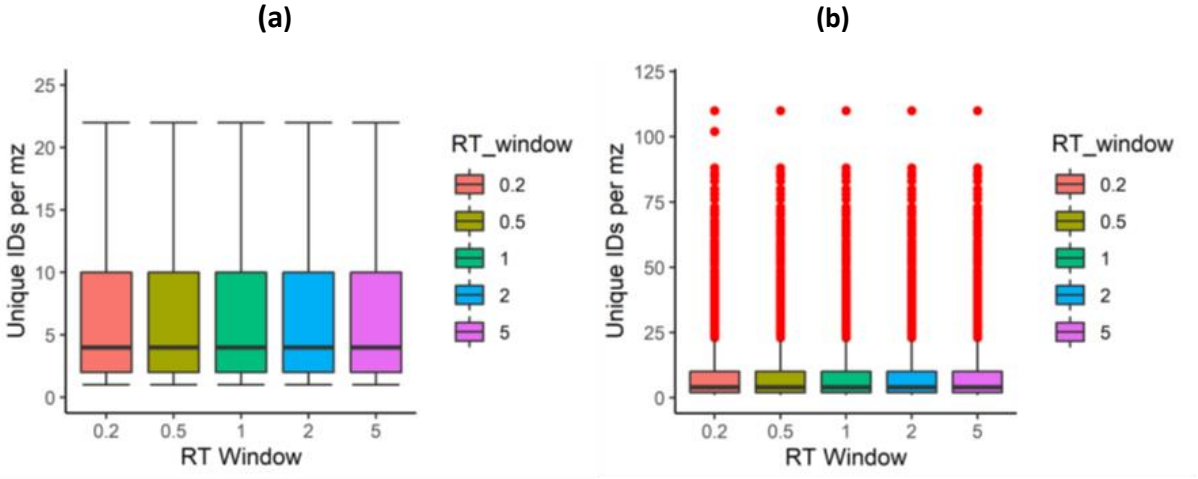


Figure 37: For the HILIC negative ion-mode dataset, the median number of unique *m/z* features per RT group was four, and the interquartile range was from 2-10. The outliers had values as high as 117. However, no change in the number unique *m/z* per RT group was observed when the multiple values of the maxRT parameter were tested (0.2s, 0.5s, 1s, 2s, and 5s).

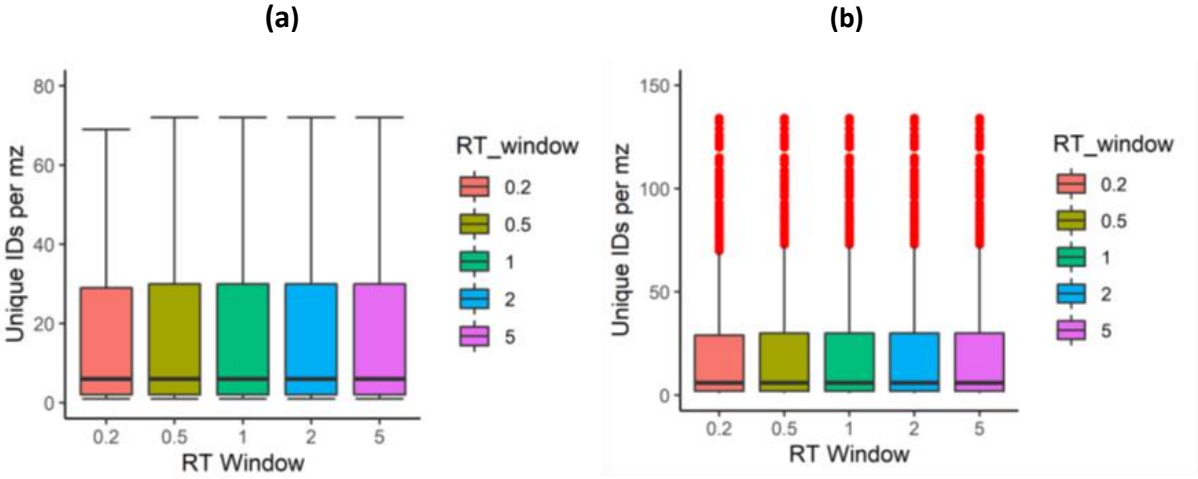


Figure 38: For the Lipids positive ion-mode dataset, the median number of unique *m/z* features per RT group was six, and the interquartile range was from 2-30. The outliers had values as high as 130. However, no change in the number unique *m/z* per RT group was observed when the multiple values of the maxRT parameter were tested (0.2s, 0.5s, 1s, 2s, and 5s).

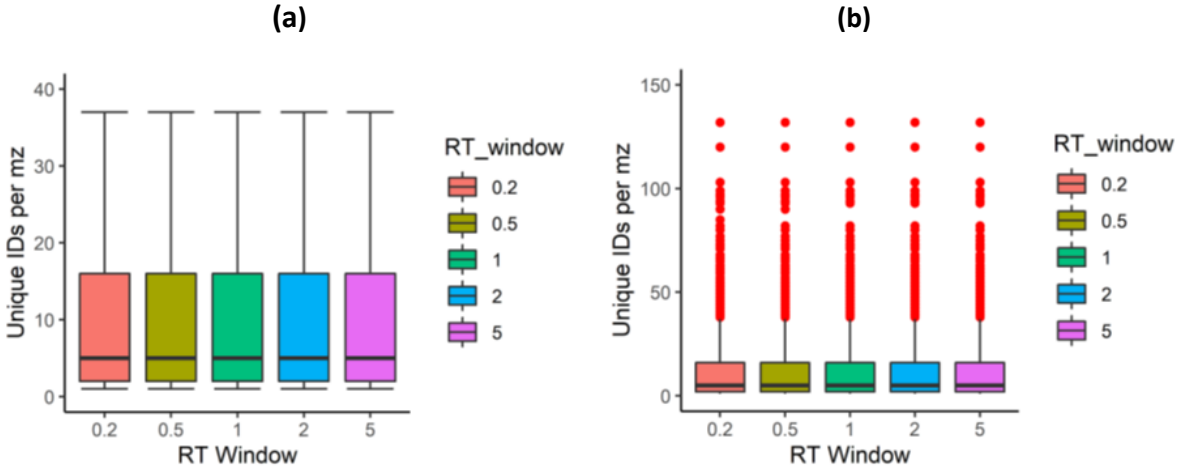


Figure 39: For the Lipids negative ion-mode dataset, the median number of unique *m/z* features per RT group was five, and the interquartile range was from 2-15. The outliers had values as high as 130. However, no change in the number unique *m/z* per RT group was observed when the multiple values of the maxRT parameter were tested (0.2s, 0.5s, 1s, 2s, and 5s).

5.2.1.5 Total number of unique *m/z* features grouped according to retention time similarity

The total number of unique *m/z* features grouped was calculated for each of 5 maximum retention difference windows (0.2s, 0.5s, 1s, 2s, and 5s). In this calculation, a unique *m/z* appearing three times with different RTs was only counted once. This calculation sought to determine the total number of unique *m/z* features assigned to a group with other *m/z* features through correlation analysis and RT similarity. This calculation informed on whether the maxRT parameter affected the number of *m/z* grouped overall.

Results for the HILIC and Lipids positive and negative ion-mode datasets are shown in Figure 40 and Figure 41. Degenerate features can inflate the percentage of unannotated *m/z* data in metabolomics. Moreover, considering *m/z* features of the same metabolite as separate metabolites can also increase FP annotation. It is therefore important to group *m/z* features accordingly. Results presented here demonstrate that the 0.2s maxRT is too small, as the smallest number of total *m/z* features grouped is achieved using this parameter. This aligns well with expectation. Degenerate features of the same metabolite are expected to co-elute, and thus must have the same RT. However, the peak apex for each degenerate feature may vary, hence the lower number of *m/z* features grouped when a 0.2s maxRT is applied,

suggesting a peak apex deviation larger than 0.2s. The changes between 1-5s are minimal, thus an appropriate optimal maxRT must be between 1-5s.

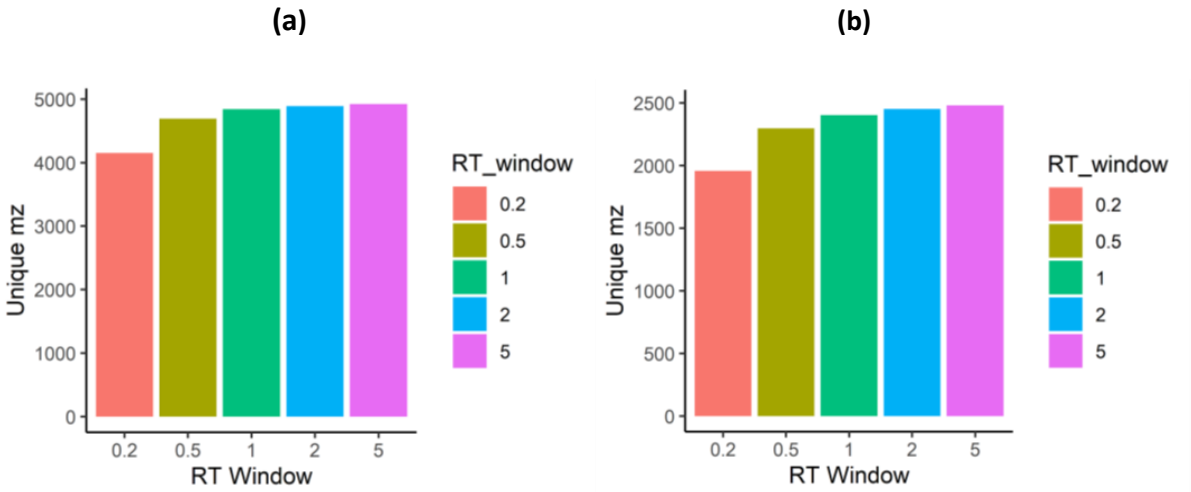


Figure 40: For the HILIC positive ion-mode dataset on the left-hand side (a), the total number of unique *m/z* grouped increased with widening maxRT windows, with the largest increase observed between 0.2s and 0.5s. Further increases are observed between 0.5s and 5s, but these are much smaller. Indeed, counts between 1-5s are minimal, and an optimal maxRT value can be selected from any of those values. For the HILIC negative ion-mode dataset on the right-hand side (b), the total number of unique *m/z* grouped increased with widening maxRT windows, with the largest increase observed between 0.2s and 0.5s. Further increases are observed between 0.5s and 5s, but these are much smaller. Indeed, counts between 1-5s are minimal, and an optimal maxRT value can be selected from any of those values.

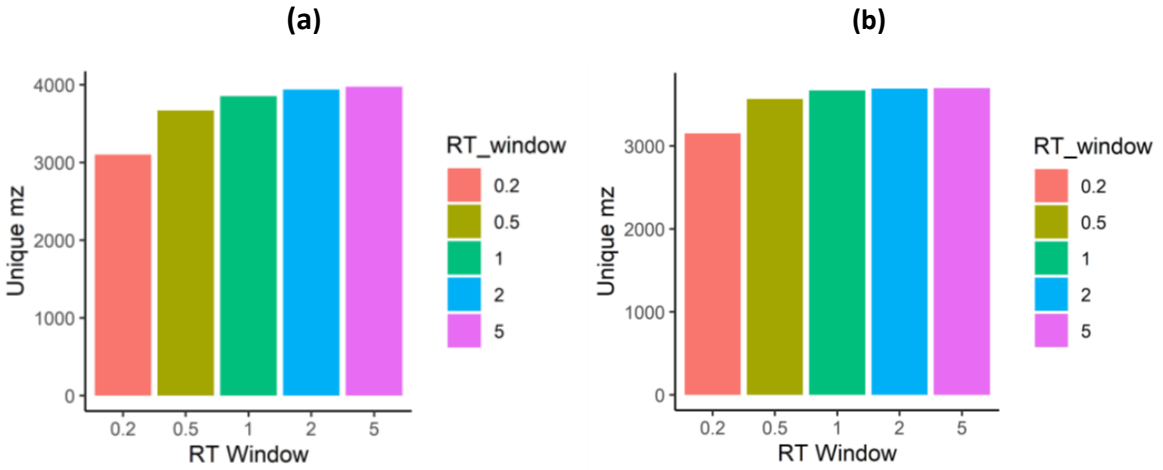


Figure 41: For the Lipids positive ion-mode dataset on the left-hand side (a), the total number of unique *m/z* grouped increased with widening maxRT windows, with the largest increase observed between 0.2s and 0.5s. Further increases are observed between 0.5s and 5s, but these are much smaller. Indeed, counts between 1-5s are minimal, and an optimal maxRT value can be selected from any of those values. For the Lipids negative ion-mode dataset on the right-hand-side (b), the total number of unique *m/z* grouped increased with widening maxRT windows, with the largest increase observed between 0.2s and 0.5s. Further increases are observed between 0.5s and 5s, but these are much smaller. Indeed, counts between 1-5s are minimal, and an optimal maxRT value can be selected from any of those values.

5.2.1.6 Unique retention time groups created

The number of RT groups created was calculated for each of five maximum retention difference windows (0.2s, 0.5s, 1s, 2s, and 5s). A single RT group was comprised of all *m/z* that had correlations above the user defined threshold (a default of r=0.7 and p=0.05 for these experiments) and all *m/z* with RTs that fall within each specified maxRT. Moreover, the adducts, isotopes, and neutral losses denoted by each correlated *m/z* difference and falling within the specified maxRT also had to be annotated. Each group therefore comprised of *m/z* features likely belonging to a single metabolite. This calculation informed on whether the maxRT parameter affected the number of RT groups created.

Results for the HILIC and Lipids positive and negative ion-mode datasets, respectively, are shown in Figure 42 and Figure 43. Generally, none of the changes observed were significant, but the jump from 0.2s to 0.5s maxRT had between 10-25% less annotations than the other RT differences tested. This suggests that the 0.2s maxRT parameter was too strict. Chromatographic peaks of the same or related analytes should exhibit similar peak shapes, influenced by various aspects such as the integration algorithm and method resolution. As

such, the peak apex of chromatograms belonging to related, co-eluting analytes can vary slightly. Results here demonstrate that a 0.2s RT similarity threshold is too strict, allowing insufficient leeway for peak apex detection-related imperfections (443). Researchers using RT similarity to group degenerate features must be wary of this.

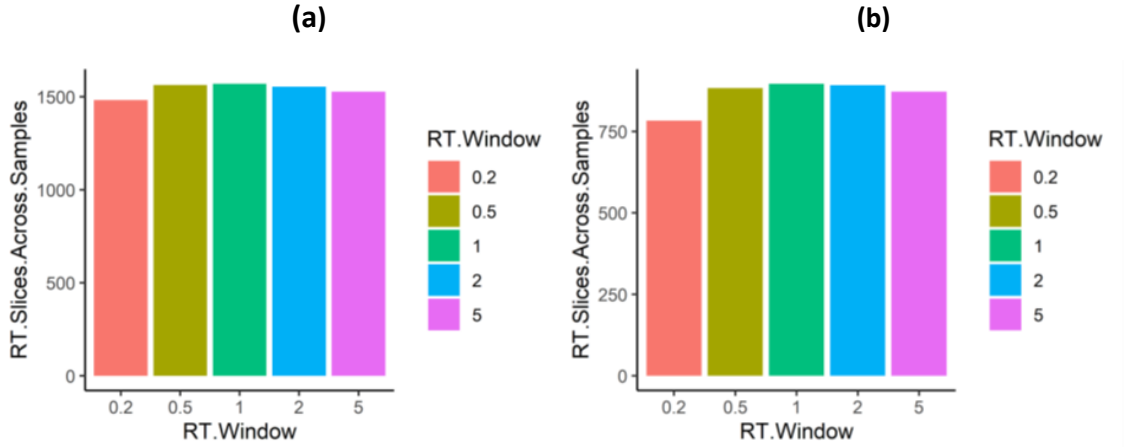


Figure 42: For the HILIC positive ion-mode on the left-hand side (a), ~1500 groups were created, and this number was similar for all RT windows, although 0.2s window had the least groups. The biggest number of groups was observed at 1s maxRT. For the HILIC negative ion-mode on the right-hand side (b), ~800 groups were created, and this number was similar for all RT windows, although 0.2s window had the least groups. The largest number of groups was observed at 1s maxRT.

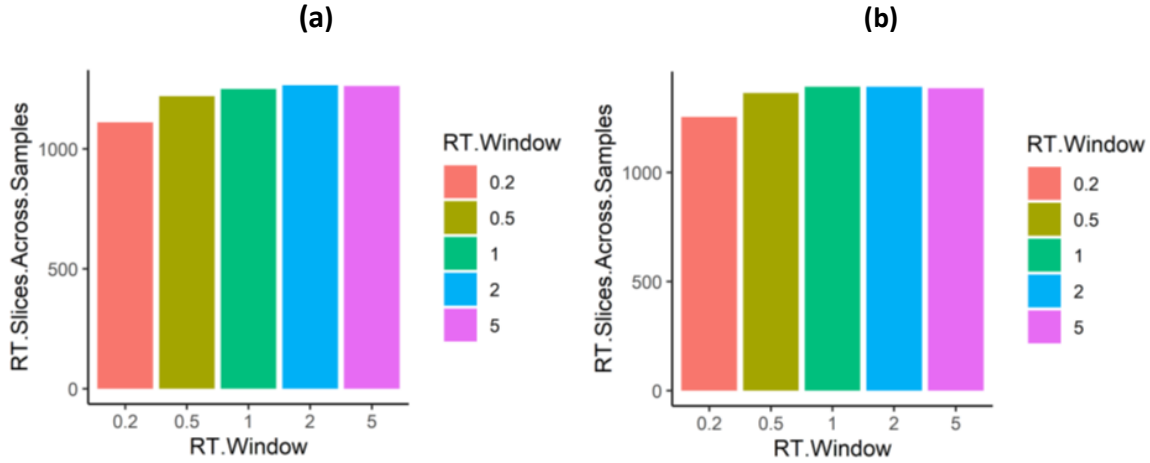


Figure 43: For the Lipids positive ion-mode on the left-hand side (a), ~1200 groups were created, and this number was similar for all RT windows, although 0.2s window had the least groups. The largest number of groups was observed at 5s maxRT, although differences between 1s and 5s were almost negligible. For the Lipids negative ion-mode on the right-hand side (b), ~1200 groups were created, and this number was similar for all RT windows, although 0.2s window had the least groups. The largest number of groups was observed at 1s and 2s maxRTs.

5.2.1.7 Annotations per unique retention time group

The number of unique annotations reported per RT group was calculated for each of 5 maximum retention difference windows (0.2s, 0.5s, 1s, 2s, and 5s). This calculation informed on whether the maxRT parameter affected the number of annotations within each RT group.

Results for each of the four assays are shown in Figure 44 and Figure 45. For most RT groups for all four assays, the number of annotations per group were largely unaffected by the maxRT applied, with the median number of annotations and the interquartile ranges remaining constant or varying only by one or two. However, for the HILIC assays, widening the maximum RT window increased the outlying number of annotations assigned to each RT group. Conversely, for the Lipid positive assay, using a maxRT of 0.2s decreased the spread of the number of annotations per group. For the HILIC assays, an increase in the maxRT was expected to result in more correlating *m/z* features falling within the RT similarity threshold. As such, it was therefore likely to increase the number of annotations. However, for the bulk of the HILIC data, no increase in the number of annotations per RT group was observed. This was because increased grouping of correlating *m/z* features based on RT similarity did not instantly translate into more annotations. As described in section 5.2, a RT group comprised of *m/z*-RT pairs assigned a group_id and sub_group_id. To be assigned a sub_group_id, the *m/z* defects of the *m/z* within a user specified RT similarity threshold had to be annotated. The *m/z* features that were highly correlated despite not belonging to the same metabolite and fell within the user defined maxRT were not counted as they were not assigned a sub_group_id. However, as the maxRT increased, so did the chance of grouping *m/z* features that were above the specified correlation threshold and within the specified RT similarity threshold but did not actually belong to the same metabolite. This undoubtedly resulted in an increase in the number of annotations for a small subset of the data (outliers), albeit incorrect ones as there are some metabolites within HMDB that have mass defects that could be annotated as different adducts (Table 78).

accession	chemical_ formula	monisotopic_molecular weight	[M+H]+	*m/z*_diff
HMDB0012025	C90H158N4O42	1967.035065	1968.042	21.88911
HMDB0033570	C92H148O46	1988.924177	1989.931	
HMDB0011965	C96H168N4O47	2129.087889	2130.095	21.88911
HMDB0033575	C98H158O51	2150.977001	2151.984	

Table 78: Two sets of metabolites in HMDB are shown. The [M+H]+ ions of these compounds have a defect that could be misannotated as a Na adduct. If these two sets of metabolites correlated above the specified threshold, and also had sufficiently similar RTs, then these could be grouped together in a RT group, having a group_id and sub_group_id. It is also possible that one of these could then be annotated as the [M+Na] adduct of the other. Instances like this are unlikely but possible, explaining why for the majority of the HILIC data, an increase in the number of annotations per RT group is not observed, but an increase in the number of annotations with increasing maxRT is observed for the outlying data.

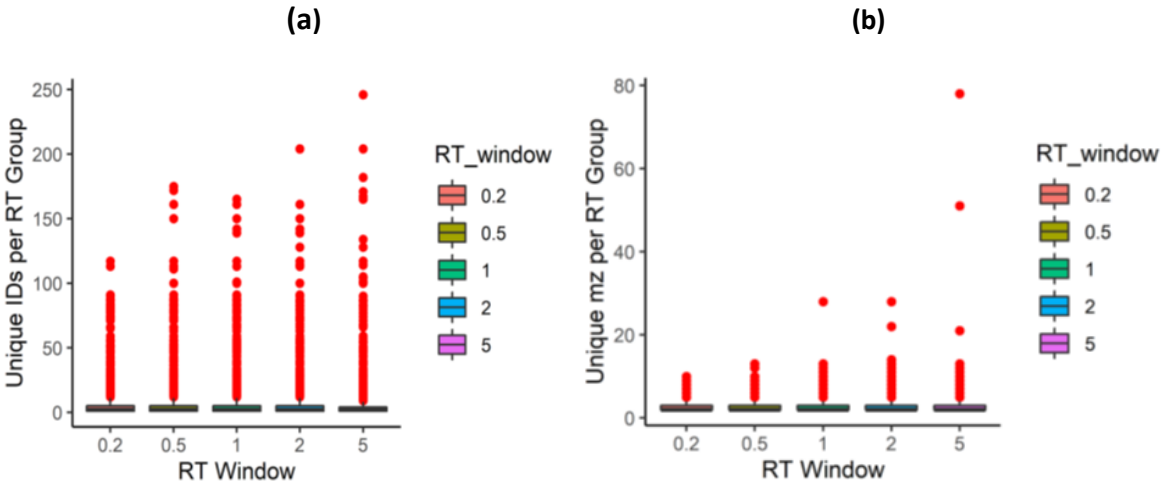


Figure 44: For the HILIC positive ion-mode dataset on the left-hand side, the median number of annotations per RT group was three and remained the same across the different maxRT parameter values applied. The interquartile range was from 1-5 unique annotations per RT group. For the HILIC negative ion-mode dataset on the right-hand side, the median number of annotations per RT group was two and remained the same across the different maxRT parameter values applied. The interquartile range was from 2-3 unique annotations per RT group.

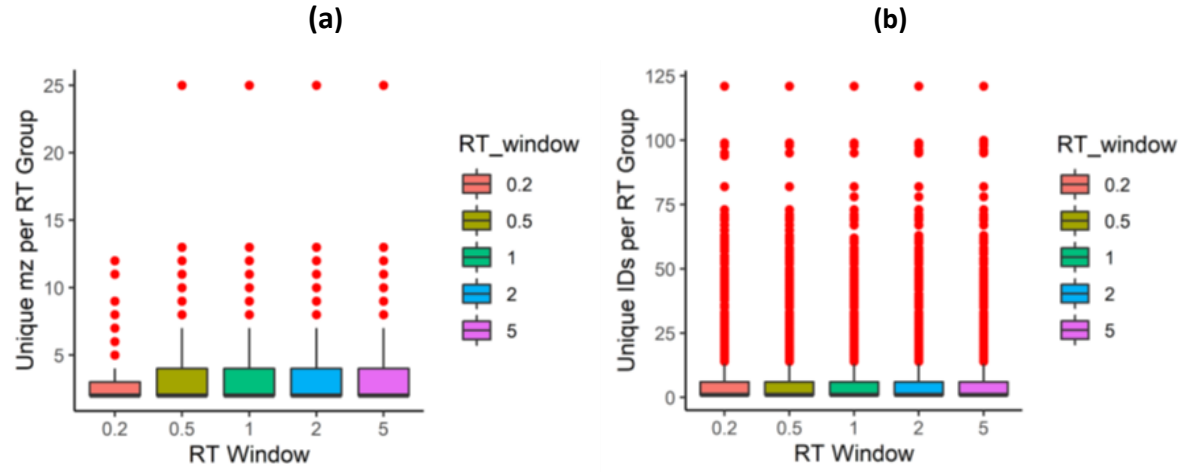


Figure 45: For the Lipids positive ion-mode dataset on the left-hand side (a), the median number of annotations per RT group was two and remained the same across the different maxRT parameter values applied. The interquartile range was from 2-3 unique annotations per RT group for the 0.2s maxRT, and 2-4 for the other maximum RT parameters tested. For the Lipids negative ion-mode dataset on the right-hand side (b), the median number of annotations per RT group was one and remained the same across the different maxRT parameter values applied. The interquartile range was from 1-6 unique annotations per RT group.

5.2.2 Maximum Retention Time Difference: Conclusion

Metabolite annotation was performed on four U(H)PLC-MS datasets (HILIC and Lipids assays in positive and negative ion-modes) applying the BEAMS software tool. To provide optimal BEAMS processing parameters, the maxRT parameter, which is applied during the grouping step of BEAMS annotation, was investigated. Five parameters were run seriatim (0.2s, 0.5s, 1s, 2s, and 5s) and results used to determine the impact, if any, of changing these parameters.

Varying the maxRT did not have much of an impact on the number of unique *m/z*-RT pairs created [1], the unique *m/z* features with at least one annotation [2], the distribution of the number of annotations per unique *m/z* [3], and the number of *m/z* features grouped together per RT group [4]. It did, however, impact the total number of unique *m/z* features grouped together [5], with more features grouped when using wider maximum RT windows. The 0.2s maximum RT window demonstrated the lowest number of groups, with an increase at 0.5s and a further increase at 1.0s. The number of groupings were similar at 1, 2 and 5s. An optimal maximum RT can therefore be selected as 1.0s, since minimal changes are observed thereafter.

Varying the maxRT also impacted the number of RT groups formed [6]. For the HILIC assays, the largest number of RT groups was observed when a maxRT of 1s was used, whilst for the Lipids positive ion-mode, the highest number of RT groups was formed when a 2s maxRT was applied, although the difference in the number of RT groups formed varied negligibly between 1-5s. For the Lipids negative ion-mode the largest number of RT groups was formed jointly by either a 1s or 2s maxRT. Indeed, for all assays, the variation in the number of RT groups formed was negligible between maxRTs of 1-5s.

Finally, the number of unique annotations per RT group [7] was also not impacted by varying the maxRT applied, with median values staying constant with each assay. However, for the HILIC assays, increasing the maxRT values also increased the number of unique annotations in a small subset of the data considered to be outliers.

Based on the total number of unique *m/z* grouped together, and the number of RT groups formed, a 1s maximum retention difference was selected as optimal. It was comforting to observe that application of varying maximum retention differences did not have a huge impact, thus application of different parameters is acceptable and does not come at a huge

detriment to the results obtained. Indeed, RT similarity filtering applied in BEAMS is not the most impactful parameter, but rather simply a filtering step accompanying *m/z* intensity correlation analyses. Moreover, it is left to the user to determine what is most important to them. Applying a tight maxRT can aid in data reduction, and indeed may reduce FP annotation. Conversely, one may choose to use wider maxRTs to increase the number of RT groups obtained, and potentially the number of annotations, particularly when MS1 data is accompanied by MS2 for increased confidence.

5.2.3 Correlation Threshold and p-value

Four U(H)PLC-MS methods, HILIC positive and negative ion-mode, and lipids positive and negative ion-modes, were applied for the analysis of human blood serum. These data were processed using XCMS (292), and then used to test the impact of changing the correlation threshold and p-value settings during the BEAMS annotation workflow. Correlation thresholds (r) of 0.00, 0.25, 0.50, 0.70, and 0.90 were investigated. For each correlation threshold, three p-value thresholds, 0.05, 0.005, and 0.0005, were also tested. For example, a correlation threshold of zero was operated three times using the three p-value thresholds (0-0.05, 0-0.005, 0-0.0005), and the same done for other correlation thresholds to be investigated. The maxRT was set to 1s, as was found to be optimal in the previous section 5.2.1, and all other parameters were set to default (see 2.4.3 for default parameters). Data analysis of the results considered the following:

1. How does changing the correlation and p-value thresholds impact a subset of the data known to be TPs (through identification applying matching to RT and MS/MS data for authentic chemical standards)?
2. How many RT groups were formed using each set of parameters, where a RT group was comprised of correlating *m/z* features with maxRTs of 1s as was found optimal in section 5.2.1?
3. How many of the formed RT groups were annotated?
4. How many unique annotations were there for each RT group?

5.2.3.1 True and false positive rates

A subset of confidently identified metabolites (described in 2.4.5), considered to be TPs, were used to investigate the impact of varying the correlation and p-value thresholds. True and FPRs were calculated as described in section 2.4.5. This calculation informed on whether the correlation and p-value threshold parameters affected the true and FPRs during BEAMS annotation.

Results for the HILIC positive and negative ion-mode assays, showing TPRs and FPRs, respectively, are shown in Figure 46 and Figure 47.

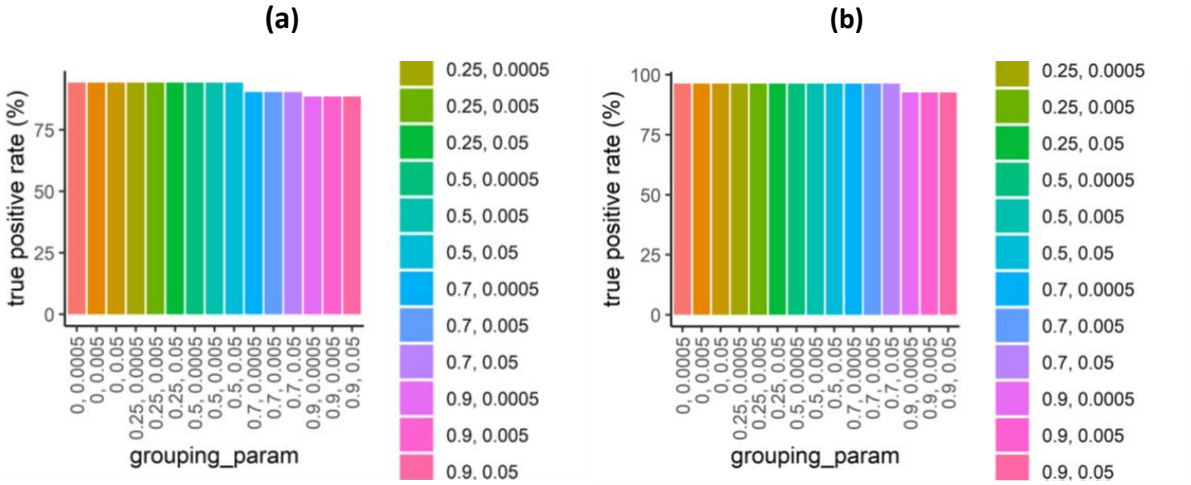


Figure 46: For the HILIC positive ion-mode dataset on the left-hand side (a), TPRs for correlation thresholds of 0-0.5 did not change, remaining at 94%. TPRs dropped to 90% when a correlation threshold of 0.7 was applied, and even further to 88% when a correlation threshold of 0.9 was applied. For all tested correlation thresholds, varying p-value thresholds did not have an impact on TPR. For the HILIC negative ion-mode dataset shown on the right-hand side (b), TPRs for correlation thresholds of 0-0.7 did not change, remaining at 94%. TPRs dropped to 93% when a correlation threshold of 0.9 was applied. For all tested correlation thresholds, varying p-value thresholds did not have an impact on TPR.

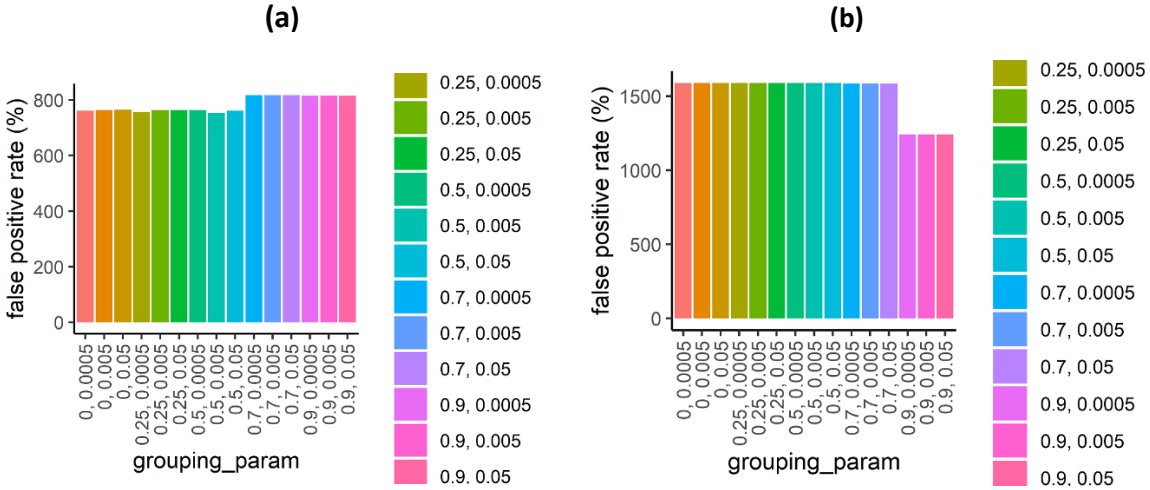


Figure 47: For the HILIC positive ion-mode dataset on the left-hand side (a), FPRs for correlation thresholds of 0-0.5 remained constant at ~760%, exhibiting only minor and negligible variations. This FPR was not affected by varying p-value thresholds. When correlation thresholds of 0.7-0.9 were applied, the FPR increased to 820%, and this was also unaffected by varying p-value thresholds. For the HILIC negative ion-mode dataset on the right-hand side (b), the FPR for correlation thresholds of 0-0.7 was 1 600%. This was unimpacted by varying p-value thresholds and decreased to 1 200% when correlation thresholds of 0.9 were applied. Once again, this FPR was not impacted by carrying p-value thresholds.

It was expected that applying stricter correlation thresholds (thereby grouping only features with high correlation coefficients) would decrease FP annotation, but this was not the case for the HILIC positive ion-mode, where the FPR increased by 40% for correlation thresholds of 0.7-0.9. This was investigated by taking the list of all correct *m/z*-RT pairs and searching for these in the BEAMS outputs when correlations of 0.5 and 0.7 were applied with a p value threshold of 0.05. From this subset, a randomly selected *m/z*-RT pair was used for investigative purposes (M809T259). It was immediately noted that applying the higher correlation threshold of 0.7 resulted in assignment of sodium adducts which were not assigned when a correlation threshold of 0.5 was applied. When correlation thresholds of 0.5 are applied, the sodium adduct may end up correlated to another feature, thus becomes unavailable for assignment as a sodium adduct. When the correlation threshold is increased, the previously accepted incorrect annotation is now unaccepted, and this sodium adducts are "free" to be assigned as such. This assignment of sodium adducts increased the number of annotations assigned to *m/z*-RT pair M809T259. Indeed, when a correlation threshold of 0.5 was applied, this *m/z*-RT pair had a total of 51 annotations and only the protonated forms of

each assignment were present. However, when a correlation threshold of 0.7 was applied, both protonated and sodiated adducts were present, and the *m/z*-RT pair M809T259 had a total of 80 annotations. It was not entirely clear if this was a positive or a negative result. One the one hand, it could be that applying higher correlation thresholds excluded grouping of degenerate features of the same metabolite that had lower correlations, and instead only grouped features that were likely biologically related. As has already been shown in this chapter (5.2.3.4, Table 78), some HMDB entries have *m/z* defects between their protonated ions that could be annotated as sodium adducts. This would explain the increase in sodium adducts observed when higher correlation thresholds were applied, which in turn resulted in an increase in FP annotation. On the other hand, perhaps the lower correlation thresholds were not sufficiently high, increasing the number of *m/z* grouped together as degenerate features of the same metabolite. Perhaps this higher number of *m/z* was too noisy and masked the annotation of sodium adducts, thus reducing the number of FP annotations. Either way, the behaviour of the HILIC positive ion-mode dataset could be explained, and the HILIC negative ion-mode dataset behaved as expected, exhibiting a reduction in FP annotation when higher correlation thresholds were applied.

Results for the Lipids positive and negative ion-mode assays, showing TPRs and FPRs, respectively, are shown in Figure 48 and Figure 49.

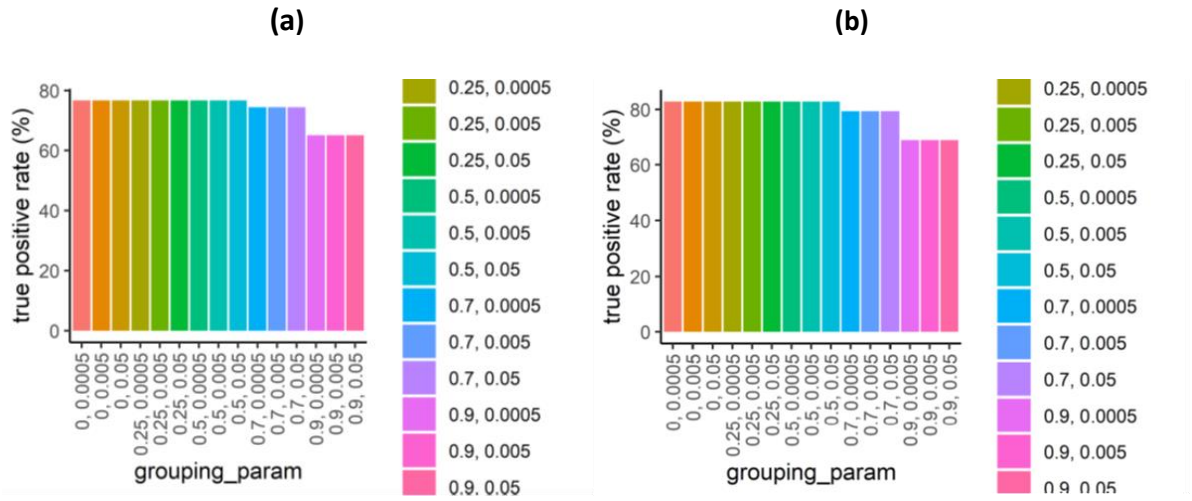


Figure 48: For the Lipids positive ion-mode dataset shown on the left-hand side (a), TPRs for correlation thresholds of 0-0.5 did not change, remaining at 77%. TPRs dropped to 74% when a correlation threshold of 0.7 is applied, and even further to 65% when a correlation threshold of 0.9 was applied. For all tested correlation thresholds, varying p-value thresholds did not have an impact on TPR. For the Lipids negative ion-mode dataset shown on the right-hand side (b), TPRs for correlation thresholds of 0-0.5 did not change, remaining at 83%. TPRs dropped to 79% when a correlation threshold of 0.7 is applied, and even further to 69% when a correlation threshold of 0.9 was applied. For all tested correlation thresholds, varying p-value thresholds did not have an impact on TPR.

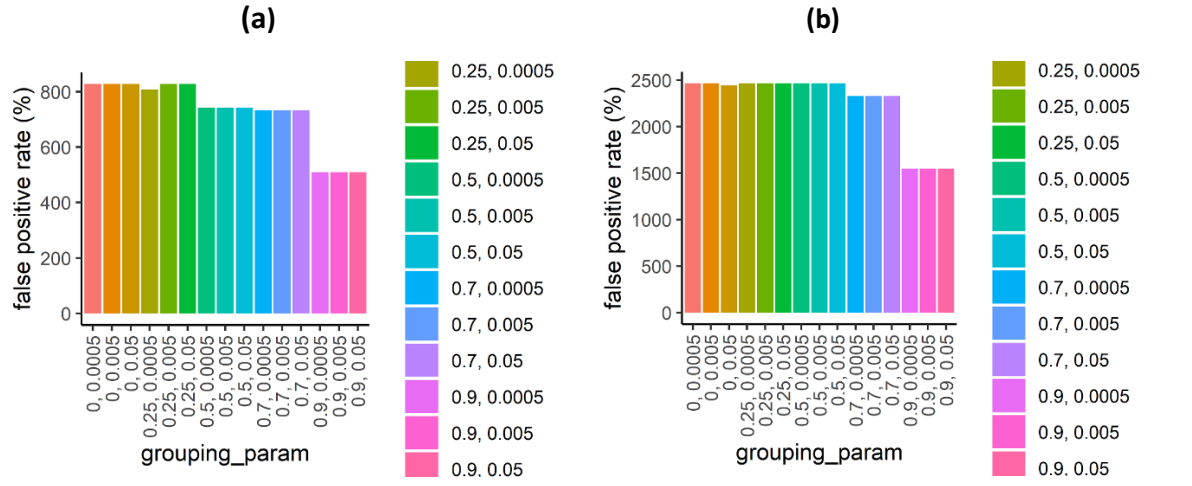


Figure 49: For the Lipids positive ion-mode dataset on the right-hand side (a), the FPR for correlation thresholds of 0-0.25 was 820%. This was unimpacted by varying p-value thresholds and decreased to 730% when correlation thresholds of 0.5-0.7 were applied. Once again, this FPR was not impacted by varying p-value thresholds. Finally, then a correlation threshold of 0.9 was applied, the FPR was at its lowest at 520% and was not changed by varying p value thresholds. For the Lipids negative ion-mode dataset on the left-hand side (b), FPRs for correlation thresholds of 0-0.5 remained constant at ~2 450%, exhibiting only minor and negligible variations. This FPR was not affected by varying p-value thresholds. When a correlation threshold of 0.7 was applied, the FPR decreased to 2 300%, and this was also unaffected by varying p-value thresholds. Finally, at a correlation threshold of 0.9, the FPR dropped to its lowest at 1 500%.

Overall, TPRs were constant and highest between r = 0 to 0.5 for the HILIC positive and Lipids positive and negative ion-modes. The HILIC negative ion-mode had the highest TPRs for correlation thresholds 0 to 0.7. The FPRs were lowest when r =0.9 for the HILIC negative, and Lipids positive and negative ion-modes. For the HILIC negative ion-mode, an opposite trend was observed, with lower correlation thresholds of 0-0.5 yielding the lowest FPRs. For all assays, the p-value thresholds did not have an impact on TPRs. The HILIC negative and Lipids positive and negative ion-mode datasets all exhibited a reduction in FPR with increasing correlation threshold parameters. However, the HILIC positive ion -mode dataset showed an increase in FPR as higher correlation thresholds were applied. An investigation of this phenomenon showed that application of the higher correlation thresholds in this dataset introduced annotation of sodium adducts that were otherwise not present at lower correlation thresholds. It was unclear whether this was due to an increase in grouping of different metabolites as a single metabolite, and thus an increase in FP annotation, or if the stricter correlation parameters resulted in less noisy metabolite spectra that revealed otherwise unannotated adducts. Based on these results, a correlation threshold of 0.5 could be applied for all assays to yield the highest TPRs. Although this yielded higher FPRs, additional strategies such as MS/MS and RT database matching could be used to further reduce FP annotation. Since the p-value thresholds did not impact TPRs, a liberal p-value threshold of 0.05 would suffice.

5.2.3.2 Unique retention time groups created

The number of RT groups formed when applying different correlation and p-value thresholds was calculated. A RT group comprised of *m/z* features with RTs within 1s of one another, as was found to be optimal in section 5.2.1. This calculation informed on whether the correlation and p-value thresholds affected the number of RT groups formed.

Results for the HILIC and Lipids positive and negative ion mode assays, respectively, are shown in Figure 50 and Figure 51. For all assays, the number of groups formed dropped at r =0.7 and was lowest for r =0.9. Where p values affected the number of RT groups, the p value threshold of 0.05 yielded the largest number. The number of groups formed dropped more drastically at r =0.9 for the Lipids assays, decreasing by approximately half for both ion-modes. Based on these results, a correlation threshold of 0.5 using a p value threshold of 0.05 would yield the

largest number of RT groups and was therefore concluded to be optimal. However, it is once again demonstrated that BEAMS parameters can be a very powerful tool for data reduction through grouping, and thus it is up to the user to determine what is most important outcome for their results.

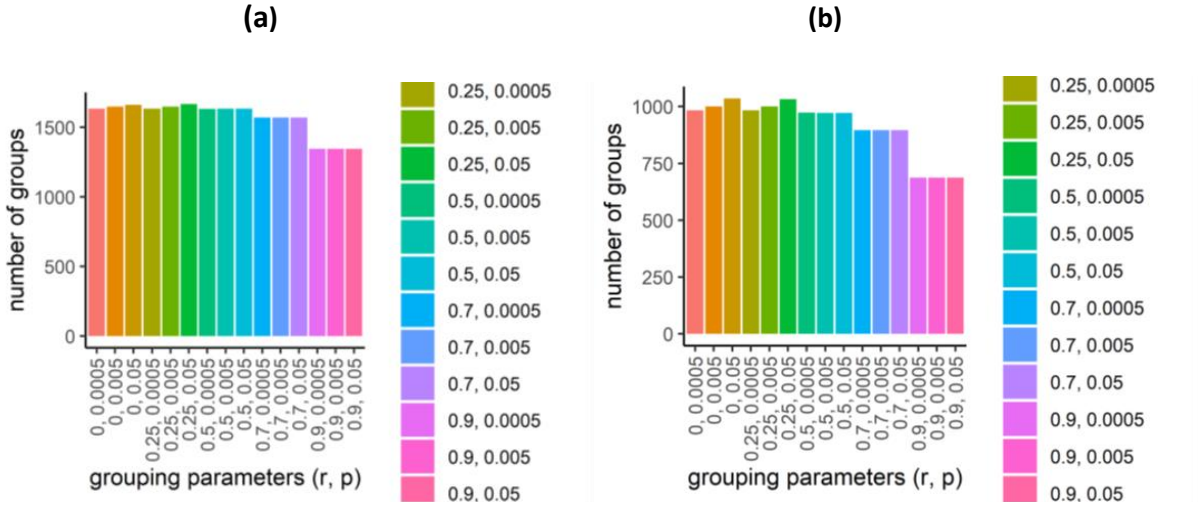


Figure 50: For the HILIC positive ion-mode (a), the number of RT groups formed is ~1,800 and does not change much for correlation thresholds ranging from 0 to 0.5. However, a minimal increase in the number of RT groups formed is observed as larger p-value thresholds are applied. The number of RT groups formed decreases to 1 600 when a correlation threshold of 0.7 is applied, and further still to 1 400 when a correlation threshold of 0.9 is applied. For both these larger correlation thresholds (0.7 and 0.9), the p value thresholds tested (0.05, 0.005, and 0.0005) did not affect the number of RT groups formed. The highest number of groups therefore resulted from correlation thresholds anywhere between 0 and 0.5, and a p value threshold of 0.05.

For the HILIC negative ion-mode (b), the number of RT groups formed was ~1 000 and did not change much for correlation thresholds ranging from 0 to 0.25. However, an increase in the number of RT groups formed was observed as larger p-value thresholds were applied. The number of RT groups formed decreased to 900 when a correlation threshold of 0.5 was applied, and further still to 850 and then 700 when correlation thresholds of 0.7 and 0.9 were applied, respectively. For these larger correlation thresholds (0.5, 0.7, and 0.9), the p value thresholds tested (0.05, 0.005, and 0.0005) did not affect the number of RT groups formed. The highest number of groups therefore resulted from correlation thresholds anywhere between 0 and 0.25, and p value thresholds of 0.05.

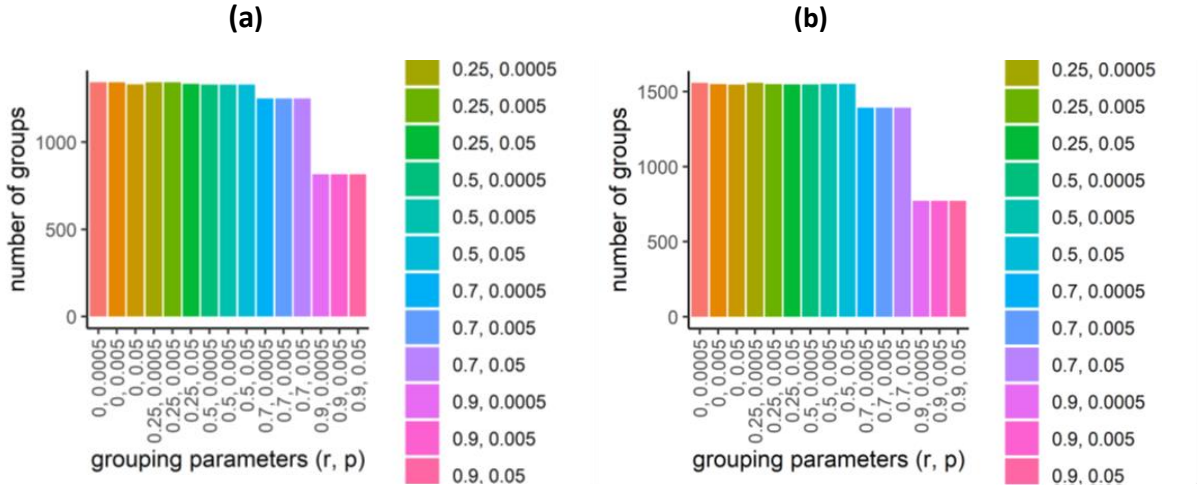


Figure 51: For the Lipids positive ion-mode (a), the number of RT groups formed was ~1 500 and did not change much for correlation thresholds ranging from 0 to 0.5. The p value threshold setting in this range also did not impact the number of RT groups formed, remaining constant at ~ 1 500 groups. The number of RT groups formed decreased to 1 300 when a correlation threshold of 0.7 was applied, and further still to 800 when a correlation threshold of 0.9 was applied. For both these larger correlation thresholds (0.7 and 0.9), the p value thresholds tested (0.05, 0.005, and 0.0005) did not affect the number of RT groups formed. The highest number of groups therefore resulted from correlation thresholds anywhere between 0 and 0.5, and a p value threshold of 0.05.

For the Lipids negative ion-mode (b), the number of RT groups formed was ~1 550 and did not change much for correlation thresholds ranging from 0 to 0.5. The p value threshold setting in this range also did not impact the number of RT groups formed, remaining constant at ~ 1 550 groups. The number of RT groups formed decreased to 1 400 when a correlation threshold of 0.7 was applied, and further still to 700 when a correlation threshold of 0.9 was applied. For both these larger correlation thresholds (0.7 and 0.9), the p value thresholds tested (0.05, 0.005, and 0.0005) did not affect the number of RT groups formed. The highest number of groups therefore resulted from correlation thresholds anywhere between 0 and 0.5, and a p value threshold of 0.05.

5.2.3.3 Retention time groups annotated

The number of RT groups with at least one annotation when applying different correlation and p-value thresholds was calculated. A RT group comprised of *m/z* features with RTs within 1s of one another, as was found to be optimal in section 5.2.1. This calculation informed on whether the correlation and p-value thresholds affected the number of RT groups which include one or more metabolite annotations.

Results for the HILIC and Lipids positive and negative ion mode assays, respectively, are shown in Figure 52 and Figure 53. For all assays, a general trend was observed. As the correlation threshold increased, the percentage of RT groups annotated decreased. N.B. that this may have been due to less RT groups formed with increasing correlation thresholds. The p value thresholds impacted the number of groups annotated at lower correlation thresholds (0 and 0.25) and in the negative ion-mode, although any changes were observed were only minor. Where p value thresholds impacted results, only small increases were observed as the p value threshold decreased, and ultimately the changes in counts were never >100 when compared for the same correlation coefficient. Based on these results, a correlation threshold between 0 and 0.5 would be optimal. Although not yielding the largest percentage of the number of RT groups annotated, these lower correlation thresholds resulted in more RT groups being formed, and thus could be more suitable. This result brings up an interesting point of discussion. Although the highest number of RT groups annotated occurs at correlation thresholds between 0-0.5 for all assays, earlier investigations into an optimal maxRT revealed that the number of unique *m/z* grouped into a single RT group based on RT similarity did not increase in any meaningful way for data within the interquartile range for all four assays. This demonstrates that grouping of degenerate features of a single metabolite relies heavily on correlation analysis. Indeed, without correlation analysis, and using only similarity in RT to group degenerate features, issues can arise resulting from co-eluting analytes. To minimise such issues, a 0.5 correlation threshold is recommended over 0 and 0.25.

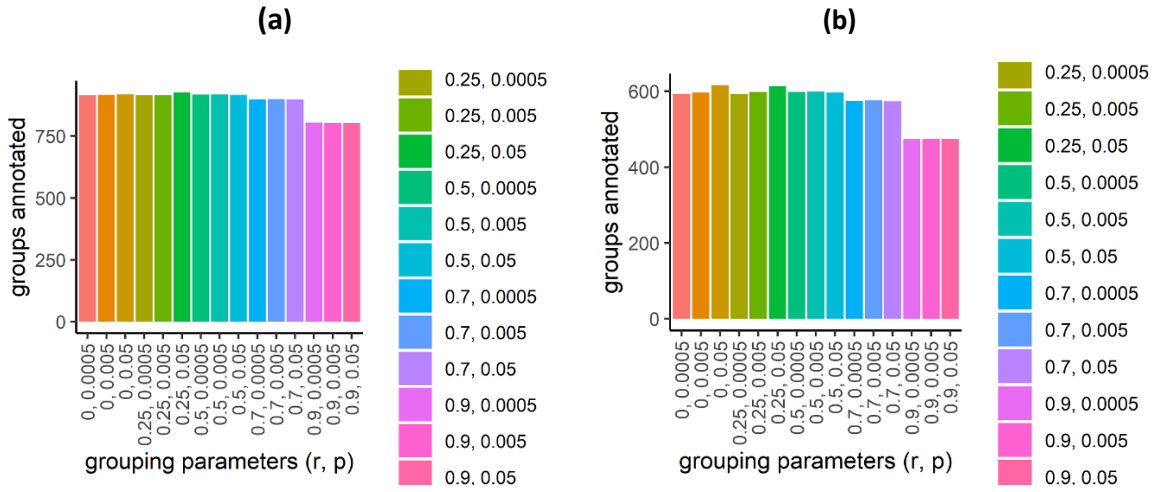


Figure 52: For the HILIC positive ion-mode dataset (a), ~900 RT groups were annotated out of the ~1 800 RT groups formed (~50%) for correlation thresholds of 0 to 0.5. The p value thresholds applied (0.05, 0.005, and 0.0005) did not have an impact on the number of RT groups annotated. Application of a correlation threshold of 0.7 resulted in the annotation of 900 out of a total of 1 600 RT groups (56%). Once again, the p value thresholds applied had no impact on the number of RT groups annotated. Finally, application of a correlation threshold of 0.9 resulted in the annotation of 850 out of a total of 1 400 RT groups (61%), and this number was not impacted by varying p value thresholds.

For the HILIC negative ion-mode dataset (b), ~600 RT groups were annotated out of the ~1 000 RT groups formed (~60%) for correlation thresholds of 0 to 0.25. The p value thresholds applied (0.05, 0.005, and 0.0005) yielded the highest numbers of RT groups annotated for 0.05. Application of a correlation threshold of 0.5 resulted in the annotation of 580 out of a total of 900 RT groups (64%). The p value thresholds applied had no impact on the number of RT groups annotated. The application of a correlation threshold of 0.7 resulted in 850 out of 850 RT groups being annotated (100%), with p value thresholds having no impact on this annotation rate. Finally, application of a correlation threshold of 0.9 resulted in the annotation of 700 out of a total of 700 RT groups (100%), and this number was also not impacted by varying p value thresholds.

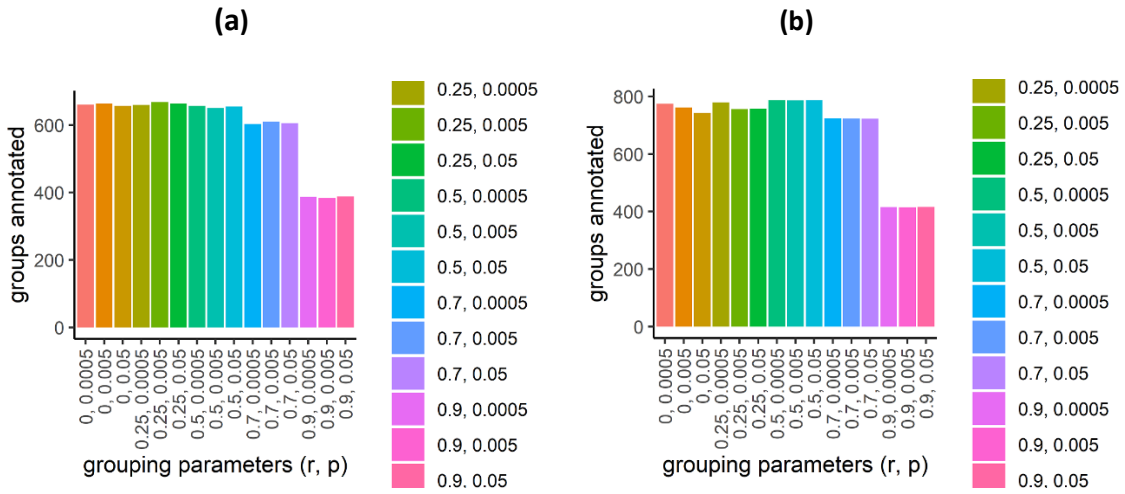


Figure 53: For the Lipids positive ion-mode dataset (a), ~650 RT groups were annotated out of the ~1 500 RT groups formed (~43%) for correlation thresholds of 0 to 0.5. The p value thresholds applied (0.05, 0.005, and 0.0005) did not have an impact on the number of RT groups annotated. Application of a correlation threshold of 0.7 resulted in the annotation of 600 out of a total of 1 300 RT groups (46%). Once again, the p value thresholds applied had no impact on the number of RT groups annotated. Finally, application of a correlation threshold of 0.9 resulted in the annotation of 380 out of a total of 800 RT groups (48%), and this number was not impacted by varying p value thresholds.

For the Lipids negative ion-mode dataset (b), ~780 RT groups were annotated out of the ~1 550 RT groups formed (~50%) for correlation thresholds of 0 to 0.5. The p value thresholds applied (0.05, 0.005, and 0.0005) for correlation thresholds of 0 and 0.25 resulted in a small increase in the number of annotated RT groups, with the highest number observed at p value threshold = 0.0005. For the correlation threshold of 0.5, the p value thresholds applied did not have an impact on the number of RT groups annotated. Application of a correlation threshold of 0.7 resulted in the annotation of 780 out of a total of 1 400 RT groups (56%). Once again, the p value thresholds applied had no impact on the number of RT groups annotated. Finally, application of a correlation threshold of 0.9 resulted in the annotation of 400 out of a total of 700 RT groups (57%), and this number was not impacted by varying p value thresholds.

5.2.3.4 Unique annotations per retention time group

The number of unique annotations per RT group when applying different correlation and p-value thresholds was calculated. This calculation informed on whether the different correlation and p-value threshold parameters affected the number of annotations in each RT group.

Results for HILIC positive and negative ion mode datasets are shown in Figure 54 and Figure 55. The maximum number of annotations per RT group were between correlation thresholds and 0 and 0.5 (~300 unique annotations per RT group). A threshold of 0.7 yielded a maximum number of annotations per group of 180, whilst this number was reduced to 120 when a correlation threshold of 0.9 was applied.

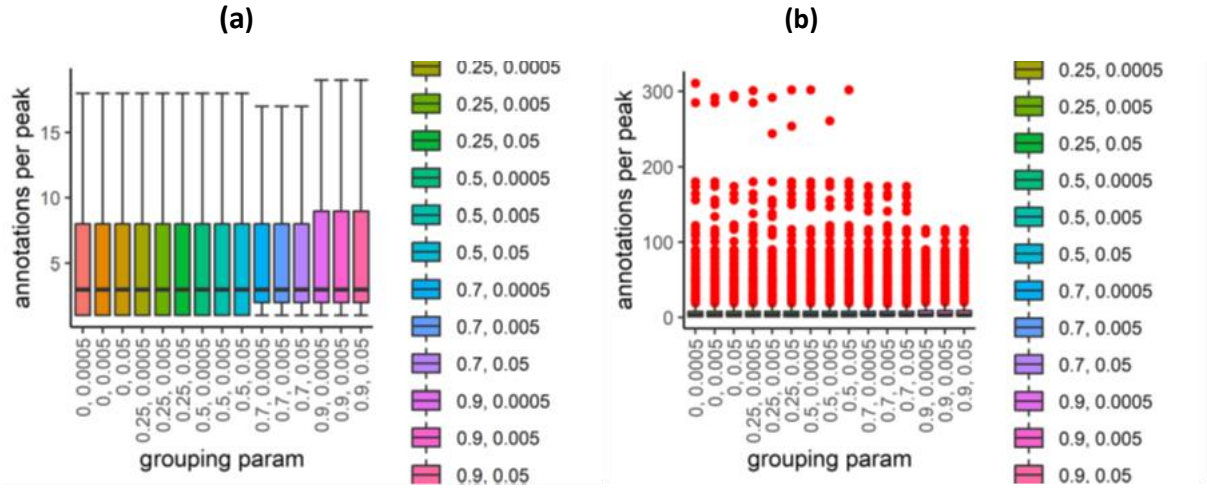


Figure 54: For the HILIC positive ion-mode, the median number of annotations for correlation thresholds of 0-0.5 for all three p value thresholds was three and did not change. The interquartile ranges 5 also did not change, ranging from 1-8, and were unaffected by varying p value thresholds. For the correlation threshold of 0.7, the interquartile range was 2-5, and was unaffected by varying p value thresholds. For the correlation threshold of 0.9, the interquartile range was 1-6, and was unaffected by varying p value thresholds. The spread of the number of annotations per RT group decreased as higher correlation thresholds were applied, but these values were not impacted by changing p value thresholds. The maximum number of annotations per RT group were obtained between correlation thresholds and 0 and 0.5 (~300 unique annotations per RT group). A threshold of 0.7 yielded a maximum number of annotations per group of 180, whilst this number was reduced to 120 when a correlation threshold of 0.9 was applied.

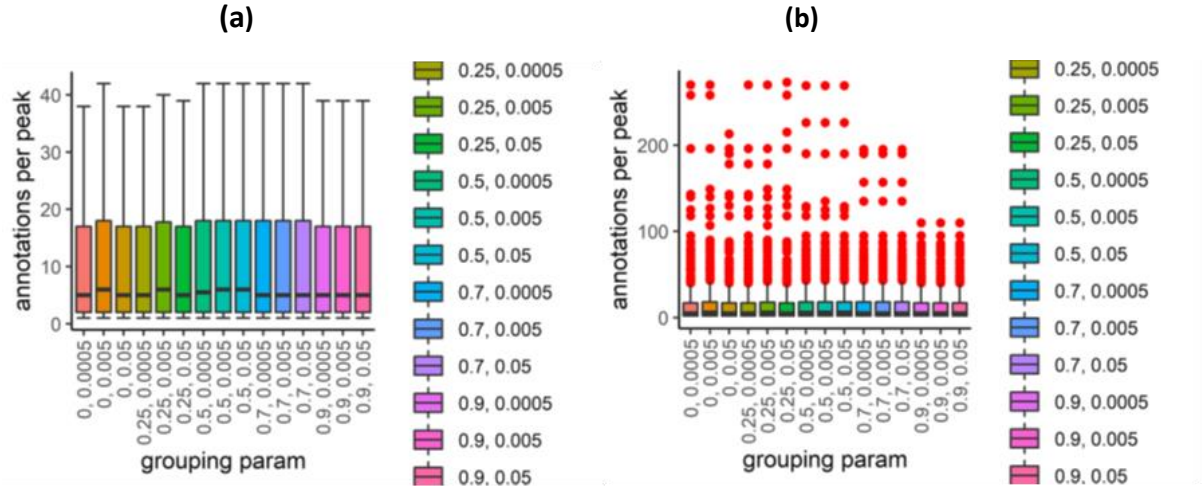


Figure 55: For the HILIC negative ion-mode, and for correlation thresholds of 0 and 0.25, the median numbers of annotations per RT group were four, five, and four for p value thresholds of 0.0005, 0.005, and 0.05, respectively. For the correlation threshold of 0.5, the median numbers of annotations per RT group were five, six, and six for p value thresholds of 0.0005, 0.005, and 0.05, respectively. For correlations of 0.7 and 0.9, the median number of annotations per RT group remained the same at four. The interquartile ranges for correlation thresholds from 0 to 0.25 ranged from 2-17, 2-18, and 2-17 for p value thresholds of 0.0005, 0.005, and 0.05, respectively. The interquartile range for the correlation thresholds of 0.5 and 0.7 were 2-18, whilst for 0.9 these were 2-17. The maximum number of annotations for correlation thresholds of 0 and 0.5 were ~300 unique annotations per RT group. A threshold of 0.7 yielded a maximum number of annotations of 180, whilst this number was reduced to 120 when a correlation threshold of 0.9 was applied.

For the Lipids positive ion-mode dataset, the median number of annotations for each tested correlation and p value threshold varied slightly. For correlation thresholds of 0, the median numbers of annotations per RT group were ten, ten, and eight for p value thresholds of 0.0005, 0.005, and 0.05, respectively. For the correlation threshold of 0.25, these were ten, nine, and, and nine for p value thresholds of 0.0005, 0.005, and 0.05, respectively. For the correlation threshold of 0.5, the median numbers of annotations per RT group were nine, ten, and nine for p value thresholds of 0.0005, 0.005, and 0.05, respectively. For correlations of 0.7 the number of annotations per RT group were ten, nine, and ten for p value thresholds of 0.0005, 0.005, and 0.05, respectively. For the correlation threshold of 0.9, the numbers of annotations per RT group were ten, ten, and nine for p value thresholds of 0.0005, 0.005, and 0.05, respectively. The interquartile ranges for correlation threshold of 0 were 2-29 for the p value threshold of 0.0005 and 3-31 for p value thresholds of 0.005 and 0.05. This interquartile range was the same for the correlation threshold of 0.25. For the 0.5 correlation threshold,

the interquartile range is from 2-29, whilst for the 0.7 correlation threshold it is 3-29, 2-27, and 3-29 for p value thresholds of 0.005 and 0.05, respectively. For correlation threshold 0.9, the interquartile range is from 3-29 for all applied p value thresholds. The maximum number of annotations per RT group is highest for correlations of 0 and 0.25, both of which have 320 annotations for all applied p-value thresholds, with minimal changes resulting from changing the p value thresholds. This value dropped slightly when a correlation threshold of 0.5 and 0.7 was applied, yielding maximum numbers of annotations of 300. These drop once more to 250 when a correlation threshold of 0.9 is applied (Figure 56).

For the Lipids negative ion-mode dataset, the median number of annotations for each tested correlation and p value threshold remained the same at three for correlation thresholds between 0 and 0.5 and were unimpacted by varying p value thresholds. When a correlation threshold of 0.7 was applied, the median numbers of annotations per RT group were three, five, and five for p value thresholds of 0.0005, 0.005, and 0.05, respectively. When a correlation threshold of 0.9 was applied, the median numbers of annotations per group were three, five, and four for p value thresholds of 0.0005, 0.005, and 0.05, respectively.

The interquartile ranges for the correlation threshold of 0 ranged from 2-18, 2-18, and 2-20 for p value thresholds of 0.0005, 0.005, and 0.05, respectively. The interquartile range for the correlation threshold of 0.25 was 2-18, whilst for correlation thresholds 0.5 and 0.7 it was 2-17 and 2-16 respectively. For these correlation thresholds, no variation occurred in response to changing p value thresholds. For correlation threshold 0.9, the interquartile range was from 2-15, 2-15, and 2-15 for p value thresholds of 0.0005, 0.005, and 0.05, respectively. The highest maximum numbers of annotations per RT group were obtained between correlation thresholds and 0 and 0.25 (~480 unique annotations per RT group). A threshold of 0.5 yielded a maximum number of annotations per group of 300, whilst this number was reduced to ~120 when a correlation threshold of 0.7 and 0.9 was applied (Figure 57).

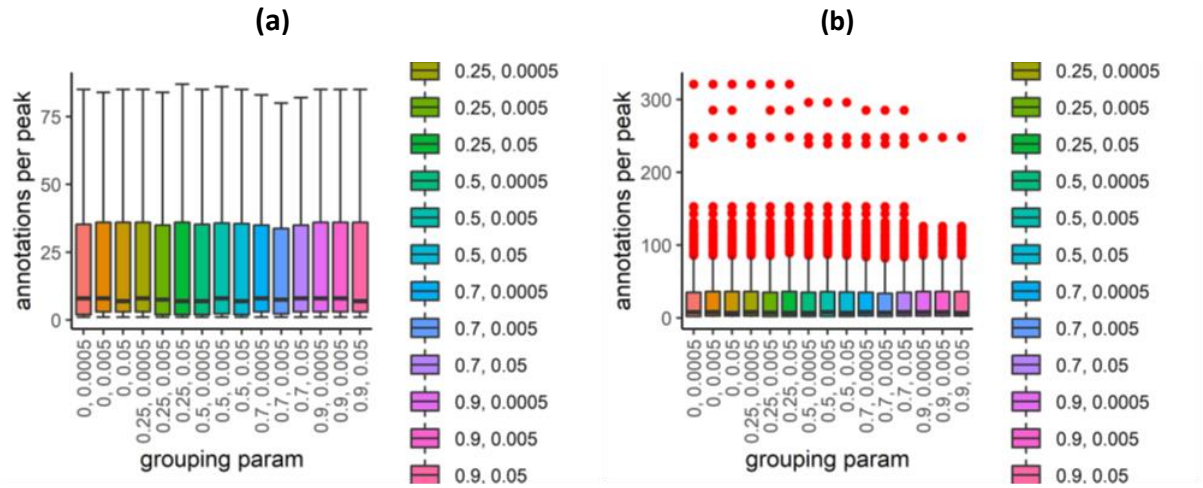


Figure 56: For the Lipids positive ion-mode, the median numbers of annotations per RT group vary depending on the varying p-value thresholds. However, it must be noted that any deviations are always small, and generally there are no major differences in the median number of annotations per RT group with changing correlation and p value thresholds. The median number of annotations for all tested conditions is around 9, whilst the interquartile range is approximately 2-31 for all tested parameters. The spread of the numbers of annotations per RT group were more informative, with the highest maximum numbers obtained for the 0 and 0.25 correlation thresholds at 320. This value dropped to 300 for the 0.5 and 0.7 correlation thresholds, and further to 250 for a correlation threshold of 0.9.

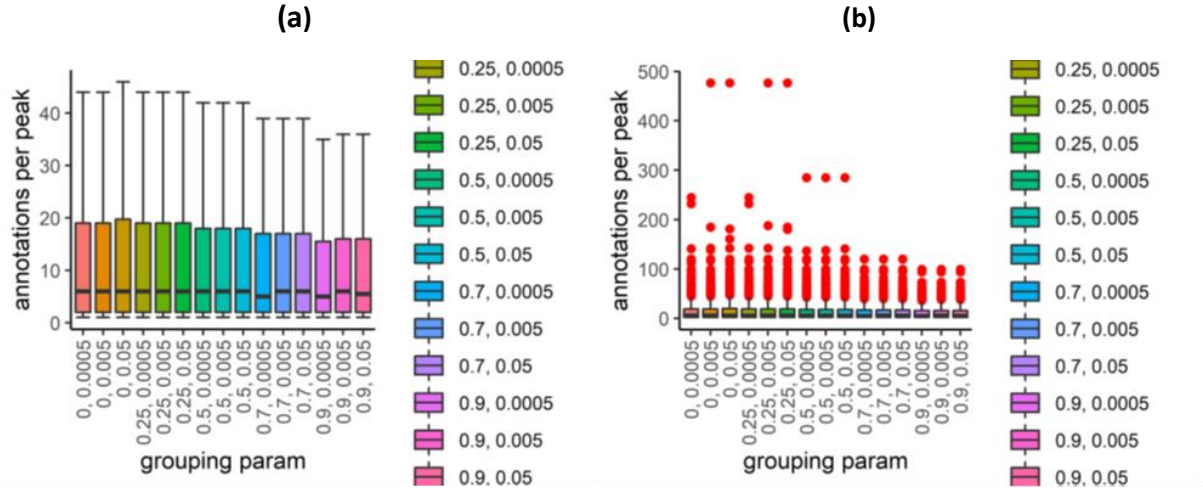


Figure 57: For the Lipids negative ion-mode, the median numbers of annotations per RT group varied depending on the varying p-value thresholds. However, it must be noted that any deviations were always small, and generally there were no major differences in the median number of annotations per RT group with changing correlation and p value thresholds. The median number of annotations for all tested conditions was around 5, whilst the interquartile range was approximately 2-18 for all tested parameters. The spread of the numbers of annotations per RT group were more informative, with the highest maximum numbers obtained for the 0 and 0.25 correlation thresholds at ~480. This value dropped to 300 for the 0.5 and 0.7 correlation thresholds, and further to 120 for a correlation threshold of 0.9.

5.2.4 Correlation Threshold and p value: Conclusion

Four U(H)PLC-MS methods, HILIC positive and negative ion-mode, and Lipids positive and negative ion-modes, were applied for the analysis of human blood serum. These data were processed using XCMS (292), and then used to test the impact of changing the correlation threshold and p-value settings during the BEAMS annotation workflow. Correlation thresholds of 0, 0.25, 0.5, 0.7, and 0.9 were investigated. For each correlation threshold, three p-value thresholds, 0.05, 0.005, and 0.0005, were also tested. The maxRT was set to 1s, as was found to be optimal in the previous section, and all other parameters were set to default.

The highest TPRs [1] were found at correlation thresholds between 0 to 0.5 for the HILIC positive and Lipids positive and negative ion-mode, whilst this range was between 0 and 0.7 for the HILIC negative ion-mode dataset. The lowest FPRs were found at correlation thresholds of 0.7-0.9 for the HILIC negative and Lipids positive and negative ion-mode datasets, whilst for the HILIC positive ion-mode dataset the lowest FPR was found between correlation thresholds of 0-0.5. Varying p value thresholds did not have an impact on TPR, meaning the significance threshold of the correlations did not vary much and thus had no impact on the results. Many a word of caution has been given about the relationship between p value and sample size, and this must be noted here. Although the datasets used in this chapter had >100 samples (including QC samples), which is well above the minimum of 10 samples recommended for a good p value calculation, some researchers recommend sample sizes as high as 250 for better accuracy (444), but such large-scale studies come with their own challenges, including batch to batch variation and subsequent challenges in dealing with this variation (445,446). It must also be noted that Pearson correlation assumes linearity and does not give a good p value estimate for non-linear relationships (447). Although Spearman's rank performs better for non-linear relationships and is more rugged towards outliers, it relies even more heavily on a larger sample size than Pearson's correlation analysis (447). It is possible that most results presented here were not impacted/were only impacted negligibly by varying p value thresholds because the sample sizes (2017 participants were included in this study but only ~100 samples were selected to increase computational efficiency) were insufficiently large and thus p values calculated were inaccurate. There are tools available that can be used to estimate appropriate sample sizes for metabolomics datasets that consider intended data analysis techniques (448).

Based on TPR and FPRs of annotation presented here, an optimal correlation threshold must be between 0 and 0.5 for these datasets. The highest number of RT groups formed [2] was also found between correlation thresholds of 0 and 0.5 for the HILIC positive and Lipids positive and negative ion-modes. For the HILIC negative ion-mode dataset, the largest number of RT groups was found between correlation thresholds of 0 to 0.25. Once again, p value thresholds did not impact results in any meaningful way, hence a correlation threshold of 0.5 would be optimal. The number of RT groups annotated [3] increased when higher correlation thresholds were used. However, using higher correlation thresholds was shown to decrease the number of groups formed, and most importantly the TPR. As such, a lower correlation threshold remained in favour. The median number of annotations per RT group and the interquartile ranges for each correlation and p value threshold demonstrated no clear and large differences. However, the distribution of the outlying data demonstrated that lower correlation thresholds yielded more annotations per RT group. The optimal correlation threshold was selected as 0.5 based on the results presented herein, and a p value threshold of 0.05 was also selected as optimal. It could be considered to use different parameters for the different ion-modes, but this introduces more computational tax to an already extensive workflow, thus applying the same parameters for both ion-modes is more time efficient. Moreover, although the difference between methods is not huge, it is still sensible to change parameters for each assay to improve the accuracy of results, even if only moderately.

5.2.5 Mass Error

Four U(H)PLC-MS methods, HILIC positive and negative ion-mode, and lipids positive and negative ion-modes, were applied for the analysis of human blood serum. These data were processed using XCMS (292), and then used to test the impact of the mass error setting during the BEAMS annotation workflow. Mass errors of 0.5, 1.0, 3.0, 5.0, 10.0, and 25.0 ppm were investigated. Although accurate mass instruments were used, it is still worth considering large mass errors as true instrument accuracy cannot be known in NTA in the absence of internal standards. The maxRT was set to 1s, and the correlation and p value thresholds were set to 0.5 and 0.05, respectively, as was found to be optimal in sections 5.2.1 and 5.2.3. All other parameters were set as default and only the mass error was varied. Data analysis of the results considered the following:

1. How was the TP and FP rate of metabolites identified using MS/MS and RT matching to data for authentic chemical standard affected by changing mass error?
2. How many RT groups were annotated with varying mass error?
3. How many annotations were there for each RT group with varying mass error?

5.2.5.1 True and false positive rate

A subset of confidently identified metabolites (described in 2.4.5), considered to be TPs were used to investigate the impact of varying the mass error tolerance. TPRs and FPRs were calculated as described in 2.4.5.This calculation informed on whether the mass error threshold affected annotation during the BEAMS workflow and would facilitate selection of an appropriate mass error tolerance for datasets collected with the ACMG.

TPR and FPR results for the HILIC and Lipids positive and negative ion-mode assays, respectively, are shown in Figure 58 to Figure 61. For all four assays, the highest TPRs were found at mass error tolerances between 3-10ppm. For both HILIC assays, the FPR increased with widening mass error tolerances. However, the rate of increase was smaller for mass error tolerances between 3-10ppm, meaning these mass error tolerances were not introducing large amounts of false annotations. Complementary to the mass error tolerances yielding the highest TPRs, a mass error tolerance of 5ppm was selected as optimal. This seemed appropriate considering that the mass error of the Q Exactive was 3ppm with external calibration. For both Lipids assays the smallest increase in FPRs was observed between 3-10ppm. Complementary to the TPR, a 5ppm mass error tolerance would be optimal for all four assays.

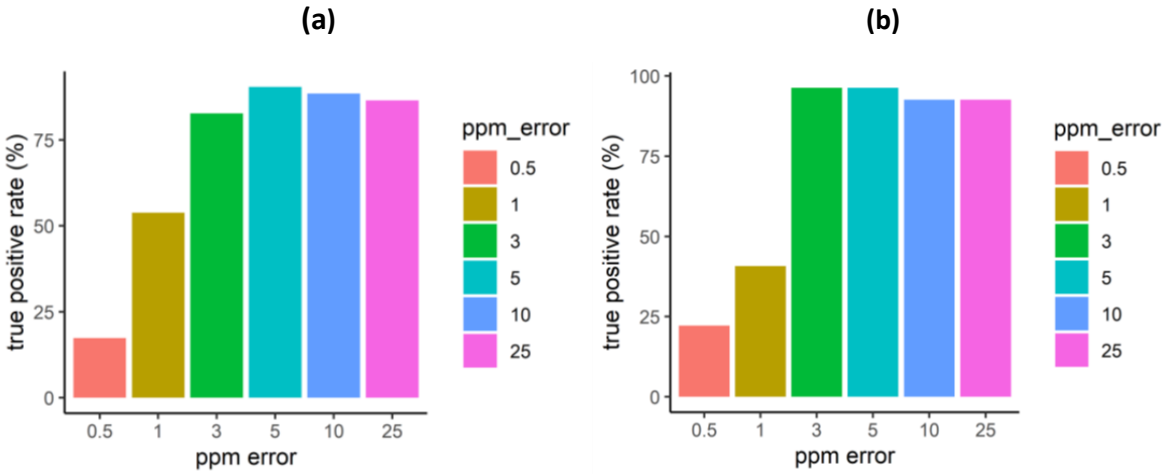


Figure 58: For the HILIC positive ion-mode dataset (a), the lowest TPR was 18% and was observed when a mass error tolerance of 0.5ppm was applied. The highest TPR was 96% and was observed at a mass error tolerance of 5ppm. The TPR decreased when mass error tolerances of 10 ad 25ppm were applied, thus a 5ppm error tolerance was selected to be optimal, maximising the TPR within the dataset. For the HILIC negative ion-mode dataset (b), the lowest TPR was 22% and was observed when a mass error tolerance of 0.5ppm was applied. The highest TPR was 96% and was observed at mass error tolerances of 3 and 5ppm. The TPR decreased when mass error tolerances of 10 ad 25ppm were applied, thus a 5ppm error tolerance was selected to be optimal, maximising the TPR within the dataset.

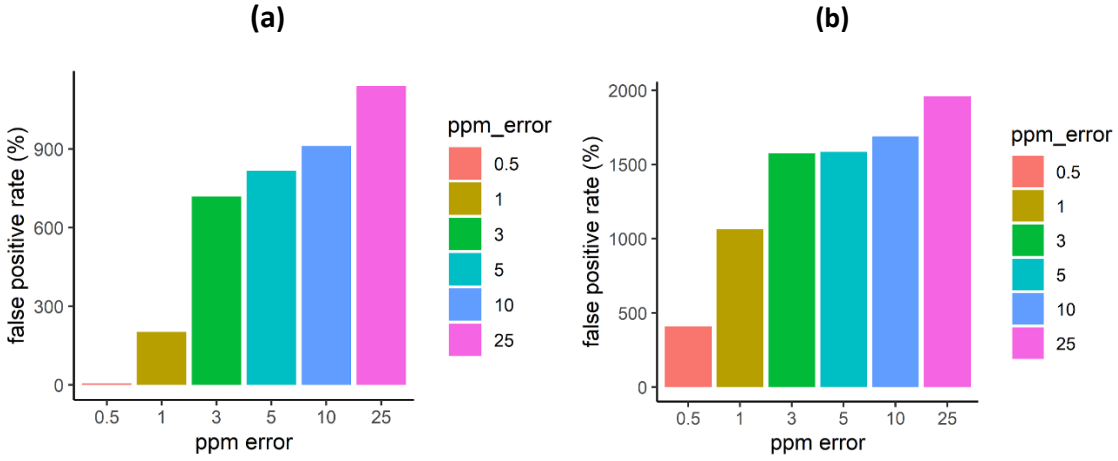


Figure 59: For both HILIC assays (positive ion-mode (a), negative ion-mode (b)), the FPR increased with widening mass error tolerances. The lowest FPR was observed when a mass error tolerance of 0.5ppm was applied, and the highest obtained when a mass error tolerance of 25ppm was applied. The rise in the FPR was steep and large between 0.5-3ppm, and between 10-25ppm. However, between 3-10ppm, the rise in FPR was much lower, and thus mass error tolerances in this region did not introduce a very large amount of false positive annotation. As such, complementary to the TPR, a mass error tolerance of 5ppm was selected as optimal.

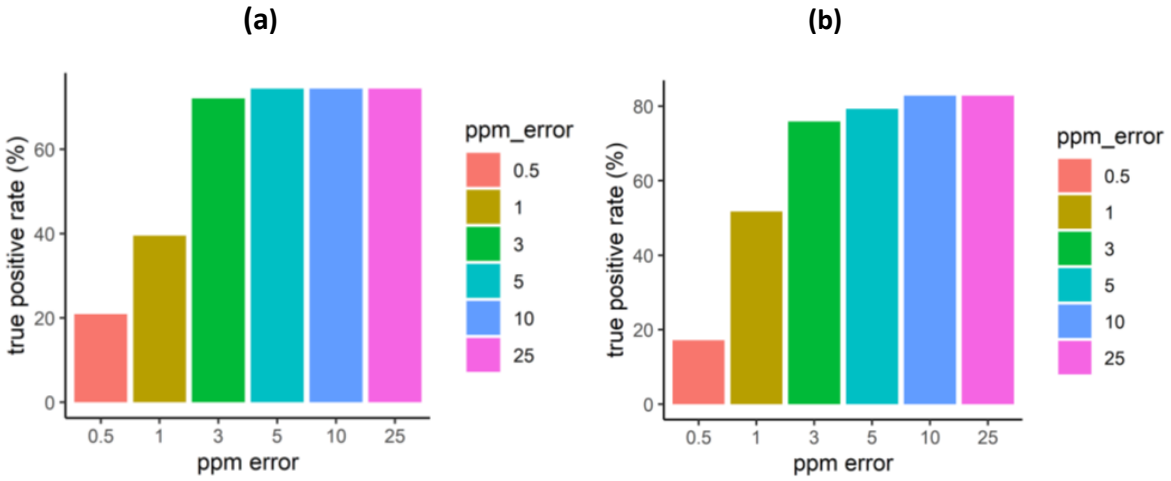


Figure 60: For the Lipids positive ion-mode dataset (a), the lowest TPR was 23% and was observed when a mass error tolerance of 0.5ppm was applied. The highest TPR was 84% and was observed at mass error tolerances of 5-25ppm. For the Lipids negative ion-mode dataset (b), the lowest TPR was 17% and was observed when a mass error tolerance of 0.5ppm was applied. The highest TPR was 83% and was observed at mass error tolerances of 10-25ppm.

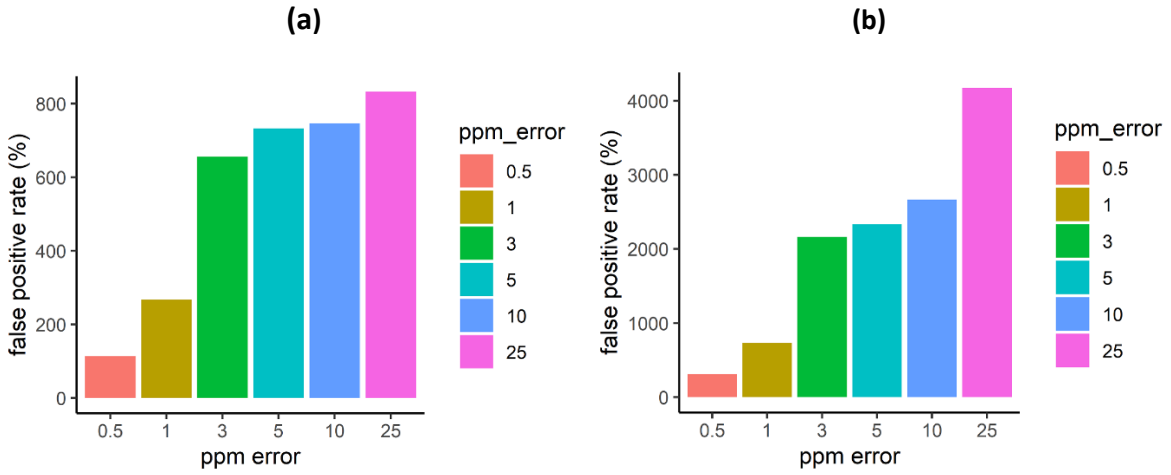


Figure 61: For both Lipids assays (positive ion-mode (a), negative ion-mode (b)), the FPR increased with widening mass error tolerances. The lowest FPR was observed when a mass error tolerance of 0.5ppm was applied, and the highest obtained when a mass error tolerance of 25ppm was applied. The rise in the FPR was steep and large between 0.5-3ppm, and between 10-25ppm. However, between 3-10ppm, the rise in FPR was much lower, and thus mass error tolerances in this region did not introduce a very large amount of false positive annotation. As such, complementary to the TPR, a mass error tolerance of 5ppm was selected as optimal.

5.2.5.2 Retention time groups annotated

The number of RT groups with at least one annotation was calculated for each of 6 mass error tolerances (0.5, 1, 3, 5, 10, and 25ppm). This was done for the 4 U(H)PLC-MS datasets, HILIC positive and negative ion-mode, and Lipids positive and negative ion-modes. This calculation demonstrated the impact of using increasing mass error tolerances on the number of RT groups annotated.

Results for the HILIC and Lipids positive and negative ion mode datasets, respectively, are shown in Figure 62 and Figure 63. This increase in annotation with increasing mass error tolerance undoubtedly also coincides with an increase in FP annotation, and a compromise is required to maximise annotations but considering the increase in the FPR. The smallest increases in the number of RT groups annotated occurred between 3-10ppm for the Lipids positive mode dataset, and 3-25ppm for the Lipids negative ion-mode dataset. Thus, an optimal mass error tolerance could be selected anywhere from these ranges.

For all four assays, increasing the mass error tolerance increased the number of RT groups annotated. Testing these mass error tolerances was required without additional information about the actual mass error achieved for the datasets collected in NTA. In particular, this is important for low intensity features, whose mass accuracies can be impacted by imprecise apex selection. These low intensity features could be crucial information, as such just filtering them out is not a reasonable approach. Such features can increase the mass accuracy of the raw data irrespective of instrument specifications, thus evaluating wider mass errors was necessary. This increase in annotation undoubtedly coincides with an increase in false positive annotation, and a compromise is required to maximise annotations but considering the increase in the FPR. Although the highest number of RT groups annotated was found at a mass error tolerance of 25ppm, results in section 5.2.5.1 show that applying a mass error tolerance of 25ppm yielded the highest FPR. As such, more RT groups were annotated at this mass error tolerance, but a large proportion of them incorrectly so.

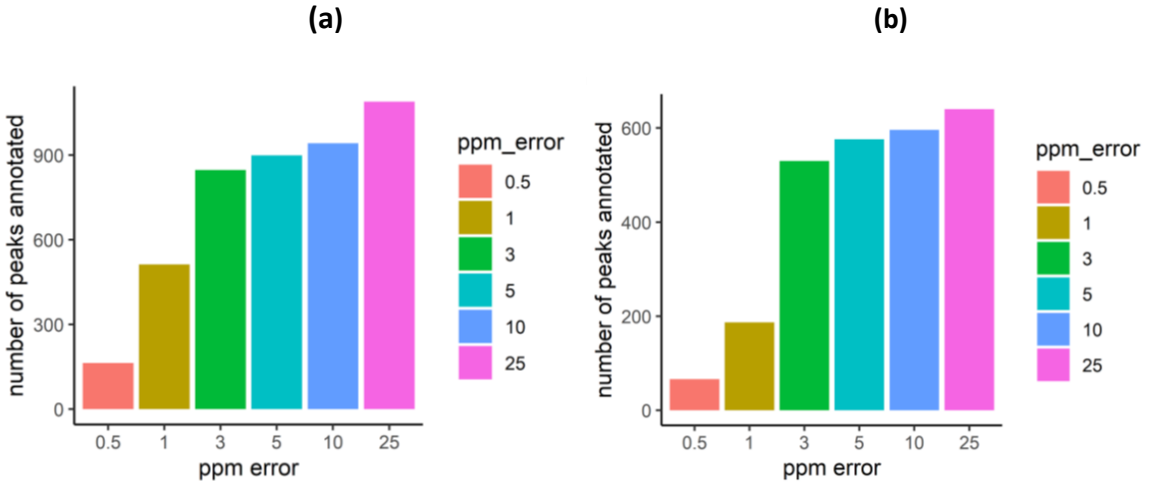


Figure 62: For the HILIC positive ion-mode dataset, the smallest number of RT groups annotated was 150 and was observed when a mass error tolerance of 0.5ppm was applied, whilst the highest number of annotated RT groups was 1 200 and occurred when a mass error tolerance of 25ppm was applied. For the HILIC negative ion-mode dataset, the smallest number of RT groups annotated was 50 and was observed when a mass error tolerance of 0.5ppm was applied, whilst the highest number of annotated RT groups was 650 and occurred when a mass error tolerance of 25ppm was applied.

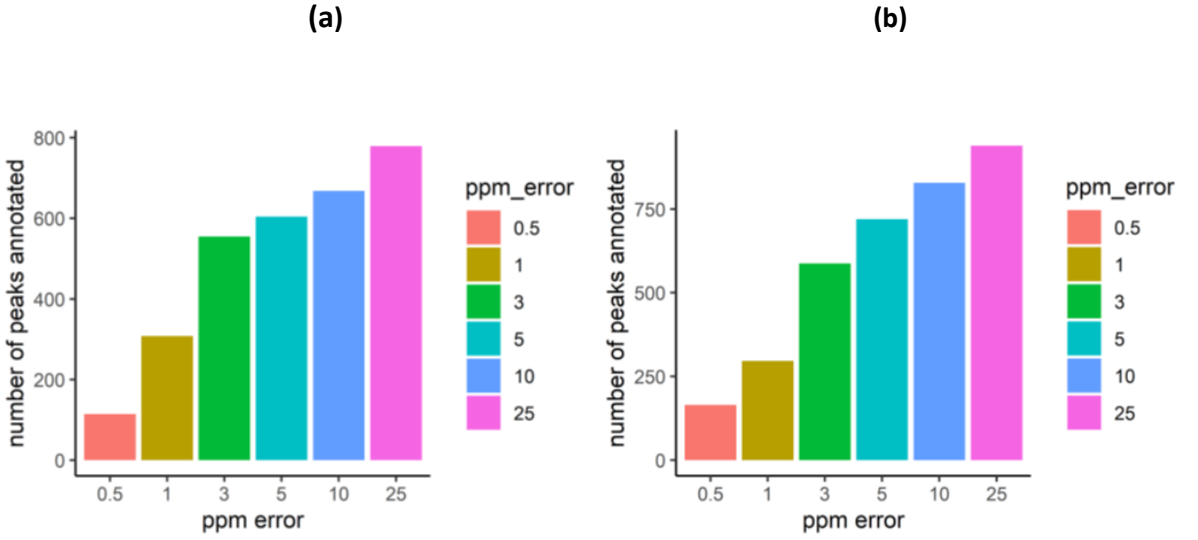


Figure 63: For the Lipids positive ion-mode dataset, the smallest number of RT groups annotated was 100 and was observed when a mass error tolerance of 0.5ppm was applied, whilst the highest number of annotated RT groups was 800 and occurred when a mass error tolerance of 25ppm was applied. For the Lipids negative ion-mode dataset, the smallest number of RT groups annotated was 180 and was observed when a mass error tolerance of 0.5ppm was applied, whilst the highest number of annotated RT groups was 900 and occurred when a mass error tolerance of 25ppm was applied.

5.2.5.3 Annotations per peak

The number of annotations per RT group was calculated for six mass error tolerances (0.5, 1.0, 3.0, 5.0, 10.0, and 25.0ppm). This was done for the 4 U(H)PLC-MS datasets, HILIC positive and negative ion-mode, and Lipids positive and negative ion-modes. This calculation informed on whether the mass error tolerance parameter affected the number of annotations per RT group.

Results for the HILIC and Lipids positive and negative ion mode datasets, respectively, are shown in Figure 64 and Figure 65. Increasing the mass error tolerance increased the median number of annotations and the interquartile range of annotations per RT group, with the highest median and interquartile range observed at 25ppm. However, as described in 5.2.5.1, a mass error tolerance of 25ppm yields the highest true positive rates. Minimal changes in the annotations per RT group were observed between 1-5ppm, as such an optimal mass error tolerance can be selected for any of those mass errors.

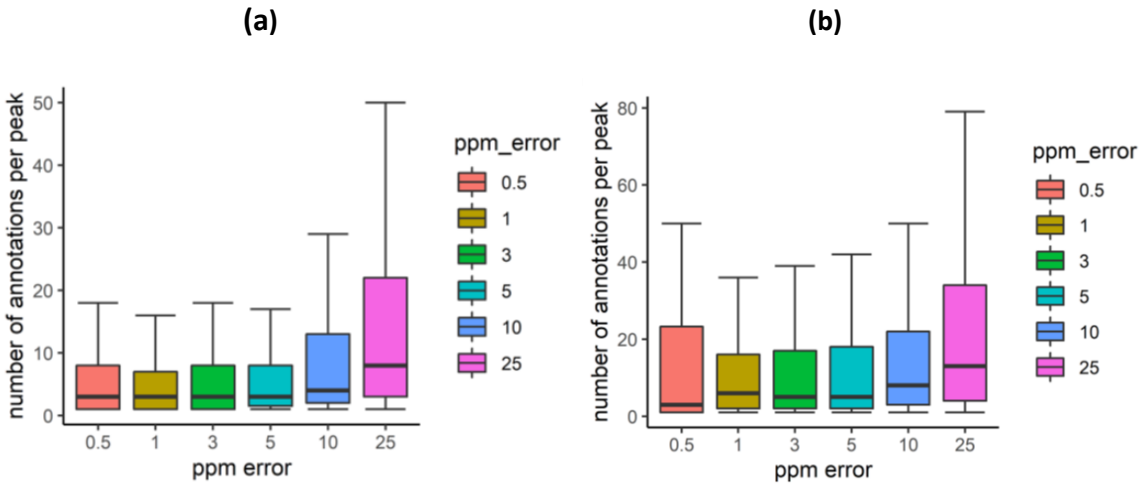


Figure 64: For the HILIC positive ion-mode dataset (a), the median number of annotations per RT group for mass error tolerances of 0.5-5ppm was three. This increased to four and eight when mass error tolerances of 10 and 25ppm were applied, respectively. The interquartile ranges were 1-8, 1-7, 1-8, and 1-8 for mass error tolerances of 0.5, 1, 3, and 5, respectively. These increased to 2-13 when a mass error tolerance of 10ppm was applied, and further to 3-22 when a mass error tolerance of 25ppm was applied. For the HILIC negative ion-mode dataset (b), the median number of annotations per RT group for the 0.5ppm mass error tolerance was two. This increased to seven when a mass error tolerance of 1ppm was applied and then decreased slightly to six when mass error tolerances of 3 and 5ppm were applied. A median of eight and 11 was obtained when mass error tolerances of 10 and 25ppm were applied. The interquartile range for the 0.5ppm mass error tolerance was 1-22. The interquartile ranges for mass error tolerances between 1 and 5ppm were 2-16, 2-17, and 2-18 for mass error tolerances of 1, 3, and 5, respectively. The interquartile ranges for mass error tolerances of 10 and 25pmm were 3-22 and 4-33, respectively.

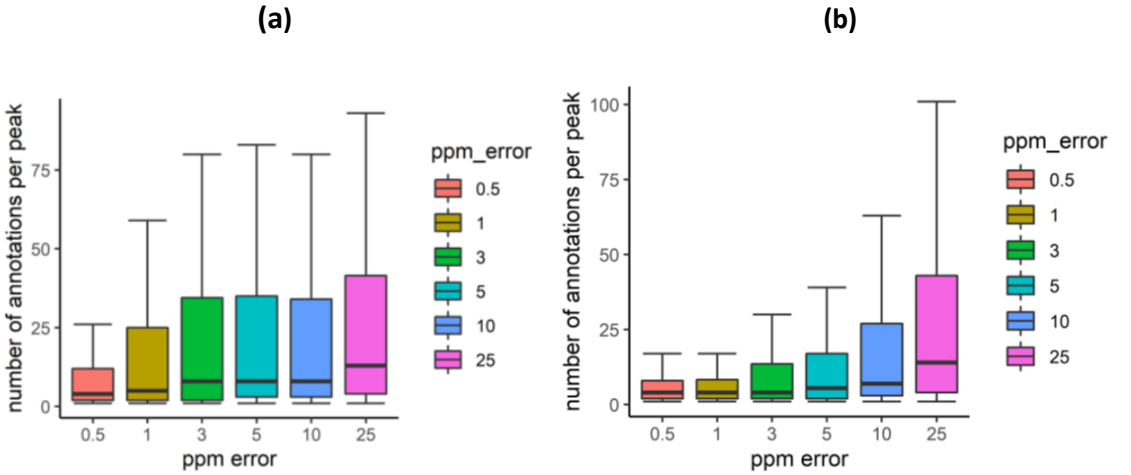


Figure 65: For the Lipids positive ion-mode dataset (a), the median number of annotations per RT group when a mass error tolerance of 0.5ppm was three. This value increased to six when a mass error tolerance of 1ppm was applied, and further to ten for mass error tolerances of 3, 5, and 10ppm, respectively. The median number of annotations per RT group was 11 when a mass error tolerance of 25ppm was applied. The interquartile ranges were 1-11, 1-25, 1-33, 3-33, 3-31, and 5-44 when mass error tolerances of 0.5, 1, 3, 5, 10, and 25ppm, respectively. For the Lipids negative ion-mode dataset (b), the median number of annotations per RT group when mass error tolerances of 0.5-3ppm were applied was five. This value increased to eight when a mass error tolerance of 5ppm was applied, and further to 10 for the mass error tolerances of 10ppm. The median number of annotations per RT group was 13 when a mass error tolerance of 25ppm was applied. The interquartile ranges were 1-11, 1-11, 1-13, 1-18, 3-27, and 5-45 when mass error tolerances of 0.5, 1, 3, 5, 10, and 25ppm were applied, respectively.

5.2.6 Mass Error: Conclusion

Four U(H)PLC-MS methods, HILIC positive and negative ion-mode, and lipids positive and negative ion-modes, were applied for the analysis of human blood serum. These data were processed using XCMS (292), and then used to test the impact of the mass error setting during the BEAMS annotation workflow. Mass errors of 0.5, 1, 3, 5, 10, and 25 ppm were investigated. The maxRT was set to 1s, and the correlation and p value thresholds were set to 0.5 and 0.05, respectively, as was found to be optimal in sections 5.2.1 and 5.2.3. All other parameters were set as default and only the mass error was varied.

The highest TPRs [1] were observed at mass error tolerances of 5ppm for the HILIC positive ion-mode dataset, and 3 and 5ppm for the HILIC negative ion-mode dataset. The highest TPRs were observed at mass error tolerances of 5, 10, and 25ppm for the Lipids positive ion-mode dataset, and 10 and 25ppm for the Lipids negative ion-mode dataset, likely due to imprecise peak apex selection as discussed in section 5.2.5.2. The biggest changes (increases) in TPRs were observed for mass error tolerances of 0.5-3ppm for all assays. Thereafter, minimal changes in TPRs occurred between 5-25ppm. Therefore, an optimal mass error tolerance could be selected anywhere between 5-25ppm. However, for the HILIC assays, a slight decline in TPR was observed when mass error tolerances of 10 and 25ppm were applied. As such, 5ppm was selected as the optimal mass error tolerance for all assays. The FPRs for all four assays increased when mass error tolerances were increased, with the lowest FPRs observed at 0.5ppm and the highest at 25pm. The increase in FPR between 0.5-3pm was steep, whilst FPRs between 3-10ppm exhibited much smaller increases as mass error tolerance increased. Thereafter, between 10-25pm, steep increases in FPRs were observed once more. With the smallest rate of increase in FPRs observed between 3-10ppm, an optimal mass error tolerance was selected as optimal, complementing the mass error tolerance yielding the highest TPR as well. The largest number of RT groups with at least one annotation [2] was observed when a mass error tolerance of 25ppm was applied for all four assays. The biggest increase in the number of RT groups annotated occurred from 0.5-3ppm for all assays. For the HILIC assays, the increase in the number of groups annotated is smallest between 3-10ppm. Thereafter, a larger increase is observed when a mass error tolerance of 25ppm is applied. This trend was similar for the Lipids positive ion-mode, with the smallest increases in the number of RT groups annotated observed when mass error tolerances of 3, 5, and 10ppm were applied.

When a mass error tolerance of 25ppm was applied, a larger jump in the number of RT groups annotated was observed. Finally, the Lipids negative ion-mode dataset generally had a constant and linear increase in the number of RT groups annotated. For three out of the four assays, the steeper increase in the number of RT groups annotated between 0.5-3ppm suggested that the lower mass error tolerances were likely too low. Changes in mass error between 3-10ppm did not result in any major increase in the number of RT groups annotated, suggesting an optimal mass error tolerance could be selected anywhere in this range. As such, a 5ppm mass error tolerance was selected as optimal, yielding many annotated RT groups, and resulting in the highest TPR. The number of annotations within each RT group [3] did not vary in any significant way for mass error tolerances between 0.5-5ppm for the HILIC positive ion-mode dataset, and 1-5ppm for the HILIC negative ion-mode dataset. For the Lipids positive ion-mode dataset, the number of annotations per RT group also did not vary significantly for mass error tolerances of 3-10ppm for the Lipids positive ion-mode dataset, and 0.5-3ppm for the Lipids negative ion-mode dataset. Therefore, for all assays, a mass error tolerance of 5ppm could be selected as optimal to coincide with the mass error tolerance yielding the highest TPR and a mid-range FPR, and this was recommended for application within the ACMG.

5.2.7 Adduct, Isotope, and Neutral Loss Lists

Four U(H)PLC-MS methods, HILIC positive and negative ion-mode, and lipids positive and negative ion-modes, were applied for the analysis of human blood serum. These data were processed using XCMS (292), and then used to test the impact of customising the adduct, isotope, and neutral loss list settings during the BEAMS annotation workflow. Default adduct, isotope, and neutral loss lists were applied first and then customised lists were created and applied as described in the methods section (2.4.3).

Data analysis of the results considered the following:

1. How many TP and FPs (%) were there using (i) the default adduct, isotope, and neutral loss lists (ii) the adduct, isotope and neutral loss lists produced by an R package created within the ACMG used to calculate what adducts, isotopes, and neutral losses are present within a dataset?

2. How many unique *m/z* features were grouped together overall applying (i) the default adduct, isotope, and neutral loss lists (ii) the adduct, isotope and neutral loss lists produced by an R package created within the ACMG used to calculate what adducts, isotopes, and neutral losses are present within a dataset?
3. How many RT groups were formed, and how many *m/z* features were in each RT group applying (i) the default adduct, isotope, and neutral loss lists (ii) the adduct, isotope and neutral loss lists produced by an R package created within the ACMG used to calculate what adducts, isotopes, and neutral losses are present within a dataset?
4. How many unique annotations were in each RT group applying (i) the default adduct, isotope, and neutral loss lists (ii) the adduct, isotope and neutral loss lists produced by an R package created within the ACMG used to calculate what adducts, isotopes, and neutral losses are present within a dataset?

5.2.7.1 True and false positive rate

A subset of confidently identified metabolites (described in 2.4.5), considered to be TPs ion- used to investigate the impact of varying the mass error tolerance. True and FPRs were calculated as described in 2.4.5. This calculation informed on whether the adduct, isotope, and neutral loss lists applied affected annotation during the BEAMS workflow and would facilitate selection of appropriate lists for datasets collected within the ACMG.

For the HILIC positive ion-mode dataset, using the customised adduct, isotope, and neutral loss lists (see 2.4.3) yielded an impressive 98% TPR, whilst the default adduct, isotope, and neutral loss lists yielded a TPR of 94% (Figure 66). However, for the HILIC negative ion-mode dataset the use of the customised adduct, isotope, and neutral loss lists resulted in a TPR of 59% whilst the default adduct, isotope, and neutral loss lists yielded a TPR of 96%. The FPRs for both HILIC assays increased from 750% to 1,600% and 1,600% to 2,200% for the HILIC positive and HILIC negative ion-modes, respectively (Figure 67).

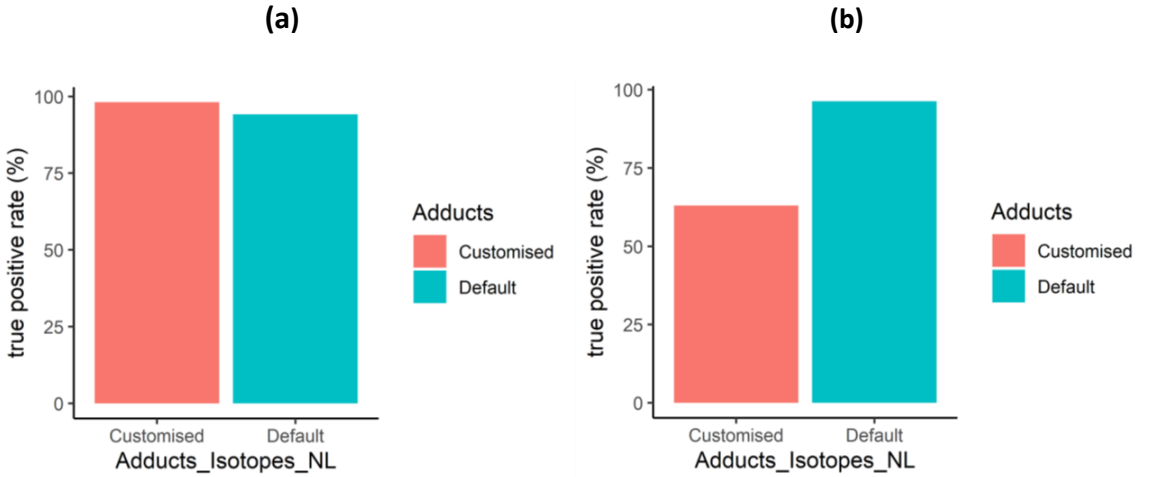


Figure 66: For the HILIC positive ion-mode dataset (a), application of customised adduct, isotope, and neutral loss lists yielded a TPR of 98%, whilst the default adduct, isotope, and neutral loss lists yielded TPRs of 94%. For the HILIC negative ion-mode dataset (b), application of customised adduct, isotope, and neutral loss lists yielded a TPR of 59%, whilst the default adduct, isotope, and neutral loss lists yielded TPRs of 96%.

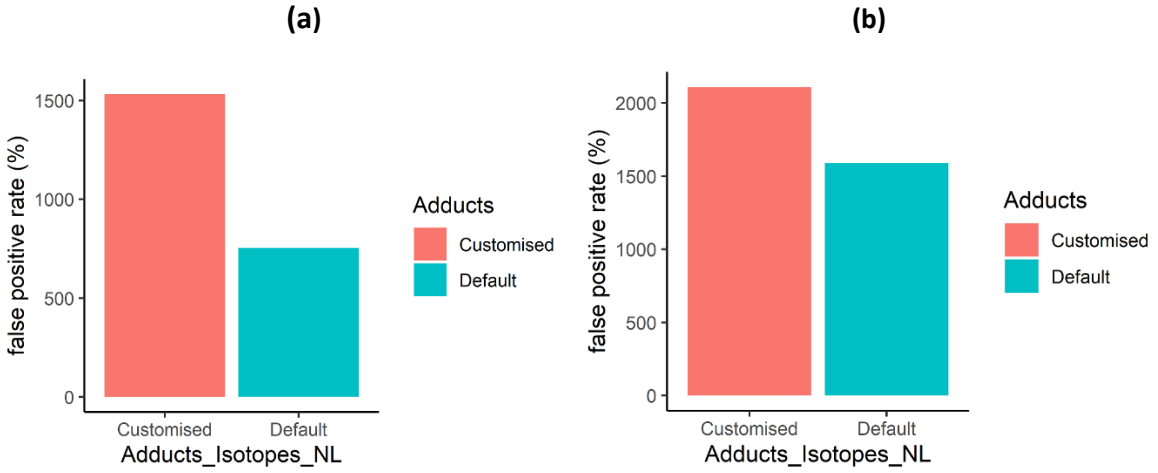


Figure 67: For the HILIC positive ion-mode dataset on the left-hand side (a), application of customised adduct, isotope, and neutral loss lists yielded a FPR of 1 600%, whilst the default adduct, isotope, and neutral loss lists yielded TPRs of 750%. For the HILIC negative ion-mode dataset on the right-hand side (b), application of customised adduct, isotope, and neutral loss lists yielded a TPR of 2 200%, whilst the default adduct, isotope, and neutral loss lists yielded TPRs of 1 600%.

The lower TPR observed when the non-default lists were applied for the HILIC negative ion-mode dataset was unexpected. The prevalence of degenerate features and the advantages of more exhaustively annotating them have been documented (322,449,450). It was therefore expected that the use of longer lists would improve TP annotation, or at the very least have

no impact. It was not expected to decrease the TPR. Data spot checking was carried out to determine why the TPR had decreased so drastically when customised lists were applied. The metabolite 3-methylxanthine (HMDB0001886) was identified through RT and MS/MS matching and was known to be present within the HILIC negative ion-mode dataset. This was manually searched for in the BEAMS outputs where default and customised adduct, isotope, and neutral loss lists were applied. In the former output, 3-methylxanthine was annotated in the [M-H] adduct form, assigned a single ^{13}C isotope, and was not assigned any neutral losses (Table 79).

name	M165T81	M166T81_2
m/z	165.0420	166.0454
RT(s)	81	81
intensity	69 431.1	8 930 783
isotope_labels_a	^{12}C	(^{13}C)
isotope_ids	M166T81_2	M165T81
isotope_labels_b	(^{13}C)	^{12}C
nl_labels		
exact_mass	165.0418	166.0452
ppm_error	1.4	1.5
rt_diff		
adduct	$[M-H]^-$	$[M-H]^-$
molecular_formula	$C_6H_6N_4O_2$	$C_6H_6N_4O_2$
compound_name	3-Methyxanthine	3-Methyxanthine
compound_id	HMDB0001886	HMDB0001886

Table 79: 3-methylxanthine (HMDB0001886) was correctly annotated when default adduct, isotope, and neutral loss lists were used as an [M-H] ion, was assigned a single 13C isotope, and was not assigned any neutral losses.

However, when the customised adduct, isotope, and neutral loss lists were applied, the same *m/z* feature (M165T81), at the same RT of 81.235145s, was not annotated as 3-methylxanthine. It was, in-fact not annotated at all, instead being assigned 3 isotopes (^{13}C,

^{15}N, and ^{17}O), and a neutral loss of H_2. Since the difference between the two sets of annotations is the additional assignment of isotopes and neutral losses, it can be posited that these impeded the annotation of 3-methylxanthine when the longer customised adduct, isotope, and neutral loss lists were applied (Table 80).

name	M165T81	M165T81
m/z	165.0420	165.0420
rt	81	81
intensity	69,431	69,431
isotope_labels_a	C, N, O	C, N, O
isotope_ids	M166T81_2, M166T81_1, M166T81_2	M166T81_2, M166T81_1, M166T81_2
isotope_labels_b	(^{13}C), (^{15}N), (^{17}O)	(^{13}C), (^{15}N), (^{17}O)
nl_labels	H_2	H_2
exact_mass		
ppm_error		
rt_diff		
adduct		
molecular_formula		
compound_name		
compound_id		

Table 80: The *m/z* feature 165.042022, at RT 81.235145s, was not annotated at all when customised adduct, isotope, and neutral loss lists were applied. The difference between the assignments here and the assignments yielded by applying the default adduct, isotope, and neutral loss lists is that when the customised lists were applied, 3 isotopes were assigned to this feature (where only one was assigned using the default lists), and a neutral loss was assigned using the customised lists whereas no neutral losses were assigned using the default lists. Since these are the only differences between the two results, it can be posited that the assignment of additional isotopes and neutral losses impeded the annotation of this peak.

Although the [M-H]- ion form of 3-methylxanthine was not annotated using the customised adduct, isotope, and neutral loss lists described above, the [M+Br]- ion form was annotated instead (Table 81).

name	M245T75	M247T75
m/z	244.9683	246.9663
rt	74	74
intensity	125,276	1,996,395
isotope_labels_a	Cl, K, S	(^{34}S), (^{37}Cl), (^{41}K)
isotope_ids	M247T75, M247T75, M247T75	M245T75, M245T75, M245T75
isotope_labels_b	(^{34}S), (^{37}Cl), (^{41}K)	Cl, K, S
nl_labels		
exact_mass	244.9682989	246.966255
ppm_error	3.61	2.92
rt_diff		
adduct	$[M+Br]^-$	$[M+Br]^-$
molecular_formula	$C_6H_6N_4O_2$	$C_6H_6N_4O_2$
compound_name	3-Methyxanthine	3-Methyxanthine
compound_id	HMDB0001886	HMDB0001886

Table 81: Although the [M-H]- in form of 3-methylxanthine was not annotated when customised adduct, isotope, and neutral loss lists were applied, the [M+Br]- ion form was annotated instead.

The examples shown here highlighted an important consideration and served as a cautionary tale. To create the customised adduct, isotope, and neutral loss lists, some common and expected adducts, isotopes, and neutral losses were used to annotate the various *m/z* differences present within each dataset. Such an approach can introduce various challenges. Firstly, if the reference list used for annotation is incomplete, then degenerate features within the dataset will not be annotated since the adduct, isotope, or neutral loss will not be added to the list. Moreover, incorrect annotation of the *m/z* differences observed in each dataset can lead to poor quality lists, assigning incorrect adducts, isotopes, and neutral losses. This in turn can impact on TP annotation. The common adducts, isotopes, and neutral losses used to annotate the *m/z* differences within each dataset were by no means exhaustive as the BEAMS package was still in early development. Thus, a smaller list of adducts, isotopes, and neutral losses was used to ensure the workflow worked accurately (see lists here 2.4.3). Moreover,

this type of scenario reinforces the importance of manual curation and user expertise when evaluating U(H)PLC-MS data, especially when fully computational workflows are applied. Data must always be evaluated, and manual adjustments made if or when required. Indeed, a fully automated workflow and some manual spot-checking are complementary approaches and can reduce the likelihood of applying parameters that worsen the quality of the results (451).

For the Lipids positive ion-mode dataset, using the customised adduct, isotope, and neutral loss lists yielded a TPR of 84%, whilst the default adduct, isotope, and neutral loss lists yielded a TPR of 79% (Figure 68). For the Lipids negative ion-mode dataset, the use of the customised adduct, isotope, and neutral loss lists resulted in a TPR of 76%, whilst use of the default lists yielded a TPR of 79%. Once again, a decrease in the TPR is observed when customised lists are applied to the negative ion-mode data, although not to the extent observed in the HILIC negative ion-mode dataset. The FPRs for the Lipids positive ion-mode dataset increased from 800% to 2,800% but decreased for the Lipids negative ion-mode dataset from 2,400% to 2,100% (Figure 69).

It is interesting to note that for all four assays, the TPR was never 100% (94-98%). This may be due to the way the data were filtered to only leave *m/z*-RT features that had a group_id and sub_group_id assigned. It may have been that some of the TP annotations had no annotated degenerate features and thus no sub_group_id assigned, hence were filtered out and not counted, reducing the TP annotation rate. Nonetheless, these results demonstrate the positive impact of considering degenerate features and attempting to annotate them more comprehensively. Customised adduct, isotope, and neutral loss lists are recommended to maximise the TPR of annotation. However, use of longer lists can reduce the quality of the results, decreasing the TPR of annotation whilst increasing FP annotation rates. Caution must be taken when applying longer lists of adducts, isotopes, and neutral losses (see 2.4.3). This can be done through inspection of results and evaluation of where longer lists may mask annotation. It is not believed that customised lists will always cause issues such as those observed herein, but rather that these data may reflect on the quality of the lists applied.

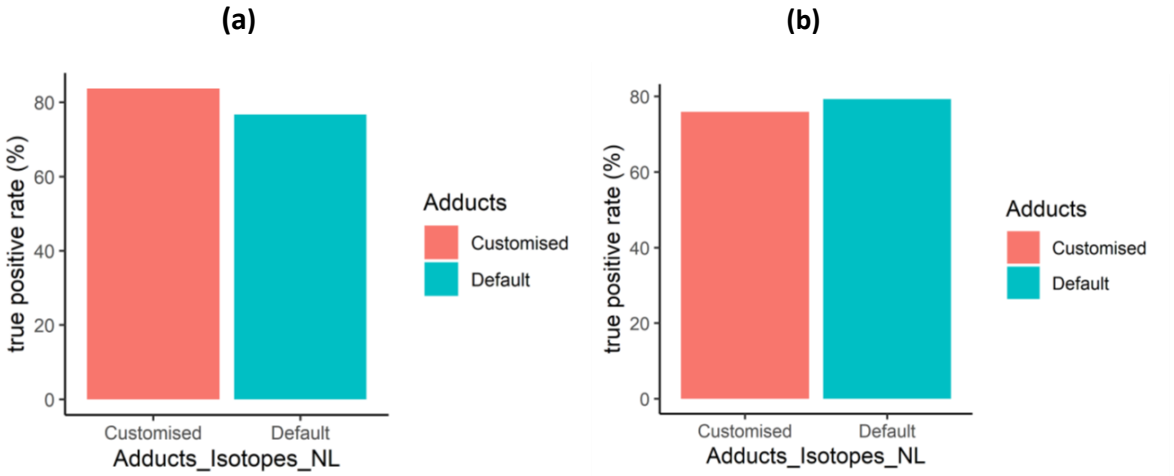


Figure 68: For the Lipids positive ion-mode dataset (a), application of customised adduct, isotope, and neutral loss lists yielded a TPR of 84%, whilst the default adduct, isotope, and neutral loss lists yielded TPRs of 79%. For the Lipids negative ion-mode dataset (b), application of customised adduct, isotope, and neutral loss lists yielded a TPR of 76%, whilst the default adduct, isotope, and neutral loss lists yielded TPRs of 79%.

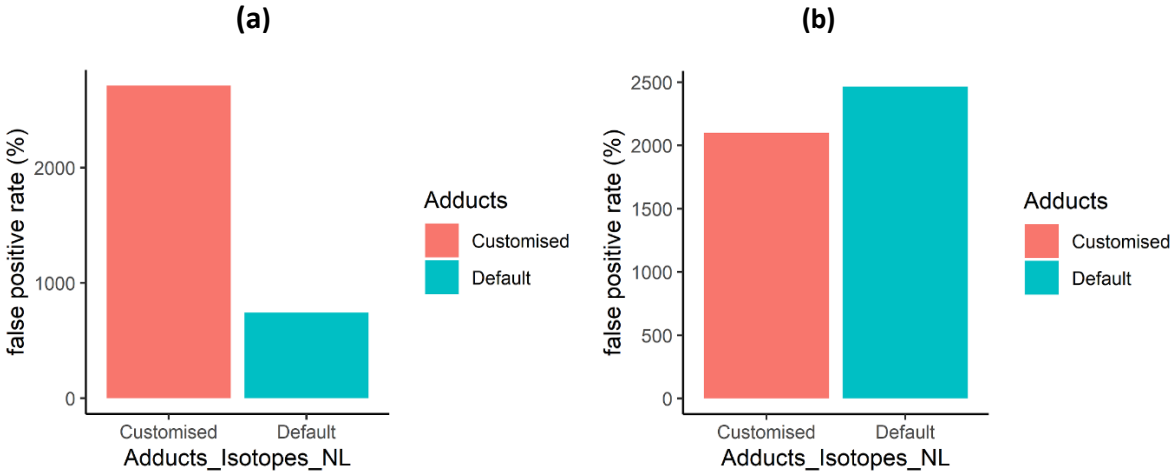


Figure 69: For the Lipids positive ion-mode dataset on the left-hand side (a), application of customised adduct, isotope, and neutral loss lists yielded a FPR of 2,800%, whilst the default adduct, isotope, and neutral loss lists yielded TPRs of 800%. For the Lipids negative ion-mode dataset on the right-hand side (b), application of customised adduct, isotope, and neutral loss lists yielded a FPR of 2,100%, whilst the default adduct, isotope, and neutral loss lists yielded TPRs of 2,400%.

5.2.7.2 Unique *m/z* features grouped according to retention time similarity

The total number of unique *m/z* features grouped when applying the default and customised adduct, isotope, and neutral loss lists was calculated. This was done for the 4 U(H)PLC-MS datasets, HILIC positive and negative ion-mode, and Lipids positive and negative ion-modes. This calculation informed on how longer and customised adduct, isotope, and neutral loss lists affected the total number *m/z* features deemed to be related to at least one other *m/z* feature.

Results for the HILIC and Lipids positive and negative ion mode assays, respectively, are shown in Figure 70 and Figure 71. For all four assays, the increase in *m/z* features grouped according to RT similarity indicated more annotation of degenerate features resulting from the customised and longer adduct, isotope, and neutral loss lists applied.

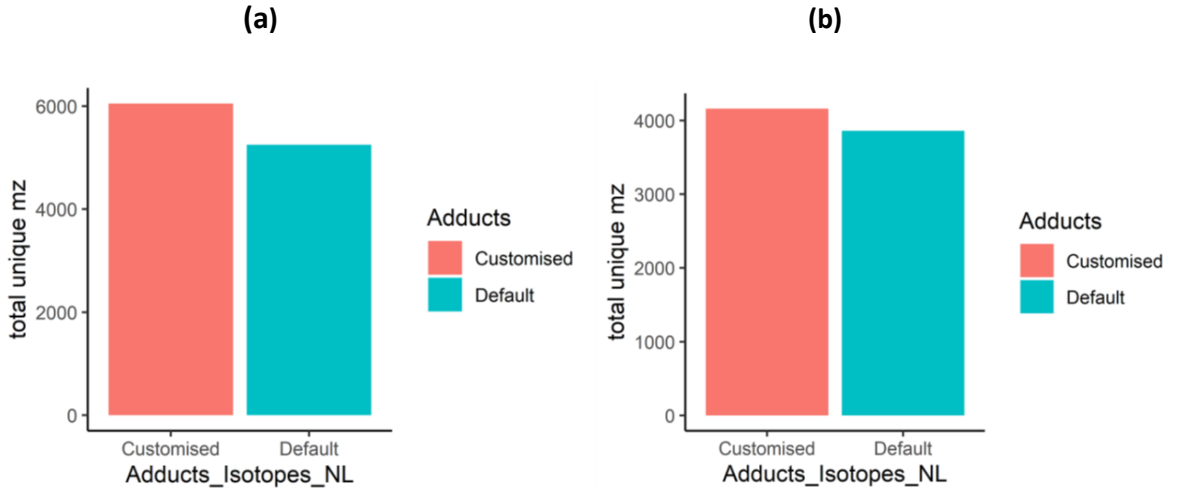


Figure 70: For the HILIC positive ion-mode dataset (a) 5,400 total grouped *m/z* features rose to 6,000 when customised adduct, isotope, and neutral loss lists were applied compared to when default adduct, isotope, and neutral loss lists were used. For the HILIC negative ion-mode dataset (b), 3,800 total grouped *m/z* features rose to 4,200 when customised adduct, isotope, and neutral loss lists were applied compared to when default adduct, isotope, and neutral loss lists were used.

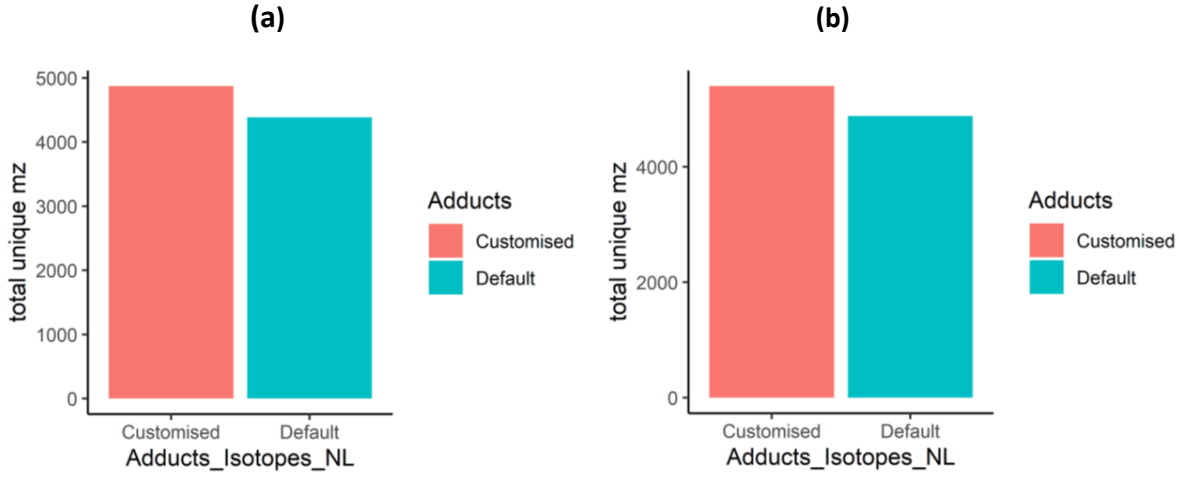


Figure 71: For the Lipids positive ion-mode dataset (a) 4,350 total grouped *m/z* features rose to 4,900 when customised adduct, isotope, and neutral loss lists were applied compared to when default adduct, isotope, and neutral loss lists were used. For the Lipids negative *ion-mode* dataset (b), 5,000 total grouped *m/z* features rose to 5,800 when customised adduct, isotope, and neutral loss lists were applied compared to when default adduct, isotope, and neutral loss lists were used.

5.2.7.3 Retention time groups formed

The total number of RT groups formed when applying the default and customised adduct, isotope, and neutral loss lists was calculated. This was done for the 4 U(H)PLC-MS datasets, HILIC positive and negative ion-mode, and Lipids positive and negative ion-modes. This calculation informed on how longer and customised adduct, isotope, and neutral loss lists affected the number of RT groups formed.

For all four assays, application of the customised adduct, isotope, and neutral loss lists resulted in a reduction in the number of RT groups formed. This corresponded to the higher number of *m/z* features deemed to be related to at least one other *m/z* feature, with less RT groups formed as more *m/z* features are grouped together and the size of each group increases. For the HILIC positive and negative ion-mode datasets, the number of RT groups formed decreased from 1,700 and to 1,100 to 1300 and 650, respectively (Figure 72). This could have resulted from the HILIC data containing broader chromatographic peaks, thus more *m/z* features could be grouped into a single, wide retention time window. For the Lipids positive and negative ion-mode datasets, the number of RT groups formed decreased from 1,850 and 1,150 to 700 and 480, respectively (Figure 73).

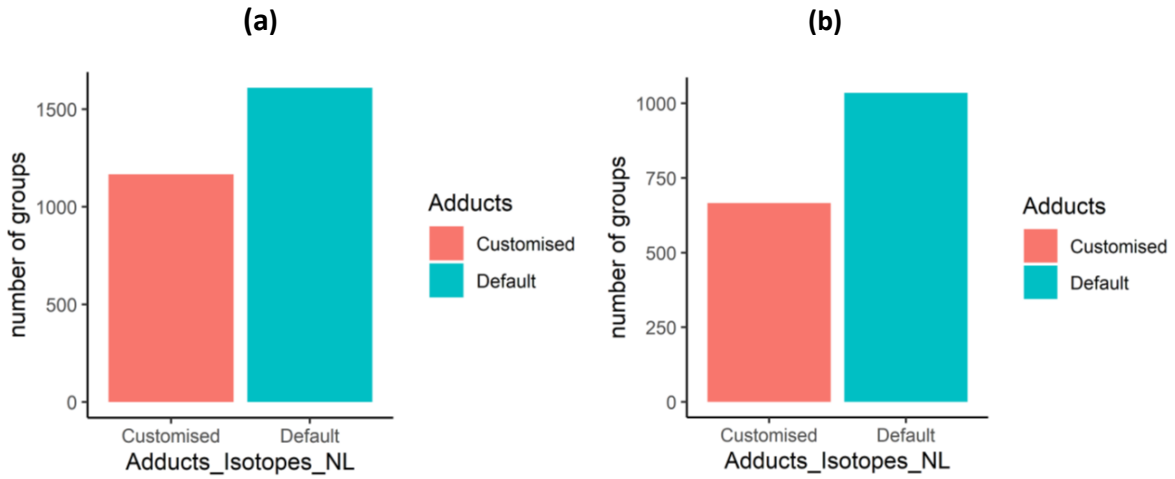


Figure 72: For the HILIC positive ion-mode dataset, the number of RT groups formed decreased from 1 700 to 1 300 when customised adduct, isotope, and neutral loss lists were applied compared to when default adduct, isotope, and neutral loss lists were applied. For the HILIC negative ion-mode dataset, the number of RT groups formed decreased from 1 100 to 650 when customised adduct, isotope, and neutral loss lists were applied compared to when default adduct, isotope, and neutral loss lists were applied.

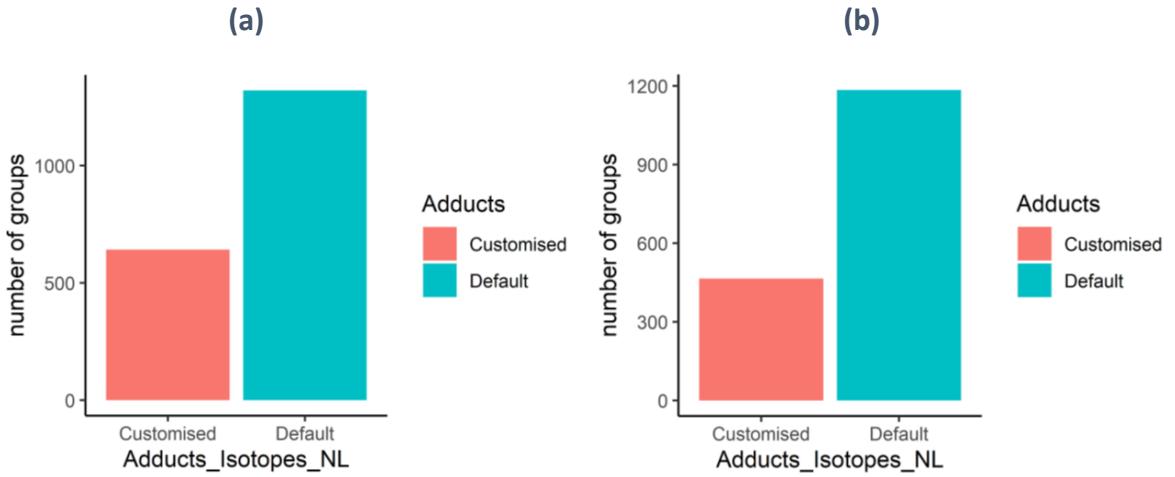


Figure 73: For the Lipids positive ion-mode dataset, the number of RT groups formed decreased from 1 850 to 700 when customised adduct, isotope, and neutral loss lists were applied compared to when default adduct, isotope, and neutral loss lists were applied. For the Lipids negative ion-mode dataset, the number of RT groups formed decreased from 1 150 to 480 when customised adduct, isotope, and neutral loss lists were applied compared to when default adduct, isotope, and neutral loss lists were applied.

5.2.7.4 Unique *m/z* per group

The number of unique *m/z* per RT group formed when applying the default and customised adduct, isotope, and neutral loss lists was calculated. This was done for the 4 U(H)PLC-MS datasets, HILIC positive and negative ion-mode, and Lipids positive and negative ion-modes. This calculation informed on how longer and customised adduct, isotope, and neutral loss lists affected the number of unique *m/z* per RT group.

For the HILIC positive ion-mode dataset, the median number of unique *m/z* per RT group increased from 2 to 3 when customised adduct, isotope, and neutral loss lists were applied. The interquartile range also increased from 2-3 to 2-4 when these customised lists were applied. The distribution of the outliers too increased when customised adduct, isotope, and neutral loss lists were applied, with the maximum number of unique *m/z* per RT group increasing from 50 to 330 (Figure 74). For the HILIC negative ion-mode dataset, the median number of unique *m/z* per RT group remained the same when customised adduct, isotope, and neutral loss lists were applied. The interquartile range also was unaffected, ranging from 1-3 when these customised lists were applied. However, the distribution of the outliers also increased when customised adduct, isotope, and neutral loss lists were applied, with the maximum number of unique *m/z* per RT group increasing from 200 to 450 (Figure 75).

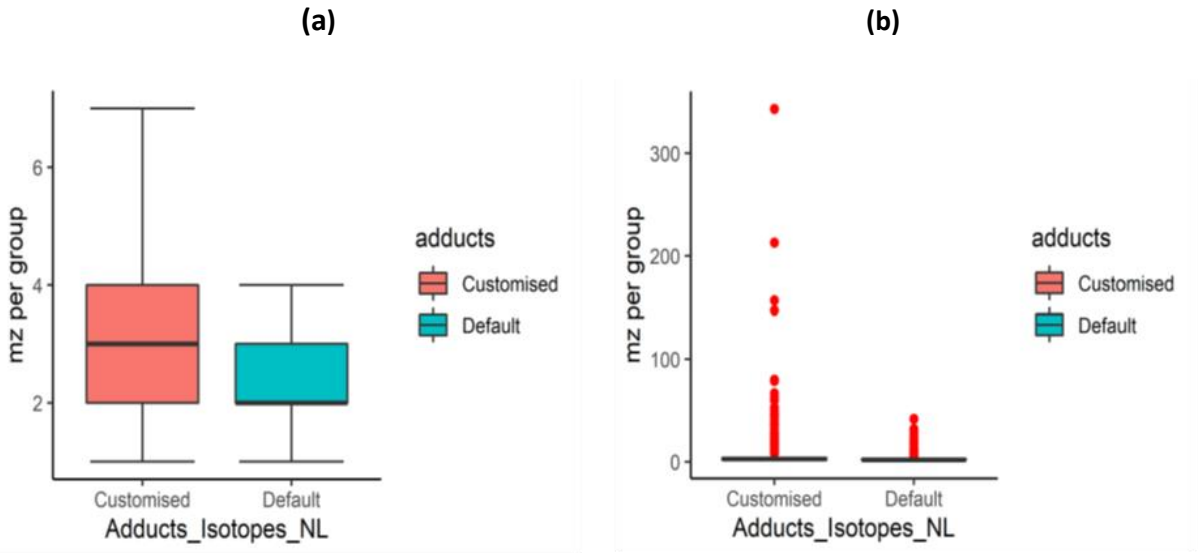


Figure 74: For the HILIC positive ion-mode dataset, the median number of unique *m/z* per retention group increased from two to three when customised adduct, isotope, and neutral loss lists were applied. The interquartile range yielded by applying the customised adduct, isotope, and neutral loss lists was 2-4, whilst this was 2-3 when default adduct, isotope, and neutral loss lists were applied. The distribution for the outliers was much higher when customised adduct, isotope, and neutral loss lists were applied, with a maximum number of unique *m/z* per RT group of 330 (compared to 50 when default adduct, isotope, and neutral loss lists were applied).

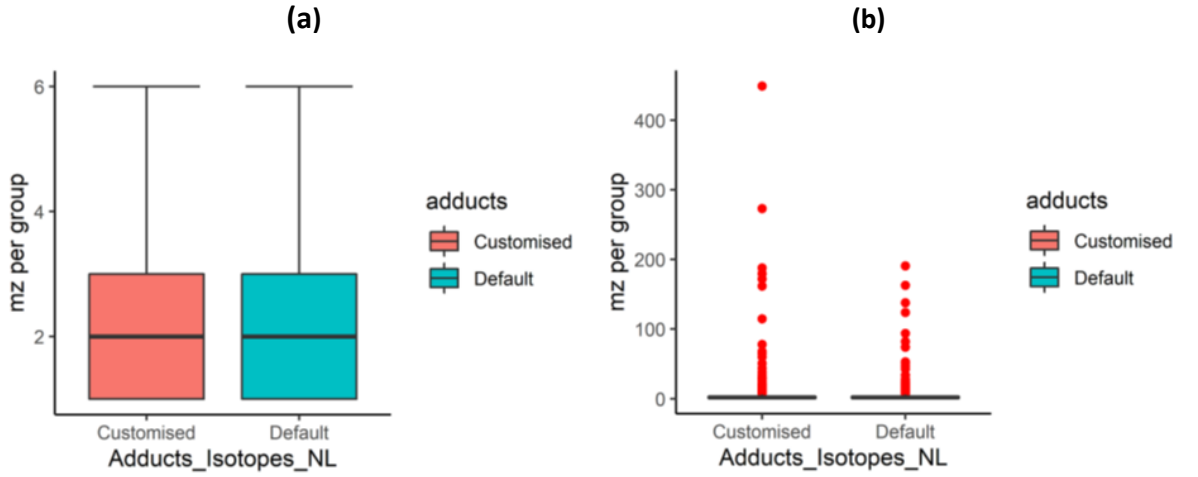


Figure 75: For the HILIC negative ion-mode dataset, the median number of unique *m/z* per retention group was unaffected when customised adduct, isotope, and neutral loss lists were applied, remaining constant at two. The interquartile range yielded by applying both the customised and default adduct, isotope, and neutral loss lists also did not change, ranging from 1-3. The distribution for the outliers was much higher when customised adduct, isotope, and neutral loss lists were applied, with a maximum number of unique *m/z* per RT group of 450 (compared to 200 when default adduct, isotope, and neutral loss lists were applied).

For the Lipids positive ion-mode dataset, the median number of unique *m/z* per RT group decreased from 3 to 2 when customised adduct, isotope, and neutral loss lists were applied. However, the interquartile range increased from 2-3 to 2-5 when these customised lists were applied. The distribution of the outliers also increased when customised adduct, isotope, and neutral loss lists were applied, with the maximum number of unique *m/z* per RT group increasing from 50 to 980 (Figure 76). For the Lipids negative ion-mode dataset, the median number of unique *m/z* per RT group remained the same at two when customised adduct, isotope, and neutral loss lists were applied. The interquartile range, however, increased from 2-3 to 2-5 when these customised lists were applied. The distribution of the outliers also increased when customised adduct, isotope, and neutral loss lists were applied, with the maximum number of unique *m/z* per RT group increasing from 470 to 790 (Figure 77).

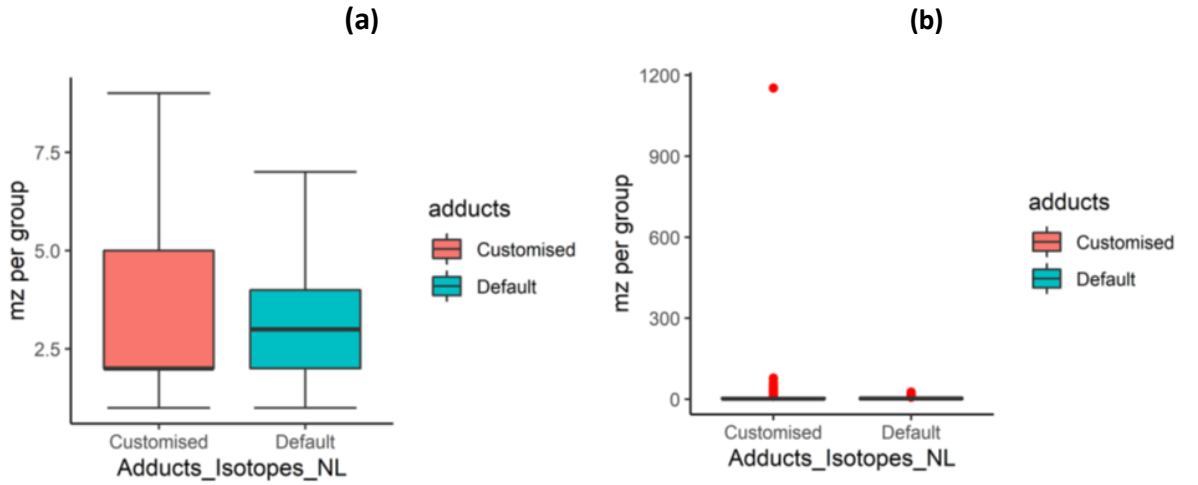


Figure 76: For the Lipids positive ion-mode dataset, the median number of unique *m/z* per retention group decreased from three to two when customised adduct, isotope, and neutral loss lists were applied. The interquartile range yielded by applying the customised adduct, isotope, and neutral loss lists, however, was 2-5, whilst this was 2-3 when default adduct, isotope, and neutral loss lists were applied. The distribution for the outliers was much higher when customised adduct, isotope, and neutral loss lists were applied, with a maximum number of unique *m/z* per RT group of 980 (compared to 50 when default adduct, isotope, and neutral loss lists were applied).

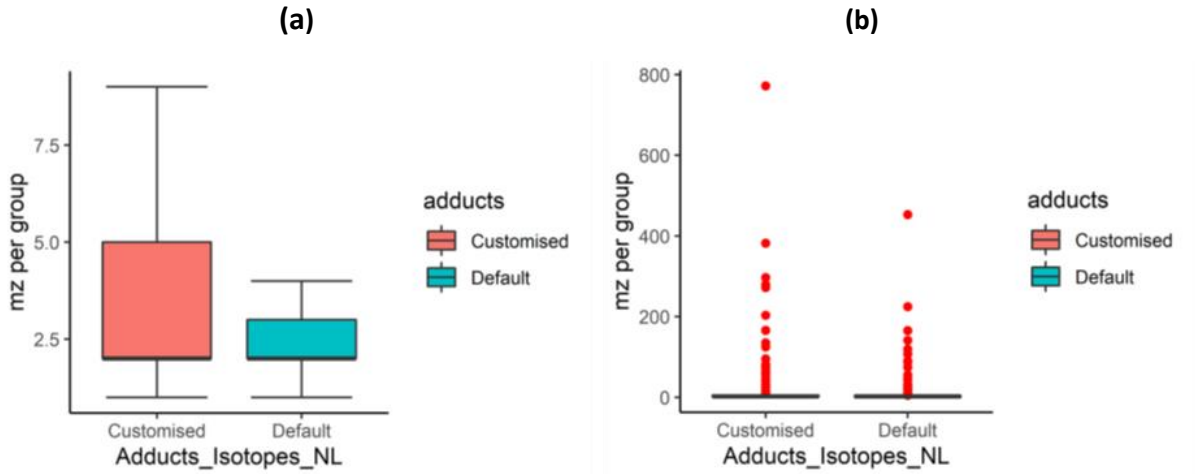


Figure 77: For the Lipids negative ion-mode dataset, the median number of unique *m/z* per retention group remained the same at two when customised adduct, isotope, and neutral loss lists were applied. The interquartile range yielded by applying the customised adduct, isotope, and neutral loss lists, however, was 2-5, whilst this was 2-3 when default adduct, isotope, and neutral loss lists were applied. The distribution for the outliers was much higher when customised adduct, isotope, and neutral loss lists were applied, with a maximum number of unique *m/z* per RT group of 790 (compared to 470 when default adduct, isotope, and neutral loss lists were applied).

5.2.7.5 Unique annotations per retention time group

The number of unique annotations per RT group formed when applying the default and customised adduct, isotope, and neutral loss lists was calculated. This was done for the 4 U(H)PLC-MS datasets, HILIC positive and negative ion-mode, and Lipids positive and negative ion-modes. This calculation informed on how longer and customised adduct, isotope, and neutral loss lists affected the number of unique annotations per RT group.

For the HILIC positive ion-mode dataset, the median number of annotations per RT group increased from 3 to 5 when customised adduct, isotope, and neutral loss lists were applied. The interquartile ranges also increased from 1-8 to 2-14 when customised adduct, isotope, and neutral loss lists were applied (Figure 78). For the HILIC negative ion-mode dataset, the median number of annotations per RT group decreased slightly from 7 to 6 when customised adduct, isotope, and neutral loss lists were applied. The interquartile ranges also decreased slightly from 2-22 to 2-20 when customised adduct, isotope, and neutral loss lists were applied. The decrease in the median number of annotations per RT group observed, was, however, only minor (Figure 78).

For the Lipids positive ion mode dataset, the median number of annotations per RT group increased from 8 to 10 when customised adduct, isotope, and neutral loss lists were applied. The interquartile ranges also increased from 3-35 to 3-38 when customised adduct, isotope, and neutral loss lists were applied (Figure 79). For the Lipids negative ion-mode dataset, the median number of annotations per RT group increased from 5 to 10 when customised adduct, isotope, and neutral loss lists were applied. The interquartile ranges also increased from 3-20 to 3-37 when customised adduct, isotope, and neutral loss lists were applied (Figure 79).

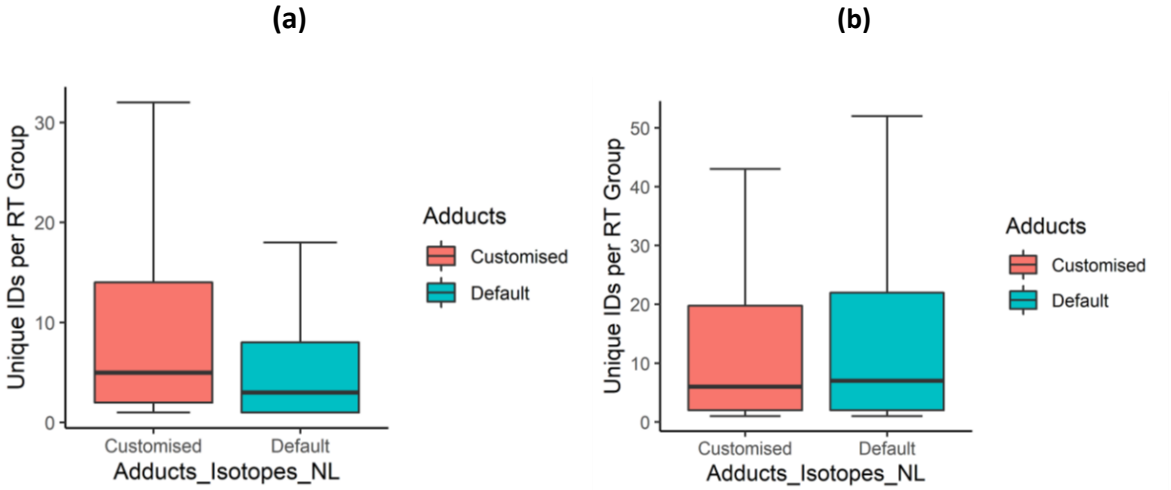


Figure 78: For the HILIC positive ion-mode dataset, the median number of annotations per RT group increased from 3 to 5 when customised adduct, isotope, and neutral loss lists were applied. The interquartile ranges also increased from 1-8 to 2-14 when customised adduct, isotope, and neutral loss lists were applied. For the HILIC negative ion-mode dataset, the median number of annotations per RT group decreased slightly from 7 to 6 when customised adduct, isotope, and neutral loss lists were applied. The interquartile ranges also decreased slightly from 2-22 to 2-20 when customised adduct, isotope, and neutral loss lists were applied.

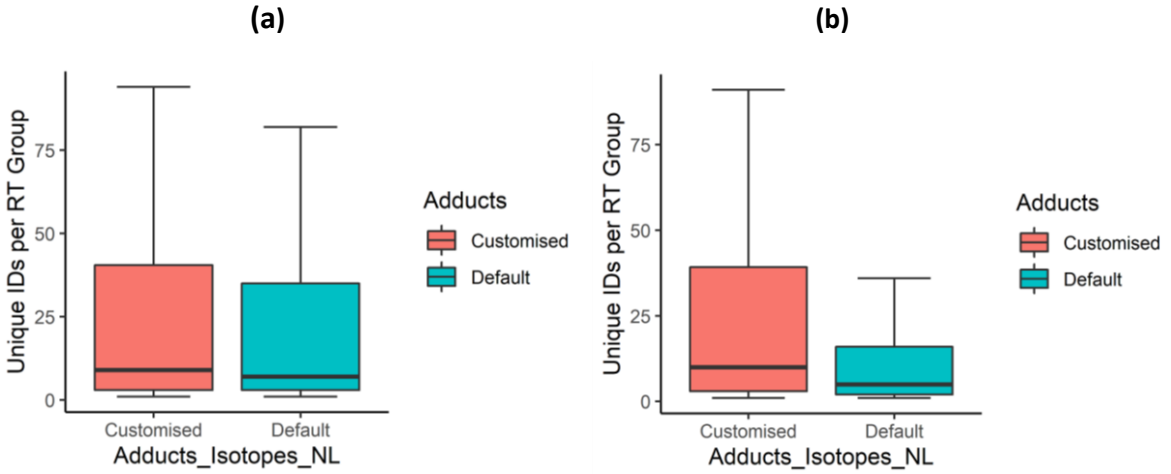


Figure 79: For the Lipids positive ion-mode dataset, the median number of annotations per RT group increased from 8 to 10 when customised adduct, isotope, and neutral loss lists were applied. The interquartile ranges also increased from 3-35 to 3-38 when customised adduct, isotope, and neutral loss lists were applied. For the Lipids negative ion-mode dataset, the median number of annotations per RT group increased from 5 to 10 when customised adduct, isotope, and neutral loss lists were applied. The interquartile ranges also increased from 3-20 to 3-37 when customised adduct, isotope, and neutral loss lists were applied.

5.2.8 Adduct, Isotope, and Neutral Loss Lists: Conclusion

Four U(H)PLC-MS methods, HILIC positive and negative ion-mode, and lipids positive and negative ion-modes, were applied for the analysis of human blood serum. These data were processed using XCMS (292), and then used to test the impact of customising the adduct, isotope, and neutral loss list settings during the BEAMS annotation workflow. Default adduct, isotope, and neutral loss lists were applied first and then customised lists were created and applied as described in the methods section (2.4.3).

The highest TPRs [1] were observed when customised adduct, isotope, and neutral loss lists were applied to the positive ion-mode data (HILIC and Lipids). However, application of these customised lists reduced the TPRs for both negative ion-mode datasets. The decrease was large in the HILIC negative ion-mode dataset (from 96% to 59%), and investigations showed that assignment of additional isotopes and neutral losses can sometimes impede annotation of some features. The reduction in the Lipids negative ion-mode dataset was only minor (from 79% to 76%), and it is important to note that only a small number of identifications were used for these calculations, thus a decrease of only one or two can translate to a large percentage difference. In particular, the negative ion-mode datasets had the lowest number of identifications, with 27 for the HILIC negative ion-mode dataset, and 29 for the Lipids negative ion-mode dataset. The FPRs increased when the customised adduct, isotope, and neutral loss lists were applied. Despite this, for the positive ion-mode datasets this increase in FP annotation did not impact the TPRs, although it did for the negative ion-mode datasets. As such, customised adduct, isotope, and neutral loss lists were concluded to be overall advantageous, although these must be applied with caution and manual checking of results is highly recommended.

For the number of adducts assigned to each of the TP annotations [2], the customised lists increased or had no impact on the median number in the positive ion-mode datasets, whilst generally less adducts were assigned to each TP annotation in the negative ion-mode datasets. For both the number of isotopes [3] and neutral losses [4] assigned to each TP annotation, an increase in the median was observed for three out of four assays (HILIC positive and Lipids positive and negative ion-modes) when customised adduct, isotope, and neutral loss lists were applied. For the HILIC negative ion-mode dataset, the median number

of isotopes and neutral losses assigned to each TP annotation was unaffected by using the customised lists and remained the same.

The number of *m/z* features grouped [5] increased for all four assays when customised adduct, isotope, and neutral loss lists were applied, whilst the number of RT groups formed [6] decreased. The number of *m/z* features within each RT group [7] increased (both the interquartile ranges and the distribution of the outliers) when customised adduct, isotope, and neutral loss lists were applied, whilst the number of annotations per RT group [8] also increased when customised adduct, isotope, and neutral loss lists were applied.

Overall, customised lists are advised as they clearly increase the annotation of isotopes and neutral losses. Moreover, more features are grouped together when such lists are applied, and more metabolite annotations are made as well. However, long lists can negatively impact annotation, increasing both FP and false negative annotation whilst decreasing TP annotation. It is therefore important not to simply apply excessively long lists for the sake of comprehensive annotation. Rather, exploration of the data in the form of discerning what *m/z* differences exist within the data must be made to determine what *m/z* differences are present within, and from this information appropriate lists can then be created.

5.3 Conclusions

BEAMS parameters were optimised using 4 U(H)PLC-MS datasets (HILIC positive and negative ion-mode, and Lipids positive and negative ion-modes). Each parameter was changed one at a time, and data evaluated by considering various calculations such as the TPRs and FPRs, the number of RT groups created, the number of unique *m/z*-rt pairs, the number of *m/z* features grouped overall, and the number of *m/z* and annotations per RT group. Generally, a lower number of RT groups with more *m/z* features grouped together within each group was desirable, highlighting more grouping of features that would otherwise be considered individual metabolites when they are in-fact degenerate features. Moreover, a high number of annotated groups would be optimal, and a high number of annotations per group was also desirable.

A 1s maxRT was found to be optimal for all four datasets, whilst optimal Pearson's correlation coefficient parameters were r=0.5 and p=0.05. A 5ppm mass error tolerance was found to maximise TPRs and therefore optimal, and the customised adduct, isotope, and neutral losses lists were advantageous in increasing annotation, particularly in the positive ion-mode datasets (Table 82).

BEAMS ANNOTATION STEP	OPTIMISED PARAMATER
Maximum RT Diff (s)	1
Pearson Correlation (r, p)	r=0.5 p=0.05
Mass error (ppm)	5
Adducts	pos=$[M+H]^+$, $[M+Na]^+$, $[M+K]^+$, $[M+NH_4]^+$ neg=$[M-H]^-$, $[M+Cl]^-$, $[M+Na-2H]^-$, $[M+K-2H]^-$, $[M+Hac-H]^-$, $[M+Br]^-$, $[M+NH_4-2H]^-$, $[M+C_2H_4O_2-H]^-$, $[M+C_2HF_3O_2-H]^-$ ***(preferably these must be customised for each dataset)***
Isotopes	pos= ^{13}C, ^{33}S, ^{34}S, ^{41}K, ^{15}N, ^{17}O, ^{18}O neg= ^{37}Cl, ^{13}C, ^{15}N, ^{17}O, ^{18}O, ^{33}S, ^{34}S
Neutral Losses	H_2O, CO, NH_3, C_2H_2, C_2H_4, CO2, C_2H_4O, C, O, Na, H_2, $CHCl_3$, CH_2O_2, HCOOK, HCOONa, $C_2F_3NaO_2$, NaCl, CH_3OH, H_3O_4P, C_2, CH_3CN, $NaNO_3$, HCl, $C_2H_4O_2$, $C_2HF_3O_2$

Table 82: Optimised and recommended BEAMS parameters for use within the ACMG.

The parameters recommended herein are suitable for data collected on the same analytical instruments and methods as the data presented here and using the same data pre-processing parameters. Application of different analytical instruments and methods is likely to yield different results. For example, different methods will have varying chromatographic peak widths and therefore resolution and require a wider or narrower maxRT. In these data, there were minimal changes in *m/z* features grouped when maxRT thresholds of >=0.5s were applied, but less when a threshold of 0.2s was applied. This reduction in grouped *m/z* results from peak apex variation. Although degenerate features of the same analyte co-elute, peak apex detection can be impeded by smoothing algorithms during peak-picking, and noise-peaks, as has been discussed (443,452). Consequently, for these data, which are expected to have peak-widths of 3-10s, a 1s maxRT was optimal as minimal increase in grouped *m/z* was observed after this. Moreover, using a wider window would increase the risk of grouping *m/z* features of different, closely eluting metabolites, thereby increasing FPRs. Correlation analysis results demonstrated correlation thresholds of 0-0.5 resulted in the highest number of

annotations, and the highest TPRs and FPRs. Applying stricter correlation thresholds reduced both TPRs and FPRs. As such, to maximise annotations, lower correlation thresholds can be applied, but strict thresholds facilitate data filtering, reducing the number of candidate annotations. P-value thresholds also had minimal impacts on the number of annotations achieved, but stricter thresholds in some datasets reduced the total number of annotations. As such, correlation analysis is an important parameter, but applying strict thresholds can reduce both TPR and FPR.

The customised adduct, isotope, and neutral loss lists are an important parameter, but the lists recommended are not fixed, and it is important that each user investigate their datasets to determine which *m/z* differences are present and use this information to guide the creation of appropriate lists. Different mobile phase compositions may give a sample type tendency towards certain adducts. Indeed, analytical sensitivity may impact the detection of low intensity isotopic peaks and thus make redundant the need to have extensive isotopic reference lists. Lists applied for this work, although longer than the default lists, were still not comprehensive. Application of longer lists may further improve annotation of degenerate features. The mass accuracy of the dataset may also impact on an appropriate mass error tolerance. Whilst not to be taken verbatim, the recommended BEAMS parameters presented are still of community wide relevance, demonstrating to anyone applying RT similarity and correlation analysis for annotation of degenerate features which annotation steps need more careful consideration.

Finally, it is important that each user carefully consider their priorities. The parameters offered herein maximise the number of annotations obtained and the grouping of degenerate features of the same metabolite. However, an increase in annotation can be a negative thing, as has been demonstrated throughout this thesis. More annotations often mean more FPs; thus, one may decide that their objectives are to filter the number of annotations and leave only those that survive this strict filtering. Results in this chapter demonstrate that smaller maxRTs, higher r and lower p-value thresholds, lower mass error tolerances, and smaller adduct, isotope, and neutral loss lists can all reduce the number of annotations yielded using BEAMS. However, this reduction in annotations is often accompanied by a reduction in TPRs.

The optimisation work carried out here highlighted an important community wide issue. The use of slightly longer adduct, isotope, and neutral loss lists resulted in more *m/z* features

being annotated as such. This demonstrates clearly that there are many degenerate features often not looked for during metabolomics annotation workflows. The issue is not availability of tools to do this, but rather lies in the multi-faceted nature of metabolomics. Many users lack sufficient computational skills to delve deeply into the annotation process, and often default parameters are applied for ease. Although these parameters do indeed yield viable results, the future of metabolomics must include further computational training. Moreover, the use of different tools, and therefore different approaches, further complicates matters, making it more challenging to amalgamate and integrate community wide knowledgebases. Indeed, the use of different tools means many different platforms are used for data annotation, such as tools based in R and Python, adding further complexity. Online resources such as the new GNPS (453), and MetaboAnalyst (454) which offer a single open-source platform for various processing steps are desirable. Last, but certainly not least, is the need for more integration of sources, especially when constructing customised adduct, isotope, and neutral loss lists. Constructing an infinitely long list is not useful without some other information. For example, it is not clear how prevalent some ion forms may be, and indeed many *m/z* differences found within datasets are unannotated. This type of information could help guide the formation of more sensible lists. This is particularly demonstrated in the negative ion-mode datasets herein, where longer adduct lists resulted in a reduction in the TPRs.

To conclude, optimal BEAMS parameters for processing polar and non-polar datasets are recommended to the reader, but caution is advised when applying such parameters, and community wide efforts are required to further address the issues of degenerate features and how best to annotate them, and more computational training would be beneficial within the metabolomics community. Further work is required using other datasets from different instruments and labs to ascertain which parameters are dataset-dependent and which parameters are dataset-independent.

6 Conclusions

6.1 Limitations and Future Work

6.1.1 (nano)ESI-DIMS Method Performance

Application of (nano)ESI-DIMS comes with inherent limitations. The lack of separation means FPRs are often very high and TPRs are relatively low, particularly in biological and environmental matrices, and as such it is next to impossible to derive conclusions with much confidence. The increased sensitivity compounds the matter, resulting in detection of a larger number of *m/z* that cannot confidently be annotated. A potential reason for the lower TPRs in the biological and environmental matrices presented here was the lower concentration at which the xenobiotics were in the sample extracts due to both the experimental design and the extraction process. This experimental design reveals an important lesson for exposomics. Xenobiotics in biological and environmental samples are unlikely to be highly concentrated (455). As such, their detection may be inhibited. This means that development of future (nano)ESI-DIMS NTA methods for the detection of xenobiotics in such matrices must carefully consider analyte concentration. The use of nano(ESI) already serves to improve sensitivity, but if collection of MS^n data is considered to improve annotation confidence, albeit its pitfalls (456), then this issue must also be considered. (nano)ESI-DIMS NTA for chemical analysis in the serum matrix is an excellent candidate for detection of the chemical exposome in clinical samples. Indeed, an increasing number of publications exist analysing the blood exposome (455,457,458), and as such makes (nano)ESI-DIMS methods an exceptionally attractive high throughput option offering decent TPRs of annotation. Moreover, (nanoESI-DIMS has already been applied for analysis of roach (458), human cancer cells (459), mosquito cells (460), and Daphnia magna (461), widening its potential for detection of xenobiotics in various matrices.

To alleviate issues relating to the lack of separation, results presented in this thesis applied a data correction. This had its limitations as many assumptions were made about mass accuracy. Mass error histograms were plotted, and it was assumed that the mass error region within which most annotations clustered likely contained TP annotations. To correct data, the median ppm of this region was calculated and then added or subtracted (whichever direction moved towards 0ppm) from each recorded *m/z* within the dataset. This assumed that all mass errors were equal for all *m/z*. Moreover, it also shifted other *m/z* away from 0ppm. This theory was not validated due to time constraints, and as such it is unknown how accurate it was. However, based on the number of true annotations still found, particularly in the clean

solvent matrix, there is some indication that the data correction was effective. Future work could further investigate this approach to data correction and validate this strategy. However, post-acquisition data correction is fast becoming a dated strategy. In the last few years, new Orbitrap instruments have been released (e.g., Orbitrap IDX, Fusion, and Exploris) which have an internal calibrant (462). In these instruments, a second ion source, called the reagent ion source (RIS), continuously introduces a regulated number of calibrant ions into the larger population of analyte ions, whilst the location of the RIS allows the quadrupole mass filter to isolate these reference ions to be used as a lock mass. This significantly improves instrument mass accuracy to less than 1ppm, rendering any post-acquisition correction moot. Future studies must therefore make use of these newer instruments if feasible, as these instruments offer much improved mass accuracies.

Annotation of (nano)ESI-DIMS using MI-pack did not apply correlation analysis to attempt to group degenerate features. Future work must also investigate the impact of such a grouping strategy on these data as this may further reduce FPRs. There is a large influx of computational tools increasingly focused on grouping of degenerate features such as MSClust (327) and RAMClust (332). However, most of these tools focus on U(H)PLC-MS data, where retention time similarity can be used to aid grouping. There is currently little research focused on grouping feature degeneracy in DIMS data, and indeed how accurate or effective such grouping is. This must be investigated to further improve annotation of (nano)ESI-DIMS data.

Finally, Phase II annotations were carried out on databases containing <400 chemicals. Such a scenario is highly unfeasible in real life, and although the results demonstrate the value of such small databases, future work must design more appropriate strategies to create smaller, more accurate databases that do not miss out on vital information. One novel strategy to create smaller databases is to apply (nano)ESI-DIMS methods to screen large databases such as KEGG (463) and ChemSpider (464) to create smaller suspect lists or databases relevant to each sample type with the lowest number of irrelevant components. These methods are quick, sensitive, and will yield TPRs of at least ~40%. Although this will cost time and money, the benefits of using such smaller databases are undeniable, and this must be investigated further. There is also scope for building species-specific databases (465–467), although this strategy remains challenged by how common metabolites appear across various species, and by current analytical capabilities as metabolome coverage is still not complete. Creating more

specific databases, e.g., the blood exposome database (468) is difficult as there are many unknown unknowns that constitute the chemical exposome, and thus cannot be added to databases or reference lists. Moreover, sensitivity to detect low abundance chemicals in biological and environmental samples, and confident identification of such chemicals, impedes the creation of specific databases. Moreover still, most NTA are exploratory, making it challenging to create any specific databases without removing potentially relevant chemicals from such databases. Despite this, these results should at least discourage the use of excessively large databases and continue to encourage more collaborative work for the continued sharing of resources and data to create more accurate databases. Exposomics has an advantage over metabolomics as there are likely to be more standards available, so level 1 identification of chemicals can be achieved, and databases can be constructed more confidently. Nonetheless, steps towards building such databases would serve to begin reducing current search spaces, and therefore FDR of annotation.

6.1.2 U(H)PLC-MS Method Performance

Application of the HILIC assays showed that most chemicals eluted in the void volume. Although still detected and annotated, these results are not an accurate representation of the benefits of HILIC separation. Moreover, isomers and isobars cannot always be separated just by applying a separation technique. This is where application of multiple complementary assays is vital. Some assays applied in a study should offer the maximum global coverage, but some must be tailored to maximise separation of isomers and isobars which are prevalent in nature. Future work must therefore optimise the HILIC assays applied to analyse xenobiotics. Moreover, there are a wide variety of analytical strategies that are suitable for small, polar analytes such as those amenable to HILIC analyses, whose methods can be challenging to develop due to poorly understood retention mechanisms (469). For example, ion-exchange chromatography (IEC) and capillary electrophoresis (CE), which separate charged analytes, have been shown to yield more reproducible data than HILIC when analysing metabolites (470). Preliminary results for ENTACT show that neither IEC nor CE were included in this global ring trial (396), and as such there is scope for testing these methods for the analysis of xenobiotics. Irrespective of which column chemistry is selected, there are still challenges with separation of isomers. Several studies have demonstrated the utility of LC coupled first to ion

mobility (IM), which separates analytes based on their movement through an inert gas such as nitrogen, and subsequently detected using MS (LC-IM/MS) (471–474). Few studies have investigated LC-IMS/MS for analysis of xenobiotics, yet it can offer improved separation of isomers and therefore reduced FDR. There is therefore scope for development of such methods in future work.

Retention time databases applied at Phase II were created using MS1 extracted ion chromatograms due to time constraints. As such, these were not identifications but rather annotations. Inclusion of MS2 spectra for confident identification of analytes added to any retention time database is crucial, therefore future work must curate these retention time databases using MS2 spectra to increase confidence of identified analytes. Additionally, spectral libraries can be created using the standard mixtures available, creating much needed public spectral libraries. Creation of RT databases or libraries is challenging, and it is extremely difficult to create comprehensive RT libraries. As a result, care must be taken when matching against a RT database as they are unlikely to contain comprehensive lists of chemicals. In lieu of this, a lack of annotation against a RT database is not a strong enough indicator on its own that a chemical is not present in a sample, as some of the constituents of the sample are likely not in the database. This is the case for creation of all databases but is particularly exacerbated in databases that require collection of data to create as data collection can be impeded by many different factors such as chemical concentration, analyte retention, and ionisation efficiency.

6.1.3 BEAMS Optimisation

The datasets applied for BEAMS optimisations represent only a small subset of available U(H)PLC-MS systems and set-up in metabolomics. This means that optimal parameters presented are most relevant to the ACMG at Birmingham and the Birmingham Phenome Centre, which use the same analytical instruments. Future work must therefore include a much larger number of datasets across the community to derive more global recommendations on optimal processing parameters. These datasets can be gathered through public data repositories such as MetaboLights (475) and Metabolomics Workbench (476). Moreover, a few automated tools have recently emerged for parameter optimisation

such as AutoTuner and Isotopologue Paramter Optimization (IPO) (477,478) which could be used to achieve this objective.

Only a small number of identified compounds were used to calculate TPRs. Future work must use a much larger number of confidently identified metabolites to provide a more accurate estimation of TPR.

Customised adduct, isotope, and neutral loss lists applied for BEAMS optimisations were longer than the default lists, but not sufficiently long as annotation of commonly observed adducts is on-going within the ACMG. Future work must focus on extensive characterisation of the prevalence and annotation of degenerate features so longer lists can be tested. To annotate degenerate features, more comprehensive lists of expected degeneracies within each dataset must be on the reference lists, and it is recommended that this type of investigation be incorporated into the workflows applied in metabolomics. Moreover, any such approach must be manually curated by the analyst to ensure that these longer lists are not negatively impacting annotation. In future experiments, similar testing of how annotation parameters impact grouping of degenerate features, creation of more comprehensive adduct, isotope, and neutral loss lists, and strategies for determining the mass error of collected datasets must all be investigated, using a much larger number of datasets to determine which parameters are dataset dependent and which are not.

6.2 Closing Statement

Much work is required to improve annotation of NTA MS1 data. However, any such work is limited by analytical capabilities such as sensitivity and selectivity. MS-based NTA is approaching technological limits, and innovative ideas are required. With so many techniques available, NTA could benefit from combinations of technologies, such as coupling of two separation techniques like LC and ion-mobility (IM) to MS or two-dimensional separation techniques (e.g., 2D-LC). There exist many factions in metabolomics and other NTA in general. Some sing praises of the high resolution of Orbitraps, others revere the fast scan speeds of TOFs, whilst others swear loyalty to different ionisation sources. This author strongly disagrees with these divisions. All these different technologies have played major roles in how far NTA has come, and efforts are required to come up with new and interesting analytical combinations. More collaboration amongst researchers is required to unclog the metabolite and chemical annotation bottleneck and using each technologies inherent biases is surely the next step to push metabolomics above and beyond. Either that or a new high speed, high resolution, and high mass accuracy mass spectrometer!

7 Bibliography

1. Kardos N, Demain AL. Penicillin: the medicine with the greatest impact on therapeutic outcomes. Applied Microbiology and Biotechnology. 2011 Nov 2;92(4):677–87.

2. Fleming A. Penicillin. British Medical Journal. 1941;2(4210):386.

3. Tamimi NAM, Ellis P. Drug Development: From Concept to Marketing! Nephron Clinical Practice. 2009;113(3):c125–31.

4. Rashid MBMA. Artificial Intelligence Effecting a Paradigm Shift in Drug Development. SLAS TECHNOLOGY: Translating Life Sciences Innovation. 2021 Feb 17;26(1):3–15.

5. Hill-McManus D, Marshall S, Liu J, Willke RJ, Hughes DA. Linked Pharmacometric-Pharmacoeconomic Modeling and Simulation in Clinical Drug Development. Clinical Pharmacology & Therapeutics. 2021 Jul 26;110(1):49–63.

6. Broadbent A. Causation and models of disease in epidemiology. Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences. 2009 Dec 1;40(4):302–11.

7. Furman K. Mono-Causal and Multi-Causal Theories of Disease: How to Think Virally and Socially about the Aetiology of AIDS. Journal of Medical Humanities. 2020 Jun 1;41(2):107–21.

8. Winder CL, Dunn WB, Goodacre R. TARDIS-based microbial metabolomics: time and relative differences in systems. Trends in Microbiology. 2011 Jul 1;19(7):315–22.

9. Oliver S. Systematic functional analysis of the yeast genome. Trends in Biotechnology. 1998 Sep 1;16(9):373–8.

10. Tweeddale H, Notley-Mcrobb L, Ferenci T. Effect of slow growth on metabolism of Escherichia coli, as revealed by global metabolite pool ('metabolome') analysis. Journal of Bacteriology. 1998;180(19):5109–16.

11. Piersigilli F, Lam TT, Vernocchi P, Quagliariello A, Putignani L, Zubair ·, et al. Identification of new biomarkers of bronchopulmonary dysplasia using metabolomics. Metabolomics. 2019;1:20.

12. Van Belkum A, Broadwell D, Lovern D, Petersen L, Weinstock G, Dunne WM. Proteomics and metabolomics for analysis of the dynamics of microbiota. Expert Review of Proteomics. 2018 Feb 28;15(2):101–4.

13. Trivedi DK, Sinclair E, Xu Y, Sarkar D, Walton-Doyle C, Liscio C, et al. Discovery of Volatile Biomarkers of Parkinson's Disease from Sebum. ACS Central Science. 2019 Apr 24;5(4):599–606.

14. Bekri S. The role of metabolomics in precision medicine. Expert Review of Precision Medicine and Drug Development. 2016 Nov 23;1(6):517–32.

15. Matich EK, Chavez Soria NG, Aga DS, Atilla-Gokcumen GE. Applications of metabolomics in assessing ecological effects of emerging contaminants and pollutants on plants. Journal of Hazardous Materials. 2019 Jul 5;373:527–35.

16. Jones OAH, Maguire ML, Griffin JL, Dias DA, Spurgeon DJ, Svendsen C. Metabolomics and its use in ecology. Austral Ecology. 2013 Sep 1;38(6):713–20.

17. Ladewig SM, Bao S, Chow AT. Natural Fibers: A Missing Link to Chemical Pollution Dispersion in Aquatic Environments. Environmental Science & Technology. 2015 Nov 3;49(21):12609–10.

18. Sumner LW, Mendes P, Dixon RA. Plant metabolomics: large-scale phytochemistry in the functional genomics era. Phytochemistry. 2003 Mar;62(6):817–36.

19. Bundy JG, Davey MP, Viant MR. Environmental metabolomics: A critical review and future perspectives. Vol. 5, Metabolomics. Springer; 2009. p. 3–21.

20. Chen D-Q, Chen H, Chen L, Tang D-D, Miao H, Zhao Y-Y. Metabolomic application in toxicity evaluation and toxicological biomarker identification of natural product. Chemico-Biological Interactions. 2016 May 25;252:114–30.

21. Gonzalez FJ, Fang Z-Z, Ma X. Transgenic mice and metabolomics for study of hepatic xenobiotic metabolism and toxicity. Expert Opinion on Drug Metabolism & Toxicology. 2015 Jun 3;11(6):869–81.

22. Bannuscher A, Hellack B, Bahl A, Laloy J, Herman H, Stan MS, et al. Metabolomics profiling to investigate nanomaterial toxicity in vitro and in vivo. Nanotoxicology. 2020 Jul 2;14(6):807–26.

23. Hernández-Mesa M, Le Bizec B, Dervilly G. Metabolomics in chemical risk analysis – A review. Analytica Chimica Acta. 2021 Apr 15;1154:338298.

24. Wild CP. Complementing the Genome with an “Exposome”: The Outstanding Challenge of Environmental Exposure Measurement in Molecular Epidemiology. Cancer Epidemiology Biomarkers & Prevention. 2005 Aug 1;14(8):1847–50.

25. Smith MT, McHale CM, de la Rosa R. Using Exposomics to Assess Cumulative Risks from Multiple Environmental Stressors. In: Unraveling the Exposome. Cham: Springer International Publishing; 2019. p. 3–22.

26. Turner MC, Vineis P, Seleiro E, Dijmarescu M, Balshaw D, Bertollini R, et al. EXPOsOMICS: final policy workshop and stakeholder consultation. BMC Public Health 2018 18:1. 2018 Feb 15;18(1):1–11.

27. Clayden J, Geeves N, Warren S. Organic Chemistry. Ed Oxford University Press. 2012;

28. Myers VH, Champagne CM. Nutritional effects on blood pressure. Current Opinion in Lipidology. 2007 Feb;18(1):20–4.

29. Oliver S, Wilson M, Kell DB, Baganz F. Systematic functional analysis of the yeast genome. Trends in Biotechnology. 1998 Sep 1;16(9):373–8.

30. Wild CP, Scalbert A, Herceg Z. Measuring the exposome: A powerful basis for evaluating environmental exposures and cancer risk. Environmental and Molecular Mutagenesis. 2013 Aug;54(7):480–99.

31. Wild CP, Scalbert A, Herceg Z. Measuring the exposome: A powerful basis for evaluating environmental exposures and cancer risk. Environmental and Molecular Mutagenesis. 2013 Aug 1;54(7):480–99.

32. Lakhtakia R. The History of Diabetes Mellitus. Sultan Qaboos University Medical Journal. 2013 Aug;13(3):368–70.

33. Nahyan Fancy. Verification and Utility in the Arabic Commentaries on the Canon of Medicine: Examples from the Works of Fakhr al-Dı¯n al-Razı¯ (d. 1210) and Ibn al-Nafı¯s (d. 1288). Journal of the History of Medicine and Allied Sciences; 2020. 361–382 p.

34. Gates SC, Sweeley CC. Quantitative Metabolic Profiling Based on Gas Chromatography. Vol. 24, Clinical Chemistry. 1978.

35. Horning EC, Horning MG. Metabolic Profiles: Gas-Phase Methods for Analysis of Metabolites. Clinical Chemistry. 1971 Aug 1;17(8):802–9.

36. Nicholson JK, Lindon JC, Holmes E. "Metabonomics": understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. Xenobiotica. 1999 Jan 22;29(11):1181–9.

37. Nicholson JK, Lindon JC. Metabonomics. Nature. 2008 Oct 22;455(7216):1054–6.

38. Ramsden JJ. Metabolomics and Metabonomics. In Springer, London; 2009. p. 1–6.

39. Pearson H. What is a gene? Nature. 2006 May 25;441(7092):398–402.

40. Hoeppner MP, Lundquist A, Pirun M, Meadows JRS, Zamani N, Johnson J, et al. An Improved Canine Genome and a Comprehensive Catalogue of Coding Genes and Non-Coding Transcripts. Chadwick BP, editor. PLoS ONE. 2014 Mar 13;9(3):e91172.

41. Jurowski K, Kochan K, Walczak J, Barańska M, Piekoszewski W, Buszewski B. Analytical Techniques in Lipidomics: State of the Art. Critical Reviews in Analytical Chemistry. 2017;47(5):418–37.

42. Moore LD, Le T, Fan G. DNA Methylation and Its Basic Function. Neuropsychopharmacology 2013 38:1. 2012 Jul 11;38(1):23–38.

43. Moore LD, Le T, Fan G. DNA methylation and its basic function. Vol. 38, Neuropsychopharmacology. Nature Publishing Group; 2013. p. 23–38.

44. Li L, Chen K, Wang T, Wu Y, Xing G, Chen M, et al. Glis1 facilitates induction of pluripotency via an epigenome–metabolome–epigenome signalling cascade. Nature Metabolism. 2020 Sep 1;2(9):882–92.

45. Chiarugi A, Dölle C, Felici R, Ziegler M. The NAD metabolome - A key determinant of cancer cell biology. Vol. 12, Nature Reviews Cancer. Nature Publishing Group; 2012. p. 741–52.

46. Kuile BH, Westerhoff H V. Transcriptome meets metabolome: Hierarchical and metabolic regulation of the glycolytic pathway. FEBS Letters. 2001 Jul 6;500(3):169–71.

47. Rossouw D, Næs T, Bauer FF. Linking gene regulation and the exo-metabolome: A comparative transcriptomics approach to identify genes that impact on the production of volatile aroma compounds in yeast. BMC Genomics. 2008 Nov 7;9(1):1–10.

48. Winkler WC, Nahvi A, Roth A, Collins JA, Breaker RR. Control of gene expression by a natural metabolite-responsive ribozyme. Nature. 2004 Mar 18;428(6980):281–6.

49. Renault H, El Amrani A, Berger A, Mouille G, Soubigou-Taconnat L, Bouchereau A, et al. γ-Aminobutyric acid transaminase deficiency impairs central carbon metabolism and leads to cell wall defects during salt stress in Arabidopsis roots. Plant, Cell and Environment. 2013 May 1;36(5):1009–18.

50. Ulvestad JS, Kumari J, Seternes T, Chi H, Dalmo RA. Studies on the effects of LPS, ß-glucan and metabolic inhibitors on the respiratory burst and gene expression in Atlantic salmon macrophages. Journal of Fish Diseases. 2018 Jul 1;41(7):1117–27.

51. Hargreaves M, Spriet LL. Skeletal muscle energy metabolism during exercise. Vol. 2, Nature Metabolism. Nature Research; 2020. p. 817–28.

52. Dickerson TJ, Janda KD. Aqueous aldol catalysis by a nicotine metabolite. Journal of the American Chemical Society. 2002 Apr 3;124(13):3220–1.

53. Covarrubias AJ, Perrone R, Grozio A, Verdin E. NAD+ metabolism and its roles in cellular processes during ageing. Vol. 22, Nature Reviews Molecular Cell Biology. Nature Research; 2021. p. 119–41.

54. Gillard J, Frenkel J, Devos V, Sabbe K, Paul C, Rempt M, et al. Metabolomics enables the structure elucidation of a diatom sex pheromone. Angewandte Chemie - International Edition. 2013 Jan 14;52(3):854–7.

55. von Bertalanffy L. Metabolic Types and Growth Types. The American Naturalist. 1951 Mar 29;85(821):111–7.

56. Roberts LD, Souza AL, Gerszten RE, Clish CB. Targeted Metabolomics. Current Protocols in Molecular Biology. 2012 Apr 15;98(1):30.2.1-30.2.24.

57. Zimmer D. New US FDA draft guidance on bioanalytical method validation versus current FDA and EMA guidelines: Chromatographic methods and ISR. Vol. 6, Bioanalysis. Future Science Ltd London, UK ; 2014. p. 13–9.

58. Kaza M, Karaźniewicz-Łada M, Kosicka K, Siemiątkowska A, Rudzki PJ. Bioanalytical method validation: new FDA guidance vs. EMA guideline. Better or worse? Journal of Pharmaceutical and Biomedical Analysis. 2019 Feb 20;165:381–5.

59. EMA. Guideline on bioanalytical method validation. 2011;

60. FDA. Bioanalytical Method Validation Guidance for Industry. 2018;

61. Björkqvist M, Ohlsson M, Minthon L, Hansson O. Evaluation of a previously suggested plasma biomarker panel to identify Alzheimer's disease. PLoS ONE. 2012 Jan 18;7(1):e29868.

62. Fraser DD, Slessarev M, Martin CM, Daley M, Patel MA, Miller MR, et al. Metabolomics Profiling of Critically Ill Coronavirus Disease 2019 Patients: Identification of Diagnostic and Prognostic Biomarkers. Critical Care Explorations. 2020 Oct 21;2(10):e0272.

63. Menni C, Zhai G, MacGregor A, Prehn C, Römisch-Margl W, Suhre K, et al. Targeted metabolomics profiles are strongly correlated with nutritional patterns in women. Metabolomics. 2013 Apr 1;9(2):506–14.

64. González-Domínguez R, Jáuregui O, Mena P, Hanhineva K, Tinahones FJ, Angelino D, et al. Quantifying the human diet in the crosstalk between nutrition and health by multi-targeted metabolomics of food and microbiota-derived metabolites. International Journal of Obesity. 2020 Dec 15;44(12):2372–81.

65. Vrhovsek U, Masuero D, Gasperotti M, Franceschi P, Caputi L, Viola R, et al. A versatile targeted metabolomics method for the rapid quantification of multiple classes of phenolics in

fruits and beverages. In: Journal of Agricultural and Food Chemistry. American Chemical Society; 2012. p. 8831–40.

66. Tallapally M, Sadiq AS, Mehtab V, Chilakala S, Vemula M, Chenna S, et al. GC-MS based targeted metabolomics approach for studying the variations of phenolic metabolites in artificially ripened banana fruits. LWT. 2020 Aug 1;130:109622.

67. Varma VR, Oommen AM, Varma S, Casanova R, An Y, Andrews RM, et al. Brain and blood metabolite signatures of pathology and progression in Alzheimer disease: A targeted metabolomics study. PLoS Medicine. 2018 Jan 1;15(1):e1002482.

68. Mahajan U V., Varma VR, Griswold ME, Blackshear CT, An Y, Oommen AM, et al. Dysregulation of multiple metabolic networks related to brain transmethylation and polyamine pathways in Alzheimer disease: A targeted metabolomic and transcriptomic study. Brayne C, editor. PLOS Medicine. 2020 Jan 24;17(1):e1003012.

69. López-López Á, López-Gonzálvez Á, Barker-Tejeda TC, Barbas C. A review of validated biomarkers obtained through metabolomics. Vol. 18, Expert Review of Molecular Diagnostics. Taylor and Francis Ltd; 2018. p. 557–75.

70. Beckles DM, Roessner U. Plant metabolomics. In: Plant Biotechnology and Agriculture. Elsevier; 2012. p. 67–81.

71. Chandramouli K, Qian P-Y. Proteomics: Challenges, Techniques and Possibilities to Overcome Biological Sample Complexity. Human Genomics and Proteomics. 2009 Jan;1(1).

72. Monteiro MS, Carvalho M, Bastos ML, Guedes de Pinho P. Metabolomics Analysis for Biomarker Discovery: Advances and Challenges. Current Medicinal Chemistry. 2013;20(2):257–71.

73. Salzberg SL. Open questions: How many genes do we have? BMC Biology. 2018 Dec 20;16(1):94.

74. HMDB. Human Metabolome Database: Classification Browse [Internet]. 2021 [cited 2021 Apr 6]. Available from: https://hmdb.ca/classyfication?page=4

75. Wishart DS, Tzur D, Knox C, Eisner R, Guo AC, Young N, et al. HMDB: the Human Metabolome Database. Nucleic Acids Research. 2007 Jan 3;35(Database):D521–6.

76. Tolstikov V, Moser AJ, Sarangarajan R, Narain NR, Kiebish MA. Current Status of Metabolomic Biomarker Discovery: Impact of Study Design and Demographic Characteristics. Metabolites. 2020 May 29;10(6):224.

77. Bennett MR, Devarajan P. Characteristics of an Ideal Biomarker of Kidney Diseases. Biomarkers of Kidney Disease. 2011 Jan 1;1–24.

78. Jacob M, Malkawi A, Albast N, Al Bougha S, Lopata A, Dasouki M, et al. A targeted metabolomics approach for clinical diagnosis of inborn errors of metabolism. Analytica Chimica Acta. 2018 Sep 26;1025:141–53.

79. Koves TR, Ussher JR, Noland RC, Slentz D, Mosedale M, Ilkayeva O, et al. Mitochondrial Overload and Incomplete Fatty Acid Oxidation Contribute to Skeletal Muscle Insulin Resistance. Cell Metabolism. 2008 Jan 9;7(1):45–56.

80. Spratlin JL, Serkova NJ, Eckhardt SG. Clinical applications of metabolomics in oncology: A review. Vol. 15, Clinical Cancer Research. American Association for Cancer Research; 2009. p. 431–40.

81. Davis VW, Bathe OF, Schiller DE, Slupsky CM, Sawyer MB. Metabolomics and surgical oncology: Potential role for small molecule biomarkers. Vol. 103, Journal of Surgical Oncology. J Surg Oncol; 2011. p. 451–9.

82. Nordström A, Lewensohn R. Metabolomics: Moving to the Clinic. Journal of Neuroimmune Pharmacology. 2010 Mar 28;5(1):4–17.

83. Mamas M, Dunn WB, Neyses L, Goodacre R. The role of metabolites and metabolomics in clinically applicable biomarkers of disease. Archives of Toxicology. 2011 Jan;85(1):5–17.

84. Beger RD, Dunn W, Schmidt MA, Gross SS, Kirwan JA, Cascante M, et al. Metabolomics enables precision medicine: "A White Paper, Community Perspective." Metabolomics. 2016 Oct 1;12(10):149.

85. Daemen A, Peterson D, Sahu N, McCord R, Du X, Liu B, et al. Metabolite profiling stratifies pancreatic ductal adenocarcinomas into subtypes with distinct sensitivities to metabolic inhibitors. Proceedings of the National Academy of Sciences of the United States of America. 2015 Aug 11;112(32):E4410–7.

86. Fan Y, Zhou X, Xia TS, Chen Z, Li J, Liu Q, et al. Human plasma metabolomics for identifying differential metabolites and predicting molecular subtypes of breast cancer. Oncotarget. 2016;7(9):9925–38.

87. Zhao H, Heimberger AB, Lu Z, Wu X, Hodges TR, Song R, et al. Metabolomics profiling in plasma samples from glioma patients correlates with tumor phenotypes. Oncotarget. 2016 Apr 12;7(15):20486–95.

88. Backshall A, Sharma R, Clarke SJ, Keun HC. Pharmacometabonomic profiling as a predictor of toxicity in patients with inoperable colorectal cancer treated with capecitabine. Clinical Cancer Research. 2011 May 1;17(9):3019–28.

89. Bertini I, Cacciatore S, Jensen B V., Schou J V., Johansen JS, Kruhøffer M, et al. Metabolomic NMR fingerprinting to identify and predict survival of patients with metastatic colorectal cancer. Cancer Research. 2012 Jan 1;72(1):356–64.

90. Coen M, Ruepp SU, Lindon JC, Nicholson JK, Pognan F, Lenz EM, et al. Integrated application of transcriptomics and metabonomics yields new insight into the toxicity due to paracetamol in the mouse. Journal of Pharmaceutical and Biomedical Analysis. 2004 Apr 1;35(1):93–105.

91. Rhee EP, Ho JE, Chen MH, Shen D, Cheng S, Larson MG, et al. A genome-wide association study of the human metabolome in a community-based cohort. Cell Metabolism. 2013 Jul 2;18(1):130–43.

92. Demirkan A, van Duijn CM, Ugocsai P, Isaacs A, Pramstaller PP, Liebisch G, et al. Genome-Wide Association Study Identifies Novel Loci Associated with Circulating Phospho- and Sphingolipid Concentrations. Gibson G, editor. PLoS Genetics. 2012 Feb 16;8(2):e1002490.

93. Gieger C, Geistlinger L, Altmaier E, Hrabé de Angelis M, Kronenberg F, Meitinger T, et al. Genetics Meets Metabolomics: A Genome-Wide Association Study of Metabolite Profiles in Human Serum. Gibson G, editor. PLoS Genetics. 2008 Nov 28;4(11):e1000282.

94. Hicks AA, Pramstaller PP, Johansson Å, Vitart V, Rudan I, Ugocsai P, et al. Genetic Determinants of Circulating Sphingolipid Concentrations in European Populations. Gibson G, editor. PLoS Genetics. 2009 Oct 2;5(10):e1000672.

95. Illig T, Gieger C, Zhai G, Römisch-Margl W, Wang-Sattler R, Prehn C, et al. A genome-wide perspective of genetic variation in human metabolism. Nature Genetics. 2010 Feb 27;42(2):137–41.

96. Kettunen J, Tukiainen T, Sarin AP, Ortega-Alonso A, Tikkanen E, Lyytikäinen LP, et al. Genome-wide association study identifies multiple loci influencing human serum metabolite levels. Nature Genetics. 2012 Mar 29;44(3):269–76.

97. Suhre K, Shin SY, Petersen AK, Mohney RP, Meredith D, Wägele B, et al. Human metabolic individuality in biomedical and pharmaceutical research. Nature. 2011 Sep 1;477(7362):54–62.

98. Suhre K, Wallaschofski H, Raffler J, Friedrich N, Haring R, Michael K, et al. A genome-wide association study of metabolic traits in human urine. Nature Publishing Group. 2011;43(6).

99. Tukiainen T, Kettunen J, Kangas AJ, Lyytik̈inen LP, Soininen P, Sarin AP, et al. Detailed metabolic and genetic characterization reveals new associations for 30 known lipid loci. Human Molecular Genetics. 2012 Mar 15;21(6):1444–55.

100. Lin HM, Barnett MPG, Roy NC, Joyce NI, Zhu S, Armstrong K, et al. Metabolomic analysis identifies inflammatory and noninflammatory metabolic effects of genetic modification in a mouse model of Crohn?s disease. Journal of Proteome Research. 2010 Apr 5;9(4):1965–75.

101. Gibney MJ, Macdonald IA, Roche HM. Nutrition and metabolism. Nutrition and metabolism. 2003;

102. Schneider H, Ma L, Glatt H. Extractionless method for the determination of urinary caffeine metabolites using high-performance liquid chromatography coupled with tandem mass spectrometry. Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences. 2003 Jun 15;789(2):227–37.

103. Mullen W, Boitier A, Stewart AJ, Crozier A. Flavonoid metabolites in human plasma and urine after the consumption of red onions: Analysis by liquid chromatography with photodiode array and full scan tandem mass spectrometric detection. Journal of Chromatography A. 2004 Nov 26;1058(1–2):163–8.

104. Strickland PT, Qian Z, Friesen MD, Rothman N, Sinha R. Metabolites of 2-amino-1-methyl-6-phenylimidazo(4,5-b)pyridine (PhIP) in human urine after consumption of charbroiled or fried beef. Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis. 2002 Sep 30;506–507:163–73.

105. Genchi G, Carocci A, Lauria G, Sinicropi MS, Catalano A. Nickel: Human Health and Environmental Toxicology. International Journal of Environmental Research and Public Health. 2020 Jan 21;17(3):679.

106. Wilber CG. Toxicology of selenium: A review. Vol. 17, Clinical Toxicology. Informa Healthcare; 1980. p. 171–230.

107. Pineau A. Selenium Toxicology. Selenium in Medicine and Biology. 2019 Jul 19;345–50.

108. Yokel RA. The toxicology of aluminum in the brain: A review. In: NeuroToxicology. 2000. p. 813–28.

109. Rebellato AP, Silva JGS, de Paiva EL, Arisseto-Bragotto AP, Pallone JAL. Aluminium in infant foods: toxicology, total content and bioaccessibility. Current Opinion in Food Science. 2021 Oct 1;41:130–7.

110. van der Heijden CA, Janssen PJCM, Strik JJTWA. Toxicology of gallates: A review and evaluation. Food and Chemical Toxicology. 1986 Oct 1;24(10–11):1067–70.

111. Bampidis V, Azimonti G, de Lourdes Bastos M, Christensen H, Dusemund B, Kos Durjava M, et al. Safety and efficacy of propyl gallate for all animal species. EFSA Journal. 2020 Apr 1;18(4):e06069.

112. Bond JA, Bolt HM. Review of the toxicology of styrene. Vol. 19, Critical Reviews in Toxicology. Informa Healthcare; 1989. p. 227–49.

113. Williams JH, Phillips TD, Jolly PE, Stiles JK, Jolly CM, Aggarwal D. Human aflatoxicosis in developing countries: A review of toxicology, exposure, potential health consequences, and interventions. Vol. 80, American Journal of Clinical Nutrition. American Society for Nutrition; 2004. p. 1106–22.

114. van Ravenzwaay B, Cunha GCP, Leibold E, Looser R, Mellert W, Prokoudine A, et al. The use of metabolomics for the discovery of new biomarkers of effect. Toxicology Letters. 2007 Jul 30;172(1–2):21–8.

115. Beger RD, Sun J, Schnackenberg LK. Metabolomics approaches for discovering biomarkers of drug-induced hepatotoxicity and nephrotoxicity. Toxicology and Applied Pharmacology. 2010 Mar 1;243(2):154–66.

116. Bonneau E, Tétreault N, Robitaille R, Boucher A, De Guire V. Metabolomics: Perspectives on potential biomarkers in organ transplantation and immunosuppressant toxicity. Clinical Biochemistry. 2016 Mar 1;49(4–5):377–84.

117. Dong H, Zhang A, Sun H, Wang H, Lu X, Wang M, et al. Ingenuity pathways analysis of urine metabolomics phenotypes toxicity of Chuanwu in Wistar rats by UPLC-Q-TOF-HDMS coupled with pattern recognition methods. Molecular BioSystems. 2012 Mar 1;8(4):1206–21.

118. Anthony ML, Beddell CR, Lindon JC, Nicholson JK. Studies on the comparative toxicity of S-(1,2-dichlorovinyl)-L-cysteine, S-(1,2-dichlorovinyl)-L-homocysteine and 1,1,2-trichloro-3,3,3-trifluoro-1-propene in the Fischer 344 rat. Archives of Toxicology. 1994 Dec;69(2):99–110.

119. Beckwith-Hall BM, Nicholson JK, Nicholls AW, Foxall PJD, Lindon JC, Connor SC, et al. Nuclear magnetic resonance spectroscopic and principal components analysis investigations into biochemical effects of three model hepatotoxins. Chemical Research in Toxicology. 1998;11(4):260–72.

120. Nicholls AW, Holmes E, Lindon JC, Shockcor JP, Farrant RD, Haselden JN, et al. Metabonomic investigations into hydrazine toxicity in the rat. Chemical Research in Toxicology. 2001;14(8):975–87.

121. Bundy JG, Davey MP, Viant MR. Environmental metabolomics: a critical review and future perspectives. Metabolomics. 2009 Mar 28;5(1):3–21.

122. Morrison N, Bearden D, Bundy JG, Collette T, Currie F, Davey MP, et al. Standard reporting requirements for biological samples in metabolomics experiments: Environmental context. Metabolomics. 2007 Sep 24;3(3):203–10.

123. Simpson MJ, McKelvie JR. Environmental metabolomics: New insights into earthworm ecotoxicity and contaminant bioavailability in soil. Vol. 394, Analytical and Bioanalytical Chemistry. Springer; 2009. p. 137–49.

124. Viant MR, Bearden DW, Bundy JG, Burton IW, Collette TW, Ekman DR, et al. International NMR-based environmental metabolomics intercomparison exercise. Environmental Science and Technology. 2009 Jan 1;43(1):219–25.

125. Sampaio BL, Edrada-Ebel R, Da Costa FB. Effect of the environment on the secondary metabolic profile of Tithonia diversifolia: A model for environmental metabolomics of plants. Scientific Reports. 2016 Jul 7;6(1):1–11.

126. Viant M. Metabolomics of aquatic organisms: the new omics on the block. Marine Ecology Progress Series. 2007 Mar 5;332:301–6.

127. García-Sevillano MÁ, García-Barrera T, Gómez-Ariza JL. Environmental metabolomics: Biological markers for metal toxicity. ELECTROPHORESIS. 2015 Sep 1;36(18):2348–65.

128. Nagato EG, Lankadurai BP, Soong R, Simpson AJ, Simpson MJ. Development of an NMR microprobe procedure for high-throughput environmental metabolomics of *Daphnia magna*. Magnetic Resonance in Chemistry. 2015 Sep 1;53(9):745–53.

129. Prüss-Ustün A, Vickers C, Haefliger P, Bertollini R. Knowns and unknowns on burden of disease due to chemicals: A systematic review. Vol. 10, Environmental Health: A Global Access Science Source. BioMed Central; 2011. p. 9.

130. WHO. Air pollution [Internet]. 2021 [cited 2022 Jan 21]. Available from: https://www.who.int/health-topics/air-pollution#tab=tab_1

131. WHO. Household air pollution and health [Internet]. 2021 [cited 2022 Jan 21]. Available from: https://www.who.int/news-room/fact-sheets/detail/household-air-pollution-and-health

132. Rockström J, Steffen W, Noone K, Persson Å, Chapin FS, Lambin EF, et al. A safe operating space for humanity. Vol. 461, Nature. Nature Publishing Group; 2009. p. 472–5.

133. REACH. EU REACH Regulation (EC) No 1907/2006 [Internet]. 2006 [cited 2021 Apr 8]. Available from: https://www.chemsafetypro.com/Topics/EU/REACH_Regulation_EC_No_1907_2006.html

134. UNEP. UNEP [Internet]. 2012 [cited 2021 Apr 6]. Available from: https://wedocs.unep.org/xmlui/handle/20.500.11822/8021

135. Morton PA, Fennell C, Cassidy R, Doody D, Fenton O, Mellander P-E, et al. A review of the pesticide MCPA in the land-water environment and emerging research needs. Wiley Interdisciplinary Reviews: Water. 2020 Jan 1;7(1):e1402.

136. Fernández-Casalderrey A, Ferrando MD, Andreu-Moliner E. Effect of sublethal concentrations of pesticides on the feeding behavior of Daphnia magna. Ecotoxicology and Environmental Safety. 1994 Feb 1;27(1):82–9.

137. Ieromina O, Peijnenburg WJGM, Musters CJM, Vijver MG. The effect of pesticides on the composition of aquatic macrofauna communities in field ditches. Basic and Applied Ecology. 2016 Mar 1;17(2):125–33.

138. Narra MR, Rajender K, Reddy RR, Murty US, Begum G. Insecticides induced stress response and recuperation in fish: Biomarkers in blood and tissues related to oxidative damage. Chemosphere. 2017 Feb 1;168:350–7.

139. Loganathan BG, Khim JS, Kodavanti PRS, Masunaga S, editors. Persistent Organic Chemicals in the Environment: Status and Trends in the Pacific Basin Countries I Contamination Status. Washington, DC: American Chemical Society; 2016. (ACS Symposium Series; vol. 1243).

140. Lyche JL, Grześ IM, Karlsson C, Nourizadeh-Lillabadi R, Aleström P, Ropstad E. Parental exposure to natural mixtures of persistent organic pollutants (POP) induced changes in transcription of apoptosis-related genes in offspring zebrafish embryos. Journal of Toxicology and Environmental Health, Part A. 2016 Aug 2;79(13–15):602–11.

141. Johnson AC, Donnachie RL, Sumpter JP, Jürgens MD, Moeckel C, Pereira MG. An alternative approach to risk rank chemicals on the threat they pose to the aquatic environment. Science of the Total Environment. 2017 Dec 1;599–600:1372–81.

142. Nasser F. Secreted protein eco-corona mediates uptake and impacts of polystyrene nanoparticles on Daphnia magna. Journal of Proteomics. 2016;137:45–51.

143. Suter II GW. Ecological Risk Assessment. CRC Press; 2016.

144. Hill IR, Heimbach F, Leeuwangh P, Matthiessen P. Freshwater Field Tests for Hazard Assessment of Chemicals. Freshwater Field Tests for Hazard Assessment of Chemicals. CRC Press; 2018.

145. Bhattacharya S, Zhang Q, Carmichael PL, Boekelheide K, Andersen ME. Toxicity Testing in the 21st Century: Defining New Risk Assessment Approaches Based on Perturbation of Intracellular Toxicity Pathways. Zhang B, editor. PLoS ONE. 2011 Jun 20;6(6):e20887.

146. Zhang QQ, Ying GG, Pan CG, Liu YS, Zhao JL. Comprehensive evaluation of antibiotics emission and fate in the river basins of China: Source analysis, multimedia modeling, and linkage to bacterial resistance. Environmental Science and Technology. 2015 Jun 2;49(11):6772–82.

147. Stahl G, Koschmann T, Suthers D. Computer Supported Collaborative Learning. Cambridge University Press; 2006. 409--426 p.

148. Lee Botts PM. Project MUSE - Evolution of the Great Lakes Water Quality Agreement [Internet]. 2005 [cited 2021 Apr 8]. Available from: https://muse.jhu.edu/book/40870

149. Matthies M, Solomon K, Vighi M, Gilman A, Tarazona J V. The origin and evolution of assessment criteria for persistent, bioaccumulative and toxic (PBT) chemicals and persistent organic pollutants (POPs). Vol. 18, Environmental Science: Processes and Impacts. Royal Society of Chemistry; 2016. p. 1114–28.

150. Conference of Plenipotentiaries. Stockholm Convention Overview [Internet]. 2001 [cited 2022 Feb 22]. Available from: http://www.pops.int/TheConvention/Overview/tabid/3351/Default.aspx

151. Schierow LJ. The Toxic Substances Control Act (TSCA): Implementation and new challenges. Toxic Substances Control Act: Summary, Challenges and Reform Proposals. 2012;19–54.

152. Kallis G, Butler D. The EU water framework directive: Measures and implications. Water Policy. 2001 Jun 1;3(2):125–42.

153. Wilkinson CF, Christoph GR, Julien E, Kelley JM, Kronenberg J, McCarthy J, et al. Assessing the Risks of Exposures to Multiple Chemicals with a Common Mechanism of Toxicity: How to Cumulate? Regulatory Toxicology and Pharmacology. 2000 Feb;31(1):30–43.

154. Carvalho RN, Ceriani L, Ippolito A, Lettieri T. Development of the first Watch List under the Environmental Quality Standards Directive 2015 Report EUR xxxxx xx. 2015.

155. Brack W, Dulio V, Ågerstrand M, Allan I, Altenburger R, Brinkmann M, et al. Towards the review of the European Union Water Framework management of chemical contamination in European surface water resources. Vol. 576, Science of the Total Environment. Elsevier B.V.; 2017. p. 720–37.

156. EFSA. International Frameworks Dealing with Human Risk Assessment of Combined Exposure to Multiple Chemicals. EFSA Journal. 2013 Jul 1;11(7):3313.

157. Wild CP. The exposome: From concept to utility. Vol. 41, International Journal of Epidemiology. Int J Epidemiol; 2012. p. 24–32.

158. Vineis P, Chadeau-Hyam M, Gmuender H, Gulliver J, Herceg Z, Kleinjans J, et al. The exposome in practice: Design of the EXPOsOMICS project. International Journal of Hygiene and Environmental Health. 2017 Mar 1;220(2):142–51.

159. Smith M, de la Rosa R, Daniels S. Using exposomics to assess cumulative risks and promote health: Using Exposomics to Assess Cumulative Risks. Environmental and Molecular Mutagenesis. 2015;56(9):715–23.

160. Holland N. Future of environmental research in the age of epigenomics and exposomics. Vol. 32, Reviews on Environmental Health. Walter de Gruyter GmbH; 2017. p. 45–54.

161. Jamin EL, Bonvallot N, Tremblay-Franco M, Cravedi JP, Chevrier C, Cordier S, et al. Untargeted profiling of pesticide metabolites by LC-HRMS: An exposomics tool for human exposure evaluation. Analytical and Bioanalytical Chemistry. 2014 Feb 28;406(4):1149–61.

162. Canali S. Big Data, epistemology and causality: Knowledge in and knowledge out in EXPOsOMICS. Big Data & Society. 2016 Dec 22;3(2):205395171666953.

163. Prior L, Manley D, Sabel CE. Biosocial health geography: New 'exposomic' geographies of health and place. Progress in Human Geography. 2019 Jun 7;43(3):531–52.

164. Cooke MS, Hu CW, Chang YJ, Chao MR. Urinary DNA adductomics – A novel approach for exposomics. Vol. 121, Environment International. Elsevier Ltd; 2018. p. 1033–8.

165. Dennis KK, Marder E, Balshaw DM, Cui Y, Lynes MA, Patti GJ, et al. Biomonitoring in the era of the exposome. Vol. 125, Environmental Health Perspectives. Public Health Services, US Dept of Health and Human Services; 2017. p. 502–10.

166. Lovett GM, Burns DA, Driscoll CT, Jenkins JC, Mitchell MJ, Rustad L, et al. Environmental monitoring GM Lovett et al. 254. Vol. 5, Frontiers in Ecology and the Environment. John Wiley & Sons, Ltd; 2007 Jun.

167. Lautenbach S, Seppelt R, Liebscher J, Dormann CF. Spatial and Temporal Trends of Global Pollination Benefit. Ollerton J, editor. PLoS ONE. 2012 Apr 26;7(4):e35954.

168. Smith KM, Loh EH, Rostal MK, Zambrana-Torrelio CM, Mendiola L, Daszak P. Pathogens, pests, and economics: Drivers of honey bee colony declines and losses. Vol. 10, EcoHealth. Springer New York LLC; 2013. p. 434–45.

169. Pirk CWW, Human H, Crewe RM, VanEngelsdorp D. A survey of managed honey bee colony losses in the Republic of South Africa - 2009 to 2011. Journal of Apicultural Research. 2014;53(1):35–42.

170. Lioy PJ, Rappaport SM. Exposure science and the exposome: An opportunity for coherence in the environmental health sciences. Vol. 119, Environmental Health Perspectives. National Institute of Environmental Health Sciences; 2011. p. a466.

171. NRC. Exposure Science in the 21st Century. Exposure Science in the 21st Century: A Vision and a Strategy. Washington, D.C.: National Academies Press; 2012. 1–210 p.

172. Huhn S, Escher BI, Krauss M, Scholz S, Hackermüller J, Altenburger R. Unravelling the chemical exposome in cohort studies: routes explored and steps to become comprehensive. Vol. 33, Environmental Sciences Europe. Springer Science and Business Media Deutschland GmbH; 2021. p. 17.

173. Emwas A-H, Roy R, McKay RT, Tenori L, Saccenti E, Gowda GAN, et al. NMR Spectroscopy for Metabolomics Research. Metabolites. 2019 Jun 27;9(7):123.

174. Olsen J v., Schwartz JC, Griep-Raming J, Nielsen ML, Damoc E, Denisov E, et al. A dual pressure linear ion trap orbitrap instrument with very high sequencing speed. Molecular and Cellular Proteomics [Internet]. 2009 Dec 1 [cited 2022 Mar 13];8(12):2759–69. Available from: http://www.mcponline.org/article/S1535947620339207/fulltext

175. Koulman A, Tapper BA, Fraser K, Cao M, Lane GA, Rasmussen S. High-throughput direct-infusion ion trap mass spectrometry: a new method for metabolomics. Rapid Communications in Mass Spectrometry. 2007 Feb 15;21(3):421–8.

176. Schmidt A, Karas M, Dülcks T. Effect of different solution flow rates on analyte ion signals in nano-ESI MS, or: When does ESI turn into nano-ESI? Journal of the American Society for Mass Spectrometry. 2003 May 1;14(5):492–500.

177. Falkenby LG, Such-Sanmartín G, Larsen MR, Vorm O, Bache N, Jensen ON. Integrated solid-phase extraction-capillary liquid chromatography (speLC) interfaced to ESI-MS/MS for fast characterization and quantification of protein and proteomes. Journal of Proteome Research. 2014 Dec 5;13(12):6169–75.

178. Bleakney W. A new method of positive ray analysis and its application to the measurement of ionization potentials in mercury vapor. Physical Review. 1929 Jul 1;34(1):157–60.

179. Suizdak G. An introduction to mass spectrometry ionization: An excerpt from The Expanding Role of Mass Spectrometry in Biotechnology, 2nd ed.; MCC Press: San Diego, 2005. Journal of the Association for Laboratory Automation. 2004 Apr 27;9(2):50–63.

180. Field FH. Chemical Ionization Mass Spectrometry. Accounts of Chemical Research. 1968 Jan 1;1(2):42–9.

181. Yamashita M, Fenn JB. Electrospray ion source. Another variation on the free-jet theme. The Journal of Physical Chemistry. 1984 Sep 1;88(20):4451–9.

182. Gaskell SJ. Electrospray: Principles and practice. Vol. 32, Journal of Mass Spectrometry. 1997. p. 677–88.

183. Albert MS, Dell RB, Winters RW. Quantitative displacement of acid-base equilibrium in metabolic acidosis. Annals of internal medicine. 1967;66(2):312–22.

184. Beynon RJ, Easterby JS. Buffer Solutions. Buffer Solutions. Taylor & Francis; 2004.

185. Konermann L, Ahadi E, Rodriguez AD, Vahidi S. Unraveling the Mechanism of Electrospray Ionization. Analytical Chemistry. 2013 Jan 2;85(1):2–9.

186. Taflin DC, Ward TL, Davis EJ. Electrified Droplet Fission and the Rayleigh Limit. 1988.

187. Vrijheid M, Slama R, Robinson O, Chatzi L, Coen M, van den Hazel P, et al. The human early-life exposome (HELIX): Project rationale and design. Environmental Health Perspectives. 2014.

188. Banerjee S, Mazumdar S. Electrospray Ionization Mass Spectrometry: A Technique to Access the Information beyond the Molecular Weight of the Analyte. International Journal of Analytical Chemistry. 2012;2012:1–40.

189. Iribarne J V., Thomson BA. On the evaporation of small ions from charged droplets. The Journal of Chemical Physics. 1976 Aug 28;64(6):2287–94.

190. Thomson BA, Iribarne J V. Field induced ion evaporation from liquid surfaces at atmospheric pressure. The Journal of Chemical Physics. 1979 Dec 25;71(11):4451–63.

191. Misner CW. The method of images in geometrostatics. Annals of Physics. 1963 Oct 1;24(C):102–17.

192. Ahadi E, Konermann L. Surface charge of electrosprayed water nanodroplets: A molecular dynamics study. Journal of the American Chemical Society. 2010 Aug 18;132(32):11270–7.

193. Konermann L. A Simple Model for the Disintegration of Highly Charged Solvent Droplets during Electrospray Ionization. Journal of the American Society for Mass Spectrometry. 2009 Mar 1;20(3):496–506.

194. Wilm M. Principles of electrospray ionization. Vol. 10, Molecular and Cellular Proteomics. American Society for Biochemistry and Molecular Biology; 2011.

195. Hogan CJ, Carroll JA, Rohrs HW, Biswas P, Gross ML. Combined charged residue-field emission model of macromolecular electrospray ionization. Analytical Chemistry. 2009 Jan 1;81(1):369–77.

196. Miller PE, Denton MB. The Quadrupole Mass Filter: Basic Operating Concepts. Vol. 63. UTC; 1986.

197. March RE. An introduction to quadrupole ion trap mass spectrometry. Vol. 32, Journal of Mass Spectrometry. John Wiley & Sons, Ltd; 1997. p. 351–69.

198. Zubarev RA, Makarov A. Orbitrap mass spectrometry. Analytical Chemistry. 2013 Jun 4;85(11):5288–96.

199. Perry RH, Cooks RG, Noll RJ. Orbitrap mass spectrometry: Instrumentation, ion motion and applications. Mass Spectrometry Reviews. 2008 Nov 1;27(6):661–99.

200. Thermo Fisher Scientific Inc. Thermo Fisher :: Orbitrap :: Q Exactive Plus [Internet]. 2021 [cited 2021 Jul 13]. Available from: https://planetorbitrap.com/q-exactive-plus

201. Rader CM, Maling GC. What Is the Fast Fourier Transform? Proceedings of the IEEE. 1967;55(10):1664–74.

202. Kelstrup CD, Jersie-Christensen RR, Batth TS, Arrey TN, Kuehn A, Kellmann M, et al. Rapid and deep proteomes by faster sequencing on a benchtop quadrupole ultra-high-field orbitrap mass spectrometer. Journal of Proteome Research. 2014 Dec 5;13(12):6187–95.

203. Manz A, Verpoorte E, Effenhauser CS, Burggraf N, Raymond DE, Widmer HM. Planar chip technology for capillary electrophoresis. Fresenius' Journal of Analytical Chemistry. 1994 Aug;348(8–9):567–71.

204. Manz A, Harrison DJ, Verpoorte EMJ, Fettinger JC, Paulus A, Lüdi H, et al. Planar chips technology for miniaturization and integration of separation techniques into monitoring systems. Capillary electrophoresis on a chip. Journal of Chromatography A. 1992 Feb 28;593(1–2):253–8.

205. Culbertson CT, Jacobson SC, Ramsey JM. Microchip devices for high-efficiency separations. Analytical Chemistry. 2000 Dec 1;72(23):5814–9.

206. Effenhauser CS, Manz A, Widmer HM. Glass Chips for High-Speed Capillary Electrophoresis Separations with Submicrometer Plate Heights. Analytical Chemistry. 1993 Oct 1;65(19):2637–42.

207. Raymond DE, Manz A, Widmer HM. Continuous Sample Pretreatment Using a Free-Flow Electrophoresis Device Integrated onto a Silicon Chip. Analytical Chemistry. 1994 Sep 1;66(18):2858–65.

208. Fujita H. Recent progress of microactuators and micromotors. Microsystem Technologies 1995 1:2. 1995;1(2):93–7.

209. Licklider L, Wang XQ, Desai A, Tai YC, Lee TD. A micromachined chip-based electrospray source for mass spectrometry. Analytical Chemistry. 2000 Jan 15;72(2):367–75.

210. Holmes PJ, Snell JE. A vapour etching technique for the photolithography of silicon dioxide. Microelectronics Reliability. 1966 Nov 1;5(4):337–41.

211. Nie Z, Kumacheva E. Patterning surfaces with functional polymers. Nature Materials. 2008;7(4):277–90.

212. Kim JS, Knapp DR. Microfabricated PDMS multichannel emitter for electrospray ionization mass spectrometry. Journal of the American Society for Mass Spectrometry. 2001;12(4):463–9.

213. Nakajima A. Design of hydrophobic surfaces for liquid droplet control. NPG Asia Materials 2011 3:5. 2011 May 19;3(5):49–56.

214. Wickremsinhe. A review of nanoelectrospray ionization applications for drug metabolism and pharmacokinetics. Current drug metabolism. 2006 Nov 28;7(8):913–28.

215. Hop CECA, Chen Y, Yu LJ. Uniformity of ionization response of structurally diverse analytes using a chip-based nanoelectrospray ionization source. Rapid Communications in Mass Spectrometry. 2005;19(21):3139–42.

216. Southam AD, Weber RJM, Engel J, Jones MR, Viant MR. A complete workflow for high-resolution spectral-stitching nanoelectrospray direct-infusion mass-spectrometry-based metabolomics and lipidomics. Nature Protocols. 2017 Jan 12;12(2):255–73.

217. Southam A, Cooper HJ, Viant MR. Dynamic Range and Mass Accuracy of Wide-Scan Direct Infusion Nanoelectrospray Fourier Transform Ion Cyclotron Resonance Mass Spectrometry-Based Metabolomics Increased by the Spectral Stitching Method. Analytical Chemistry. 2007 Jun 15;79(12):4595–602.

218. Southam AD, Weber RJM, Engel J, Jones MR, Viant MR. A complete workflow for high-resolution mass-spectrometry-based metabolomics and lipidomics. Nature Protocols. 2016;12(2):255–73.

219. Taylor N, Gavin A, Viant M. Metabolomics Discovers Early-Response Metabolic Biomarkers that Can Predict Chronic Reproductive Fitness in Individual Daphnia magna. Metabolites. 2018 Jul 23;8(3):42.

220. Grintzalis K, Lawson TN, Nasser F, Lynch I, Viant MR. Metabolomic method to detect a metabolite corona on amino-functionalized polystyrene nanoparticles. Nanotoxicology. 2019 Jul 3;13(6):783–94.

221. Zhang C. Rapid evolution magnifies the metabolomic plasticity to predation in a natural Daphnia population. Authorea Preprints. 2020 Feb 6;

222. Dekkers S, Williams TD, Zhang J, Zhou J, Vandebriel RJ, De La Fonteyne LJJ, et al. Multi-omics approaches confirm metal ions mediate the main toxicological pathways of metal-bearing nanoparticles in lung epithelial A549 cells. Environmental Science: Nano. 2018 Jun 14;5(6):1506–17.

223. Benkestock K. Electrospray Ionization Mass Spectrometry for Determination of Noncovalent Interactions in Drug Electrospray Ionization Mass Spectrometry for Determination of Noncovalent Interactions in Drug Discovery Kurt Benkestock Doctoral Thesis School of Chemical Sc. 2016.

224. Falegan O, Ball M, Shaykhutdinov R, Pieroraio P, Farshidfar F, Vogel H, et al. Urine and Serum Metabolomics Analyses May Distinguish between Stages of Renal Cell Carcinoma. Metabolites. 2017 Feb 3;7(1):6.

225. Sanchez LJ. Chemical fingerprinting of goat whole milk powder during shelf-life testing: NMR- and GC-MS-based metabolomics approach. University of Otago; 2020.

226. Xi B, Luo J, Gao Y qin, Yang X ling, Guo T fen, Li W hong, et al. Transcriptome–metabolome analysis of fatty acid of Bamei pork and Gansu Black pork in China. Bioprocess and Biosystems Engineering. 2020 Nov 7;1:3.

227. Misra BB, Das V, Landi M, Abenavoli MR, Araniti F. Short-term effects of the allelochemical umbelliferone on Triticum durum L. metabolism through GC–MS based untargeted metabolomics. Plant Science. 2020 Sep 1;298:110548.

228. Diez-Simon C, Mumm R, Hall RD. Mass spectrometry-based metabolomics of volatiles as a new tool for understanding aroma and flavour chemistry in processed food products. Vol. 15, Metabolomics. Springer New York LLC; 2019. p. 41.

229. Misra BB. Advances in High Resolution GC-MS Technology: Focus on Application of GC-Orbitrap-MS in Metabolomics and Exposomics for FAIR Practices. Analytical Methods. 2021;

230. Lindahl A, Heuchel · Rainer, Forshed J, Lehtiö · Janne, Löhr M, Nordström · Anders. Discrimination of pancreatic cancer and pancreatitis by LC-MS metabolomics. Metabolomics. 2017;13:61.

231. Jiao S, Nie M, Song H, Xu D, You F. Physiological responses to cold and starvation stresses in the liver of yellow drum (Nibea albiflora) revealed by LC-MS metabolomics. Science of the Total Environment. 2020 May 1;715:136940.

232. Garcia-Aloy M, Ulaszewska M, Franceschi P, Estruel-Amades S, Weinert CH, Tor-Roca A, et al. Discovery of Intake Biomarkers of Lentils, Chickpeas, and White Beans by Untargeted LC–MS Metabolomics in Serum and Urine. Molecular Nutrition & Food Research. 2020 Jul 22;64(13):1901137.

233. Zheng X, Xie X, Liu Y, Cong J, Fan J, Fang Y, et al. Deciphering the mechanism of carbon sources inhibiting recolorization in the removal of refractory dye: Based on an untargeted LC–MS metabolomics approach. Bioresource Technology. 2020 Jul 1;307:123248.

234. Depke T, Thöming JG, Kordes A, Häussler S, Brönstrup M. Untargeted LC-MS Metabolomics Differentiates Between Virulent and Avirulent Clinical Strains of Pseudomonas aeruginosa. Biomolecules. 2020 Jul 13;10(7):1041.

235. Shulaev V, Isaac G. Supercritical fluid chromatography coupled to mass spectrometry – A metabolomics perspective. Journal of Chromatography B. 2018 Aug 15;1092:499–505.

236. Hofstetter R, Link A, Fassauer GM. Enantioselective Supercritical Fluid Chromatography (SFC) for Chiral Metabolomics. Neuromethods. 2021;159:101–12.

237. van de Velde B, Guillarme D, Kohler I. Supercritical fluid chromatography – Mass spectrometry in metabolomics: Past, present, and future perspectives. Journal of Chromatography B. 2020 Dec 15;1161:122444.

238. Zhang X, Quinn K, Cruickshank-Quinn C, Reisdorph R, Reisdorph N. The application of ion mobility mass spectrometry to metabolomics. Vol. 42, Current Opinion in Chemical Biology. Elsevier Ltd; 2018. p. 60–6.

239. Mairinger T, Causon TJ, Hann S. The potential of ion mobility–mass spectrometry for non-targeted metabolomics. Vol. 42, Current Opinion in Chemical Biology. Elsevier Ltd; 2018. p. 9–15.

240. Paglia G, Astarita G. Metabolomics and lipidomics using traveling-wave ion mobility mass spectrometry. Nature Protocols. 2017 Apr 1;12(4):797–813.

241. Levy AJ, Oranzi NR, Ahmadireskety A, Kemperman RHJ, Wei MS, Yost RA. Recent progress in metabolomics using ion mobility-mass spectrometry. Vol. 116, TrAC - Trends in Analytical Chemistry. Elsevier B.V.; 2019. p. 274–81.

242. Lai Y. Exposomics by Liquid Chromatography-ESI-Drift-Tube Ion Mobility Quadrupole Time-of-Flight Mass Spectrometry: A Proof-of-Principle Study. 2014.

243. Metz TO, Baker ES, Schymanski EL, Renslow RS, Thomas DG, Causon TJ, et al. Integrating ion mobility spectrometry into mass spectrometry-based exposome measurements: what can it add and how far can it go? Bioanalysis. 2017 Jan 1;9(1):81–98.

244. Buko A, Cheng L, Gustev A, Feldman A. Abstract 3510: The metabolomic profile of urine from prostate cancer patients using capillary electrophoresis mass spectrometry (CEMS). In: Cancer Research. American Association for Cancer Research (AACR); 2020. p. 3510–3510.

245. Sugimoto M. Capillary electrophoresis–mass spectrometry of hydrophilic metabolomics. In: Neuromethods. Humana Press Inc.; 2021. p. 113–20.

246. Molnár I, Rieger HJ, Monks KE. Aspects of the “Design Space” in high pressure liquid chromatography method development. Journal of Chromatography A. 2010 May 7;1217(19):3193–200.

247. Fountain KJ. Application Note UPLC versus UHPLC: Comparison of Loading and Peak Capacity for Small Molecule Drugs. 2011.

248. Shu S, Kobayashi H, Okubo M, Sabarudin A, Butsugan M, Umemura T. Chemical anchoring of lauryl methacrylate-based reversed phase monolith to 1/16’ o.d. polyetheretherketone tubing. Journal of Chromatography A. 2012 Jun 15;1242:59–66.

249. Schellinger AP, Carr PW. Isocratic and gradient elution chromatography: A comparison in terms of speed, retention reproducibility and quantitation. Journal of Chromatography A. 2006 Mar 24;1109(2):253–66.

250. Hermo MP, Nemutlu E, Kir S, Barrón D, Barbosa J. Improved determination of quinolones in milk at their MRL levels using LC–UV, LC–FD, LC–MS and LC–MS/MS and validation in line with regulation 2002/657/EC. Analytica Chimica Acta. 2008 Apr 14;613(1):98–107.

251. Nielsen KF, Månsson M, Rank C, Frisvad JC, Larsen TO. Dereplication of Microbial Natural Products by LC-DAD-TOFMS. Journal of Natural Products. 2011 Nov 28;74(11):2338–48.

252. HPLC Help Centre. HPLC Help Center [Internet]. 2021 [cited 2021 Aug 2]. Available from: https://www.idex-hs.com/literature-tools/educational-materials/hplc-center/

253. Xiong C, Zhou X, Chen R, Zhang Y, Peng WP, Nie Z, et al. Characterization of column packing materials in high-performance liquid chromatography by charge-detection quadrupole ion trap mass spectrometry. Analytical Chemistry. 2011 Jul 1;83(13):5400–6.

254. Jandera P, Hájek T, Škeříková V, Soukup J. Dual hydrophilic interaction-RP retention mechanism on polar columns: Structural correlations and implementation for 2-D separations on a single column. Journal of Separation Science. 2010 Mar 1;33(6–7):841–52.

255. Noga S, Jandera P, Buszewski B. Retention Mechanism Studies of Selected Amino Acids and Vitamin B6 on HILIC Columns with Evaporative Light Scattering Detection. Chromatographia 2013 76:15. 2013 Jul 10;76(15):929–37.

256. Appelblad P. ZIC ®-HILIC Chromatography for Food Safety. 1997;(April).

257. Smith CA, Want EJ, 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 Feb 1;78(3):779–87.

258. Bolten CJ, Kiefer P, Letisse F, Portais JC, Wittmann C. Sampling for metabolome analysis of microorganisms. Analytical Chemistry. 2007 May 15;79(10):3843–9.

259. Koning W de, Dam K van. A method for the determination of changes of glycolytic metabolites in yeast on a subsecond time scale using extraction at neutral pH. Analytical Biochemistry. 1992;204(1):118–23.

260. Vuckovic D. Current trends and challenges in sample preparation for global metabolomics using liquid chromatography-mass spectrometry. Vol. 403, Analytical and Bioanalytical Chemistry. Springer; 2012. p. 1523–48.

261. Büscher JM, Czernik D, Ewald JC, Sauer U, Zamboni N. Cross-platform comparison of methods for quantitative metabolomics of primary metabolism. Analytical Chemistry. 2009 Mar 15;81(6):2135–43.

262. Rammouz R El, Létisse F, Durand S, Portais JC, Moussa ZW, Fernandez X. Analysis of skeletal muscle metabolome: Evaluation of extraction methods for targeted metabolite quantification using liquid chromatography tandem mass spectrometry. Analytical Biochemistry. 2010 Mar 15;398(2):169–77.

263. Bolten CJ, Kiefer P, Letisse F, Portais JC, Wittmann C. Sampling for metabolome analysis of microorganisms. Analytical Chemistry. 2007 May 15;79(10):3843–9.

264. Deprez S, Sweatman BC, Connor SC, Haselden JN, Waterfield CJ. Optimisation of collection, storage and preparation of rat plasma for 1H NMR spectroscopic analysis in toxicology studies to determine inherent variation in biochemical profiles. Journal of Pharmaceutical and Biomedical Analysis. 2002 Nov 7;30(4):1297–310.

265. Dunn WB, Ellis DI. Metabolomics: Current analytical platforms and methodologies. TrAC - Trends in Analytical Chemistry. 2005 Apr 1;24(4):285–94.

266. Mushtaq MY, Choi YH, Verpoorte R, Wilson EG. Extraction for metabolomics: Access to the metabolome. Vol. 25, Phytochemical Analysis. John Wiley and Sons Ltd; 2014. p. 291–306.

267. Maharjan R, Ferenci T. Global metabolite analysis: The influence of extraction methodology on metabolome profiles of Escherichia coli. Analytical Biochemistry. 2003 Feb 1;313(1):145–54.

268. Viant MR. Revealing the metabolome of animal tissues using 1H nuclear magnetic resonance spectroscopy. Methods in molecular biology (Clifton, NJ). 2007;358:229–46.

269. Keun HC, Athersuch TJ. Nuclear magnetic resonance (NMR)-based metabolomics. Methods in molecular biology (Clifton, NJ). 2011;708:321–34.

270. Chen D, Han W, Huan T, Li L, Li L. Effects of Freeze-Thaw Cycles of Blood Samples on High-Coverage Quantitative Metabolomics. Analytical Chemistry. 2020 Jul 7;92(13):9265–72.

271. Müller E, Berger R, Blass E, Sluyts D, Pfennig A. Liquid-Liquid Extraction. In: Ullmann's Encyclopedia of Industrial Chemistry. Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA; 2008.

272. Romdhane M, Gourdon C. Investigation in solid-liquid extraction: Influence of ultrasound. Chemical Engineering Journal. 2002 May 28;87(1):11–9.

273. Poole CF. New trends in solid-phase extraction. Vol. 22, TrAC - Trends in Analytical Chemistry. Elsevier; 2003. p. 362–73.

274. Hawthorne SB. ANALYTICAL-SCALE SUPERCRITICAL FLUID EXTRACTION. 1990.

275. Sitnikov DG, Monnin CS, Vuckovic D. Systematic Assessment of Seven Solvent and Solid-Phase Extraction Methods for Metabolomics Analysis of Human Plasma by LC-MS. Scientific Reports 2016 6:1. 2016 Dec 21;6(1):1–11.

276. Jensen SK. Improved Bligh and Dyer extraction procedure. Lipid Technology. 2008 Dec 1;20(12):280–1.

277. Bligh EG, Dyer WJ. A rapid method of total lipid extraction and purification. Canadian journal of biochemistry and physiology. 1959;37(8):911–7.

278. Chetwynd AJ, Dunn WB, Rodriguez-Blanco G. Collection and Preparation of Clinical Samples for Metabolomics. Advances in Experimental Medicine and Biology. 2017 Jan 1;965:19–44.

279. Krumsiek J, Mittelstrass K, Do KT, Stückler F, Ried J, Adamski J, et al. Gender-specific pathway differences in the human serum metabolome. Metabolomics. 2015 Dec 1;11(6):1815–33.

280. Broadhurst D, Goodacre R, Reinke SN, Kuligowski J, Wilson ID, Lewis MR, et al. Guidelines and considerations for the use of system suitability and quality control samples in mass spectrometry assays applied in untargeted clinical metabolomic studies. Metabolomics 2018 14:6. 2018 May 18;14(6):1–17.

281. Godzien J, Alonso-Herranz V, Barbas C, Armitage EG. Controlling the quality of metabolomics data: new strategies to get the best out of the QC sample. Metabolomics 2014 11:3. 2014 Jul 27;11(3):518–28.

282. Gika HG, Theodoridis GA, Earll M, Wilson ID. A QC approach to the determination of day-to-day reproducibility and robustness of LC–MS methods for global metabolite profiling in metabonomics/metabolomics. http://dx.doi.org/104155/bio12212. 2012 Oct 9;4(18):2239–47.

283. Dunn WB, Broadhurst DI, Edison A, Guillou C, Viant MR, Bearden DW, et al. Quality assurance and quality control processes: summary of a metabolomics community questionnaire. Metabolomics 2017 13:5. 2017 Mar 6;13(5):1–6.

284. Broadhurst D, Goodacre R, Reinke SN, Kuligowski J, Wilson ID, Lewis MR, et al. Guidelines and considerations for the use of system suitability and quality control samples in mass spectrometry assays applied in untargeted clinical metabolomic studies. Metabolomics 2018 14:6. 2018 May 18;14(6):1–17.

285. Lommen A. Metalign: Interface-driven, versatile metabolomics tool for hyphenated full-scan mass spectrometry data preprocessing. Analytical Chemistry. 2009 Apr 15;81(8):3079–86.

286. Lommen A, Kools HJ. MetAlign 3.0: Performance enhancement by efficient use of advances in computer hardware. Metabolomics. 2012 Aug 8;8(4):719–26.

287. Lapierre N, Alser M, Eskin E, Koslicki D, Mangul S. Metalign: Efficient alignment-based metagenomic profiling via containment min hash. Genome Biology. 2020 Sep 10;21(1):1–15.

288. Katajamaa M, Miettinen J, Orešič M. MZmine: Toolbox for processing and visualization of mass spectrometry based molecular profile data. Bioinformatics. 2006 Mar 1;22(5):634–6.

289. 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 Jul 23;11(1):1–11.

290. Olivon F, Grelier G, Roussi F, Litaudon M, Touboul D. MZmine 2 Data-Preprocessing to Enhance Molecular Networking Reliability. Analytical Chemistry. 2017 Aug 1;89(15):7836–40.

291. Xia J, Sinelnikov I V., Han B, Wishart DS. MetaboAnalyst 3.0-making metabolomics more meaningful. Nucleic Acids Research. 2015 Jul 1;43(W1):W251–7.

292. Smith CA, Want EJ, 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 Feb 1;78(3):779–87.

293. Tautenhahn R, Patti GJ, Rinehart D, Siuzdak G. XCMS online: A web-based platform to process untargeted metabolomic data. Analytical Chemistry. 2012 Jun 5;84(11):5035–9.

294. Forsberg EM, Huan T, Rinehart D, Benton HP, Warth B, Hilmers B, et al. Data processing, multi-omic pathway mapping, and metabolite activity analysis using XCMS Online. Nature Protocols. 2018 Apr 1;13(4):633–51.

295. Sana T, Fischer S, Tichy S. An LC / MS Metabolomics Discovery Workfl ow for Malaria-Infected Red Blood Cells Using Mass Profi ler Professional Software and LC-Triple Quadrupole MRM Confi rmation. Agilent. 2010;1–12.

296. Bhatia A, Sarma SJ, Lei Z, Sumner LW. UHPLC-QTOF-MS/MS-SPE-NMR: A Solution to the Metabolomics Grand Challenge of Higher-Throughput, Confident Metabolite Identifications. In: Methods in Molecular Biology. Humana Press Inc.; 2019. p. 113–33.

297. Züllig T, Zandl-Lang M, Trötzmüller M, Hartler J, Plecko B, Köfeler HC. A metabolomics workflow for analyzing complex biological samples using a combined method of untargeted and target-list based approaches. Metabolites. 2020 Sep 1;10(9):1–12.

298. Huan T, Forsberg EM, Rinehart D, Johnson CH, Ivanisevic J, Benton HP, et al. Systems biology guided by XCMS Online metabolomics. Nature Methods 2017 14:5. 2017 Apr 27;14(5):461–2.

299. Forsberg EM, Huan T, Rinehart D, Benton HP, Warth B, Hilmers B, et al. Data processing, multi-omic pathway mapping, and metabolite activity analysis using XCMS Online. Nature Protocols 2018 13:4. 2018 Mar 1;13(4):633–51.

300. Davidson RL, Weber RJM, Liu H, Sharma-Oates A, Viant MR. Galaxy-M: a Galaxy workflow for processing and analyzing direct infusion and liquid chromatography mass spectrometry-based metabolomics data. GigaScience. 2016 Dec 23;5(1):10.

301. Weber RJM, Viant MR. MI-Pack: Increased confidence of metabolite identification in mass spectra by integrating accurate masses and metabolic pathways. Chemometrics and Intelligent Laboratory Systems. 2010 Nov 15;104(1):75–82.

302. Davidson RL, Weber RJM, Liu H, Sharma-Oates A, Viant MR. Galaxy-M: a Galaxy workflow for processing and analyzing direct infusion and liquid chromatography mass spectrometry-based metabolomics data. GigaScience. 2016 Dec 23;5(1):10.

303. Martens L, Chambers M, Sturm M, Kessner D, Levander F, Shofstahl J, et al. mzML - A community standard for mass spectrometry data. Vol. 10, Molecular and Cellular Proteomics. Elsevier; 2011. p. R110.000133.

304. Adusumilli R, Mallick P. Data conversion with proteoWizard msConvert. In: Methods in Molecular Biology. Humana Press Inc.; 2017. p. 339–68.

305. McBean GA, Hantel M, editors. Interactions Between Global Climate Subsystems: The Legacy of Hann. Washington, D. C.: American Geophysical Union; 1993. (Geophysical Monograph Series; vol. 75).

306. Scigelova M, Hornshaw M, Giannakopulos A, Makarov A. Fourier transform mass spectrometry. Vol. 10, Molecular and Cellular Proteomics. American Society for Biochemistry and Molecular Biology; 2011.

307. Ebel A, Dreher W, Leibfritz D. Effects of zero-filling and apodization on spectral integrals in discrete Fourier-transform spectroscopy of noisy data. 2006;

308. Lee JE, Hwang GS, Lee CH, Hong YS. Metabolomics reveals alterations in both primary and secondary metabolites by wine bacteria. Journal of Agricultural and Food Chemistry. 2009 Nov 25;57(22):10772–83.

309. Chaleckis R, Meister I, Zhang P, Wheelock CE. Challenges, progress and promises of metabolite annotation for LC–MS-based metabolomics. Vol. 55, Current Opinion in Biotechnology. Elsevier Ltd; 2019. p. 44–50.

310. Schuster S, Fell DA, Dandekar T. A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. Nature Biotechnology. 2000;18(3):326–32.

311. Buchholz A, Hurlebaus J, Wandrey C, Takors R. Metabolomics: Quantification of intracellular metabolite dynamics. Biomolecular Engineering. 2002 Jun 1;19(1):5–15.

312. Wu ZE, Kruger MC, Cooper GJS, Poppitt SD, Fraser K. Tissue-specific sample dilution: An important parameter to optimise prior to untargeted lc-ms metabolomics. Metabolites. 2019 Jul 1;9(7):124.

313. Sumner LW, Amberg A, Barrett D, Beale MH, Beger R, Daykin CA, et al. Proposed minimum reporting standards for chemical analysis: Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). Metabolomics. 2007;

314. Schymanski EL, Jeon J, Gulde R, Fenner K, Ruff M, Singer HP, et al. Identifying small molecules via high resolution mass spectrometry: Communicating confidence. Vol. 48, Environmental Science and Technology. American Chemical Society; 2014. p. 2097–8.

315. Zhou S, Song Q, Tang Y, Naidong W. Critical Review of Development, Validation, and Transfer for High Throughput Bioanalytical LC-MS/MS Methods. Current Pharmaceutical Analysis. 2006 Mar 4;1(1):3–14.

316. Blaženović I, Kind T, Ji J, Fiehn O. Software tools and approaches for compound identification of LC-MS/MS data in metabolomics. Vol. 8, Metabolites. MDPI AG; 2018. p. 31.

317. Kind T, Okazaki Y, Saito K, Fiehn O. LipidBlast templates as flexible tools for creating new in-silico tandem mass spectral libraries. Analytical Chemistry. 2014 Nov 18;86(22):11024–7.

318. Blaženović I, Kind T, Torbašinović H, Obrenović S, Mehta SS, Tsugawa H, et al. Comprehensive comparison of in silico MS/MS fragmentation tools of the CASMI contest: Database boosting is needed to achieve 93% accuracy. Journal of Cheminformatics. 2017 May 25;9(1):32.

319. McEachran AD, Balabin I, Cathey T, Transue TR, Al-Ghoul H, Grulke C, et al. Linking in silico MS/MS spectra with chemistry data to improve identification of unknowns. Scientific Data. 2019 Dec 1;6(1):1–9.

320. Kind T, Liu K-H, Lee DY, DeFelice B, Meissen JK, Fiehn O. LipidBlast in silico tandem mass spectrometry database for lipid identification. Nature Methods 2013 10:8. 2013 Jun 30;10(8):755–8.

321. Da Silva RR, Dorrestein PC, Quinn RA. Illuminating the dark matter in metabolomics. Vol. 112, Proceedings of the National Academy of Sciences of the United States of America. National Academy of Sciences; 2015. p. 12549–50.

322. Kachman M, Habra H, Duren W, Wigginton J, Sajjakulnukit P, Michailidis G, et al. Deep annotation of untargeted LC-MS metabolomics data with Binner. Bioinformatics. 2020 Mar 1;36(6):1801–6.

323. Brown M, Dunn WB, Dobson P, Patel Y, Winder CL, Francis-Mcintyre S, et al. Mass spectrometry tools and metabolite-specific databases for molecular identification in metabolomics. Analyst. 2009 Jun 22;134(7):1322–32.

324. Mahieu NG, Patti GJ. Systems-Level Annotation of a Metabolomics Data Set Reduces 25 000 Features to Fewer than 1000 Unique Metabolites. Analytical Chemistry. 2017 Oct 3;89(19):10397–406.

325. Zhang P, Ang IL, Lam MMT, Wei R, Lei KMK, Zhou X, et al. Susceptibility to false discovery in biomarker research using liquid chromatography–high resolution mass spectrometry based untargeted metabolomics profiling. Clinical and Translational Medicine. 2021 Jun 27;11(6).

326. Alonso A, Julià A, Beltran A, Vinaixa M, Díaz M, Ibañez L, et al. AStream: An R package for annotating LC/MS metabolomic data. Bioinformatics. 2011 May 1;27(9):1339–40.

327. Tikunov YM, Laptenok S, Hall RD, Bovy A, de Vos RCH. MSClust: A tool for unsupervised mass spectra extraction of chromatography-mass spectrometry ion-wise aligned data. Metabolomics. 2012 Aug 15;8(4):714–8.

328. Kuhl C, Tautenhahn R, Böttcher C, Larson TR, Neumann S. CAMERA: An integrated strategy for compound spectra extraction and annotation of liquid chromatography/mass spectrometry data sets. Analytical Chemistry. 2012 Jan 3;84(1):283–9.

329. Silva RR, Jourdan F, Salvanha DM, Letisse F, Jamin EL, Guidetti-Gonzalez S, et al. ProbMetab: An R package for Bayesian probabilistic annotation of LC-MS-based metabolomics. Bioinformatics. 2014 May 1;30(9):1336–7.

330. Senan O, Aguilar-Mogas A, Navarro M, Capellades J, Noon L, Burks D, et al. CliqueMS: A computational tool for annotating in-source metabolite ions from LC-MS untargeted metabolomics data based on a coelution similarity network. Bioinformatics. 2019 Oct 15;35(20):4089–97.

331. Kuhl C, Tautenhahn R, Böttcher C, Larson TR, Neumann S. CAMERA: An integrated strategy for compound spectra extraction and annotation of liquid chromatography/mass spectrometry data sets. Analytical Chemistry. 2012 Jan 3;84(1):283–9.

332. Broeckling CD, Afsar FA, Neumann ⊥ S, Ben-Hur ∇ A, Prenni JE. RAMClust: A Novel Feature Clustering Method Enables Spectral-Matching-Based Annotation for Metabolomics Data. 2014;

333. Tikunov YM, Laptenok S, Hall RD, Bovy A, de Vos RCH. MSClust: a tool for unsupervised mass spectra extraction of chromatography-mass spectrometry ion-wise aligned data. Metabolomics. 2012 Aug 15;8(4):714–8.

334. Brown M, Goodacre R, Kell D, Dunn W. Automated workflows for accurate mass-based putative metabolite identification in LC/MS-derived metabolomic datasets. Bioinformatics. 2011 Apr 15;27(8):1108–12.

335. Rogers S, Scheltema RA, Girolami M, Breitling R. Probabilistic assignment of formulas to mass peaks in metabolomics experiments. Bioinformatics. 2009 Feb 15;25(4):512–8.

336. Daly R, Rogers S, Wandy J, Jankevics A, Burgess KEV, Breitling R. MetAssign: Probabilistic annotation of metabolites from LC-MS data using a Bayesian clustering approach. Bioinformatics. 2014 Apr 2;30(19):2764–71.

337. Silva RR, Jourdan F, Salvanha DM, Letisse F, Jamin EL, Guidetti-Gonzalez S, et al. ProbMetab: An R package for Bayesian probabilistic annotation of LC-MS-based metabolomics. Bioinformatics. 2014 May 1;30(9):1336–7.

338. Kachman M, Habra H, Duren W, Wigginton J, Sajjakulnukit P, Michailidis G, et al. Deep annotation of untargeted LC-MS metabolomics data with Binner. Bioinformatics. 2020 Mar 1;36(6):1801–6.

339. Mahieu NG, Patti GJ. Systems-Level Annotation of a Metabolomics Data Set Reduces 25 000 Features to Fewer than 1000 Unique Metabolites. Analytical Chemistry. 2017 Oct 3;89(19):10397–406.

340. Kind T, Fiehn O. Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. BMC Bioinformatics. 2007 Mar 27;8(1):1–20.

341. Salek RM, Steinbeck C, Viant MR, Goodacre R, Dunn WB. The role of reporting standards for metabolite annotation and identification in metabolomic studies. GigaScience. 2013;2(1):13.

342. Vinaixa M, Samino S, Saez I, Duran J, Guinovart JJ, Yanes O. A guideline to univariate statistical analysis for LC/MS-based untargeted metabolomics-derived data. Vol. 2, Metabolites. MDPI AG; 2012. p. 775–95.

343. Broadhurst DI, Kell DB. Statistical strategies for avoiding false discoveries in metabolomics and related experiments. Metabolomics. 2006 Dec 28;2(4):171–96.

344. Hochberg Y, Benjamini Y. More powerful procedures for multiple significance testing. Statistics in Medicine. 1990 Jul 1;9(7):811–8.

345. Storey JD, Tibshirani R. Statistical significance for genomewide studies. Proceedings of the National Academy of Sciences of the United States of America. 2003 Aug 5;100(16):9440–5.

346. Sedgwick P. Multiple significance tests: The Bonferroni correction. BMJ (Online). 2012 Jan 28;344(7841).

347. Saccenti E, Hoefsloot HCJ, Smilde AK, Westerhuis JA, Hendriks MMWB. Reflections on univariate and multivariate analysis of metabolomics data. Vol. 10, Metabolomics. Springer New York LLC; 2014. p. 361–74.

348. Jolliffe IT. PRINCIPAL COMPONENT ANALYSIS: A BEGINNER'S GUIDE — I. Introduction and application. Weather. 1990 Oct 1;45(10):375–82.

349. Weller SC. Correspondence Analysis. In: Encyclopedia of Biostatistics. John Wiley & Sons, Ltd; 2005.

350. Jolliffe IT, Cadima J, Jolliffe I, Statist PJ. On Relationships Between Uncentred And Column-Centred Principal Component Analysis ON RELATIONSHIPS BETWEEN UNCENTRED AND COLUMN-CENTRED PRINCIPAL COMPONENT ANALYSIS of Statistics 473. Vol. 25. 2009.

351. Huang D, Jiang F, Tong G, Zhou G. Scaled PCA: A New Approach to Dimension Reduction. SSRN Electronic Journal. 2019 May 15;

352. Chadeau-Hyam M, Ebbels TMD, Brown IJ, Chan Q, Stamler J, Huang CC, et al. Metabolic profiling and the metabolome-wide association study: Significance level for biomarker identification. Journal of Proteome Research. 2010 Sep 3;9(9):4620–7.

353. Lachenbruch PA, Goldstein M. Discriminant Analysis. Biometrics. 1979 Mar;35(1):69.

354. Westerhuis JA, Hoefsloot HCJ, Smit S, Vis DJ, Smilde AK, Velzen EJJ van, et al. Assessment of PLSDA cross validation. Metabolomics 2008 4:1. 2008 Jan 24;4(1):81–9.

355. Cudeck R, Browne MW. Cross-Validation Of Covariance Structures. http://dx.doi.org/101207/s15327906mbr1802_2. 2010 Jan 4;18(2):147–67.

356. Hawkins DM. The Problem of Overfitting. Journal of Chemical Information and Computer Sciences. 2003 Jan;44(1):1–12.

357. Domingos P. Bayesian Averaging of Classifiers and the Overfitting Problem. 2000;

358. Subramanian J, Simon R. Overfitting in prediction models – Is it a problem only in high dimensions? Contemporary Clinical Trials. 2013 Nov 1;36(2):636–41.

359. Bilbao I, Bilbao J. Overfitting problem and the over-training in the era of data: Particularly for Artificial Neural Networks. 2017 IEEE 8th International Conference on Intelligent Computing and Information Systems, ICICIS 2017. 2017 Jul 1;2018-January:173–7.

360. Ulrich EM, Sobus JR, Grulke C, Richard AM, Newton S, Strynar M, et al. Genesis and Study Design for EPA ' s Non-Targeted Analysis Collaborative Trial (ENTACT) Drivers for Exposure Science. 2018;

361. Dix DJ, Houck KA, Martin MT, Richard AM, Setzer RW, Kavlock RJ. The toxcast program for prioritizing toxicity testing of environmental chemicals. Toxicological Sciences. 2007 Jan 1;95(1):5–12.

362. Richard AM, Judson RS, Houck KA, Grulke CM, Volarath P, Thillainadarajah I, et al. ToxCast Chemical Landscape: Paving the Road to 21st Century Toxicology. Vol. 29, Chemical Research in Toxicology. American Chemical Society; 2016. p. 1225–51.

363. Ihaka R, Gentleman R. R: A Language for Data Analysis and Graphics. Journal of Computational and Graphical Statistics. 1996;5(3):299–314.

364. Bancos I, Taylor AE, Chortis V, Sitch AJ, Jenkinson C, Davidge-Pitts CJ, et al. Urine steroid metabolomics for the differential diagnosis of adrenal incidentalomas in the EURINE-ACT study: a prospective test validation study. The Lancet Diabetes & Endocrinology. 2020 Sep 1;8(9):773–81.

365. Southam AD, Pursell H, Frigerio G, Jankevics A, Weber RJM, Dunn WB. Characterization of Monophasic Solvent-Based Tissue Extractions for the Detection of Polar Metabolites and Lipids Applying Ultrahigh-Performance Liquid Chromatography–Mass Spectrometry Clinical Metabolic Phenotyping Assays. Journal of Proteome Research. 2020 Jan 1;20(1):831–40.

366. Fiehn O, Robertson D, Griffin J, vab der Werf M, Nikolau B, Morrison N, et al. The metabolomics standards initiative (MSI). Metabolomics. 2007 Sep 17;3(3):175–8.

367. Sud M, Fahy E, Cotter D, Brown A, Dennis EA, Glass CK, et al. LMSD: LIPID MAPS structure database. Nucleic Acids Research. 2007 Jan 1;35(SUPPL. 1):D527–32.

368. Racine J. RStudio. 2012;27(1):167–72.

369. Zala SM, Penn DJ. Abnormal behaviours induced by chemical pollution: A review of the evidence and new challenges. Vol. 68, Animal Behaviour. Academic Press; 2004. p. 649–64.

370. Persson LM, Breitholtz M, Cousins IT, De Wit CA, MacLeod M, McLachlan MS. Confronting unknown planetary boundary threats from chemical pollution. Environmental Science and Technology. 2013 Nov 19;47(22):12619–22.

371. Diamond ML, de Wit CA, Molander S, Scheringer M, Backhaus T, Lohmann R, et al. Exploring the planetary boundary for chemical pollution. Vol. 78, Environment International. Elsevier Ltd; 2015. p. 8–15.

372. Dodds JN, Alexander NLM, Kirkwood KI, Foster MR, Hopkins ZR, Knappe DRU, et al. From Pesticides to Per- And Polyfluoroalkyl Substances: An Evaluation of Recent Targeted and Untargeted Mass Spectrometry Methods for Xenobiotics. Vol. 93, Analytical Chemistry. American Chemical Society; 2021. p. 641–56.

373. Bousova K, Sarikaya E, Godula M, Martins C, George E. Fast Screening and Quantification of Pesticide Residues Using a Comprehensive LC-MS Solution: The Pesticide Explorer Collection-Standard Quantitation Key Words. 2015;

374. Rumsfeld D. Known and Unknown: A Memoir - Donald Rumsfeld - Google Books. 2011.

375. Ahammed Shabeer TP, Girame R, Utture S, Oulkar D, Banerjee K, Ajay D, et al. Optimization of multi-residue method for targeted screening and quantitation of 243 pesticide residues in cardamom (Elettaria cardamomum) by gas chromatography tandem mass spectrometry (GC-MS/MS) analysis. Chemosphere. 2018 Feb 1;193:447–53.

376. de Souza RM, Seibert D, Quesada HB, de Jesus Bassetti F, Fagundes-Klen MR, Bergamasco R. Occurrence, impacts and general aspects of pesticides in surface water: A review. Vol. 135, Process Safety and Environmental Protection. Institution of Chemical Engineers; 2020. p. 22–37.

377. Samsidar A, Siddiquee S, Shaarani SM. A review of extraction, analytical and advanced methods for determination of pesticides in environment and foodstuffs. Vol. 71, Trends in Food Science and Technology. Elsevier Ltd; 2018. p. 188–201.

378. Chibwe L, Davie-Martin CL, Aitken MD, Hoh E, Massey Simonich SL. Identification of polar transformation products and high molecular weight polycyclic aromatic hydrocarbons (PAHs) in contaminated soil following bioremediation. Science of the Total Environment. 2017 Dec 1;599–600:1099–107.

379. Helm JS, Nishioka M, Brody JG, Rudel RA, Dodson RE. Measurement of endocrine disrupting and asthma-associated chemicals in hair products used by Black women. Environmental Research. 2018 Aug 1;165:448–58.

380. Hollender J, Schymanski EL, Singer HP, Ferguson PL. Nontarget Screening with High Resolution Mass Spectrometry in the Environment: Ready to Go? 2017;

381. Eggen T, Moeder M, Arukwe A. Municipal landfill leachates: A significant source for new and emerging pollutants. Science of the Total Environment. 2010 Oct 1;408(21):5147–57.

382. Petrovic M, Eljarrat E, Lopez De Alda MJ, Barceló D. Endocrine disrupting compounds and other emerging contaminants in the environment: A survey on new monitoring strategies and occurrence data. Vol. 378, Analytical and Bioanalytical Chemistry. Springer; 2004. p. 549–62.

383. Lara-Marťn PA, Li X, Bopp RF, Brownawell BJ. Occurrence of alkyltrimethylammonium compounds in urban estuarine sediments: Behentrimonium as a new emerging contaminant. Environmental Science and Technology. 2010 Oct 1;44(19):7569–75.

384. Newton SR, Mcmahen RL, Sobus JR, Mansouri K, Williams AJ, Mceachran AD, et al. Suspect screening and non-targeted analysis of drinking water using point-of-use fi lters *. Environmental Pollution. 2018;234:297–306.

385. Kaserzon S, Kokkali V, Leerdam JA Van, Mueller JF, Pijnappels M, Reid MJ, et al. Exploring the Potential of a Global Emerging Contaminant Early Warning Network through the Use of Retrospective Suspect Screening with High-Resolution Mass Spectrometry. 2018;

386. Schymanski EL, Singer HP, Slobodnik J, Ipolyi IM, Oswald P, Krauss M, et al. Non-target screening with high-resolution mass spectrometry: Critical review using a collaborative trial on water analysis. Analytical and Bioanalytical Chemistry. 2015 Aug 1;407(21):6237–55.

387. Gago-Ferrero P, Schymanski EL, Bletsou AA, Aalizadeh R, Hollender J, Thomaidis NS. Extended Suspect and Non-Target Strategies to Characterize Emerging Polar Organic Contaminants in Raw Wastewater with LC-HRMS/MS. Environmental Science and Technology. 2015 Oct 20;49(20):12333–41.

388. Albergamo V, Schollée JE, Schymanski EL, Helmus R, Timmer H, Hollender J, et al. Nontarget screening reveals time trends of polar micropollutants in a riverbank filtration system. Environmental Science and Technology. 2019 Jul 2;53(13):7584–94.

389. Newton SR, Sobus JR, Ulrich EM, Singh RR, Chao A, McCord J, et al. Examining NTA performance and potential using fortified and reference house dust as part of EPA's Non-Targeted Analysis Collaborative Trial (ENTACT). Analytical and Bioanalytical Chemistry. 2020 Jul 1;412(18):4221–33.

390. Shahlaei M. Descriptor selection methods in quantitative structure-activity relationship studies: A review study. Vol. 113, Chemical Reviews. American Chemical Society; 2013. p. 8093–103.

391. Richard AM, Williams CLR. Distributed structure-searchable toxicity (DSSTox) public database network: A proposal. Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis. 2002 Jan 29;499(1):27–52.

392. Grulke CM, Williams AJ, Thillanadarajah I, Richard AM. EPA's DSSTox database: History of development of a curated chemistry resource supporting computational toxicology research. Computational Toxicology. 2019 Nov 1;12:100096.

393. EPA. Non-Targeted Analysis Workshop [Internet]. 2015 [cited 2021 May 11]. Available from: https://sites.google.com/site/nontargetedanalysisworkshop/

394. Thermo Fisher Scientific Inc. Thermo Scientific Orbitrap Elite Product Specifications. Thermo Scientific. 2011;

395. Kind T, Fiehn O. Metabolomic database annotations via query of elemental compositions: Mass accuracy is insufficient even at less than 1 ppm. BMC Bioinformatics. 2006 Apr 28;7:234.

396. Ulrich EM, Sobus JR, Grulke CM, Richard AM, Newton SR, Strynar MJ, et al. EPA's non-targeted analysis collaborative trial (ENTACT): genesis, design, and initial findings. Analytical and Bioanalytical Chemistry. 2019;411(4):853–66.

397. Billet K, Malinowska MA, Munsch T, Unlubayir M, Adler S, Delanoue G, et al. Semi-Targeted Metabolomics to Validate Biomarkers of Grape Downy Mildew Infection Under Field Conditions. Plants 2020, Vol 9, Page 1008. 2020 Aug 10;9(8):1008.

398. Wang X, Jones DR, Shaw TI, Cho J-H, Wang Y, Tan H, et al. Target-Decoy-Based False Discovery Rate Estimation for Large-Scale Metabolite Identification. Journal of Proteome Research. 2018 Jul 6;17(7):2328–34.

399. Scheubert K, Hufsky F, Petras D, Wang M, Nothias L-F, Dührkop K, et al. Significance estimation for large scale metabolomics annotations by spectral matching. Nature Communications 2017 8:1. 2017 Nov 14;8(1):1–10.

400. Matsuda F, Shinbo Y, Oikawa A, Hirai MY, Fiehn O, Kanaya S, et al. Assessment of Metabolome Annotation Quality: A Method for Evaluating the False Discovery Rate of Elemental Composition Searches. PLOS ONE. 2009 Oct 16;4(10):e7490.

401. Fujimoto GM, Kyle JE, Lee J-Y, Metz TO, Payne SH. A generalizable method for false-discovery rate estimation in mass spectrometry-based lipidomics. bioRxiv. 2020 Feb 19;2020.02.18.946483.

402. Jones DR, Wang X, Shaw T, Cho J-H, Chen P-C, Dey KK, et al. Metabolome Identification by Systematic Stable Isotope Labeling Experiments and False Discovery Analysis with a Target-Decoy Strategy. bioRxiv. 2016 Nov 26;089904.

403. Kind T, Fiehn O. Metabolomic database annotations via query of elemental compositions: Mass accuracy is insufficient even at less than 1 ppm. BMC Bioinformatics. 2006 Apr 28;7(1):1–10.

404. Barupal DK, Fiehn O. Generating the blood exposome database using a comprehensive text mining and database fusion approach. Environmental Health Perspectives. 2019 Sep 1;127(9).

405. Singh RR, Chao A, Phillips KA, Xia XR, Shea D, Sobus JR, et al. Expanded coverage of non-targeted LC-HRMS using atmospheric pressure chemical ionization: a case study with ENTACT mixtures. Analytical and Bioanalytical Chemistry. 2020 Aug 1;412(20):4931–9.

406. Veillon L, Huang Y, Peng W, Dong X, Cho BG, Mechref Y. Characterization of isomeric glycan structures by LC-MS/MS. Vol. 38, Electrophoresis. Wiley-VCH Verlag; 2017. p. 2100–14.

407. Koster RA, Alffenaar JWC, Greijdanus B, Vandernagel JEL, Uges DRA. Fast and highly selective LC-MS/MS screening for THC and 16 other abused drugs and metabolites in human hair to monitor patients for drug abuse. Therapeutic Drug Monitoring. 2014;36(2):234–43.

408. Eliasson M, Rännar S, Madsen R, Donten MA, Marsden-Edwards E, Moritz T, et al. Strategy for optimizing LC-MS data processing in metabolomics: A design of experiments approach. Analytical Chemistry. 2012 Aug 7;84(15):6869–76.

409. McLean C, Kujawinski EB. AutoTuner: High Fidelity and Robust Parameter Selection for Metabolomics Data Processing. Analytical Chemistry. 2020 Apr 21;92(8):5724–32.

410. 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 Oct 31;1029:50–7.

411. Gustafsson Å, Krais AM, Gorzsás A, Lundh T, Gerde P. Isolation and characterization of a respirable particle fraction from residential house-dust. Environmental Research. 2018 Feb 1;161:284–90.

412. Ortmayr K, Causon TJ, Hann S, Koellensperger G. Increasing selectivity and coverage in LC-MS based metabolome analysis. Vol. 82, TrAC - Trends in Analytical Chemistry. Elsevier B.V.; 2016. p. 358–66.

413. Cajka T, Fiehn • Oliver. Increasing lipidomic coverage by selecting optimal mobile-phase modifiers in LC-MS of blood plasma. Metabolomics. 2016;12.

414. Roca M, Alcoriza MI, Garcia-Cañaveras JC, Lahoz A. Reviewing the metabolome coverage provided by LC-MS: Focus on sample preparation and chromatography-A tutorial. Analytica Chimica Acta. 2021 Feb 22;1147:38–55.

415. Tada I, Tsugawa H, Meister I, Zhang P, Shu R, Katsumi R, et al. Creating a reliable mass spectral–retention time library for all ion fragmentation-based metabolomics. Metabolites. 2019 Nov 1;9(11):251.

416. Schrimpe-Rutledge AC, Codreanu SG, Sherrod SD, McLean JA. Untargeted Metabolomics Strategies—Challenges and Emerging Directions. Journal of the American Society for Mass Spectrometry. 2016 Dec 1;27(12):1897–905.

417. Kanehisa M, Furumichi M, Sato Y, Ishiguro-Watanabe M, Tanabe M. KEGG: Integrating viruses and cellular organisms. Nucleic Acids Research. 2021 Jan 8;49(D1):D545–51.

418. Ayers M. ChemSpider: The Free Chemical Database2012312 ChemSpider: The Free Chemical Database . URL: www.chemspider.com: Royal Society of Chemistry Last visited April 2012. Gratis . Reference Reviews. 2012 Sep 14;26(7):45–6.

419. Fisher Scientific Inc T. Q Exactive Plus Orbitrap LC-MS/MS System - Characterize, quantify and confirm with unmatched confidence. 2016.

420. Travis SC, Kordas K, Aga DS. Optimized workflow for unknown screening using gas chromatography high-resolution mass spectrometry expands identification of contaminants in silicone personal passive samplers. Rapid Communications in Mass Spectrometry. 2021 Apr 30;35(8):e9048.

421. Poole CF. New trends in solid-phase extraction. TrAC Trends in Analytical Chemistry. 2003 Jun 1;22(6):362–73.

422. Zheng SJ, Liu SJ, Zhu QF, Guo N, Wang YL, Yuan BF, et al. Establishment of Liquid Chromatography Retention Index Based on Chemical Labeling for Metabolomic Analysis. Analytical Chemistry. 2018 Jul 17;90(14):8412–20.

423. Creek DJ, Jankevics A, Breitling R, Watson DG, Barrett MP, Burgess KEV. Toward global metabolomics analysis with hydrophilic interaction liquid chromatography-mass spectrometry: Improved metabolite identification by retention time prediction. Analytical Chemistry. 2011 Nov 15;83(22):8703–10.

424. Bonini P, Kind T, Tsugawa H, Barupal DK, Fiehn O. Retip: Retention Time Prediction for Compound Annotation in Untargeted Metabolomics. Analytical Chemistry. 2020 Jun 2;92(11):7515–22.

425. Domingo-Almenara X, Guijas C, Billings E, Montenegro-Burke JR, Uritboonthai W, Aisporna AE, et al. The METLIN small molecule dataset for machine learning-based retention time prediction. Nature Communications. 2019 Dec 1;10(1):1–9.

426. Aicheler F, Li J, Hoene M, Lehmann R, Xu G, Kohlbacher O. Retention Time Prediction Improves Identification in Nontargeted Lipidomics Approaches. Analytical Chemistry. 2015 Aug 4;87(15):7698–704.

427. Witting M, Böcker S. Current status of retention time prediction in metabolite identification. Vol. 43, Journal of Separation Science. Wiley-VCH Verlag; 2020. p. 1746–54.

428. Schymanski EL, Kondić T, Neumann S, Thiessen PA, Zhang J, Bolton EE. Empowering large chemical knowledge bases for exposomics: PubChemLite meets MetFrag. Journal of Cheminformatics. 2021 Dec 1;13(1):1–15.

429. Munir H, Benjamin A, Allwood J, Dunn W, He S, Nash G, et al. A1.18 Mesenchymal stem cells lose their immuno-protective effects upon changes in their local microenvironment. Annals of the Rheumatic Diseases. 2016 Feb 1;75(Suppl 1):A8–A8.

430. Mumme K, Gray C, Reynolds CM, Vickers MH, Harrison CJ, Stanley JL, et al. Maternal-fetal hepatic and placental metabolome profiles are associated with reduced fetal growth in a rat model of maternal obesity. Metabolomics. 2016;12.

431. Allwood JW, Tomasz ·, Woznicki L, Xu Y, Foito A, Aaby · Kjersti, et al. Application of HPLC-PDA-MS metabolite profiling to investigate the effect of growth temperature and day length on blackcurrant fruit. Metabolomics. 2016;15:12.

432. D. Southam A, D. Haglington L, Lukáš Najdekr, Andris Jankevics, M. Weber RJ, B. Dunn W. Assessment of human plasma and urine sample preparation for reproducible and high-throughput UHPLC-MS clinical metabolic phenotyping. Analyst. 2020 Oct 12;145(20):6511–23.

433. Heazell AEP, Bernatavicius G, Warrander L, Brown MC, Dunn WB. A Metabolomic Approach Identifies Differences in Maternal Serum in Third Trimester Pregnancies That End in Poor Perinatal Outcome. Reproductive Sciences 2012 19:8. 2012 Dec 30;19(8):863–75.

434. Bajhaiya AK, Dean AP, Driver T, Trivedi DK, Rattray NJW, Allwood JW, et al. High-throughput metabolic screening of microalgae genetic variation in response to nutrient limitation. Metabolomics 2015 12:1. 2015 Nov 13;12(1):1–14.

435. Chetwynd AJ, Ogilvie LA, Nzakizwanayo J, Pazdirek F, Hoch J, Dedi C, et al. The potential of nanoflow liquid chromatography-nano electrospray ionisation-mass spectrometry for global profiling the faecal metabolome. Journal of Chromatography A. 2019 Aug 30;1600:127–36.

436. D'Elia R V., Goodchild SA, Winder CL, Southam AD, Weber RJM, Stahl FM, et al. Multiple metabolic pathways are predictive of ricin intoxication in a rat model. Metabolomics 2019 15:7. 2019 Jul 3;15(7):1–15.

437. Considine EC;, Khashan AS;, Kenny LC, Considine EC, Khashan AS. Title Screening for preterm birth: Potential for a metabolomics biomarker Screening for Preterm Birth: Potential for a Metabolomics Biomarker Panel. 2019;

438. Pearson K. Determination of the Coefficient of Correlation Author (s): Karl Pearson Published by : American Association for the Advancement of Science Stable URL : http://www.jstor.org/stable/1635783. Science. 1909;30(757):23–5.

439. Spearman C. Footrule_for_measuring_correla.pdf. 1906.

440. Mihaleva V V., Vorst O, Maliepaard C, Verhoeven HA, de Vos RCH, Hall RD, et al. Accurate mass error correction in liquid chromatography time-of-flight mass spectrometry based metabolomics. Metabolomics. 2008 Jun 22;4(2):171–82.

441. Ma X, Chong L, Tian R, Shi R, Hu TY, Ouyang Z, et al. Identification and quantitation of lipid C=C location isomers: A shotgun lipidomics approach enabled by photochemical reaction. Proceedings of the National Academy of Sciences. 2016 Mar 8;113(10):2573–8.

442. Zhang W, Zhao PX. Quality evaluation of extracted ion chromatograms and chromatographic peaks in liquid chromatography/mass spectrometry-based metabolomics data. BMC Bioinformatics. 2014 Dec 21;15(S11):S5.

443. Schönbrodt FD, Perugini M. At what sample size do correlations stabilize? Journal of Research in Personality. 2013 Oct 1;47(5):609–12.

444. Kirwan JA, Broadhurst DI, Davidson RL, Viant MR. Characterising and correcting batch variation in an automated direct infusion mass spectrometry (DIMS) metabolomics workflow. Analytical and Bioanalytical Chemistry 2013 405:15. 2013 Mar 1;405(15):5147–57.

445. Han W, Li L. Evaluating and minimizing batch effects in metabolomics. Mass Spectrometry Reviews. 2020;

446. Camacho D, De La Fuente A, Mendes P. The origin of correlations in metabolomics data. 2004;

447. Nyamundanda G, Gormley IC, Fan Y, Gallagher WM, Brennan L. MetSizeR: selecting the optimal sample size for metabolomic studies using an analysis based approach. BMC Bioinformatics 2013 14:1. 2013 Nov 21;14(1):1–8.

448. Mahieu NG, Spalding JL, Gelman SJ, Patti GJ. Defining and Detecting Complex Peak Relationships in Mass Spectral Data: The Mz.unity Algorithm. Analytical Chemistry. 2016 Sep 20;88(18):9037–46.

449. Sindelar M, Patti GJ. Chemical Discovery in the Era of Metabolomics. Vol. 142, Journal of the American Chemical Society. American Chemical Society; 2020. p. 9097–105.

450. Li CR, Hou XH, Xu YY, Gao W, Li P, Yang H. Manual annotation combined with untargeted metabolomics for chemical characterization and discrimination of two major crataegus species based on liquid chromatography quadrupole time-of-flight mass spectrometry. Journal of Chromatography A. 2020 Feb 8;1612:460628.

451. Wang Y, Ma L, Zhang M, Chen M, Li P, He C, et al. A Simple Method for Peak Alignment Using Relative Retention Time Related to an Inherent Peak in Liquid Chromatography-Mass Spectrometry-Based Metabolomics. Journal of Chromatographic Science. 2019 Jan 1;57(1):9–16.

452. Nothias LF, Petras D, Schmid R, Dührkop K, Rainer J, Sarvepalli A, et al. Feature-based molecular networking in the GNPS analysis environment. Nature Methods. 2020 Sep 1;17(9):905–8.

453. Chong J, Soufan O, Li C, Caraus I, Li S, Bourque G, et al. MetaboAnalyst 4.0: Towards more transparent and integrative metabolomics analysis. Nucleic Acids Research. 2018 Jul 2;46(W1):W486–94.

454. Rappaport SM, Barupal DK, Wishart D, Vineis P, Scalbert A. The blood exposome and its role in discovering causes of disease. Environmental Health Perspectives. 2014;122(8):769–74.

455. Lawson TN, Weber RJM, Jones MR, Chetwynd AJ, Rodríguez-Blanco G, Guida R Di, et al. msPurity: Automated Evaluation of Precursor Ion Purity for Mass Spectrometry-Based Fragmentation in Metabolomics. Analytical Chemistry. 2017 Feb 21;89(4):2432–9.

456. Nemkov T, Stefanoni D, Bordbar A, Issaian A, Palsson BO, Dumont LJ, et al. Blood donor exposome and impact of common drugs on red blood cell metabolism. JCI Insight. 2021 Feb 8;6(3).

457. Park J, Kim J, Kim E, Kim WJ, Won S. Prenatal lead exposure and cord blood DNA methylation in the Korean Exposome Study. Environmental Research. 2021 Apr 1;195:110767.

458. Southam AD, Lange A, Al-Salhi R, Hill EM, Tyler CR, Viant MR. Distinguishing between the metabolome and xenobiotic exposome in environmental field samples analysed by direct-infusion mass spectrometry based metabolomics and lipidomics. Metabolomics [Internet]. 2014 Dec 1 [cited 2022 Mar 13];10(6):1050–8. Available from: https://link.springer.com/article/10.1007/s11306-014-0693-3

459. Southam AD, Khanim FL, Hayden RE, Constantinou JK, Koczula KM, Michell RH, et al. Drug redeployment to kill leukemia and lymphoma cells by disrupting SCD1-mediated synthesis of monounsaturated fatty acids. Cancer Research [Internet]. 2015 Jun 15 [cited 2022 Mar 13];75(12):2530–40. Available from: https://aacrjournals.org/cancerres/article/75/12/2530/599703/Drug-Redeployment-to-Kill-Leukemia-and-Lymphoma

460. Molloy JC, Sommer U, Viant MR, Sinkins SP. Wolbachia modulates lipid metabolism in Aedes albopictus mosquito cells. Applied and Environmental Microbiology [Internet]. 2016 May 1 [cited 2022 Mar 13];82(10):3109–20. Available from: https://journals.asm.org/doi/abs/10.1128/AEM.00275-16

461. Taylor NS, White TA, Viant MR. Defining the Baseline and Oxidant Perturbed Lipidomic Profiles of Daphnia magna. Metabolites 2017, Vol 7, Page 11 [Internet]. 2017 Mar 15 [cited 2022 Mar 13];7(1):11. Available from: https://www.mdpi.com/2218-1989/7/1/11/htm

462. Technical Publications L. EASY-ETD and EASY-IC Ion Sources User Guide For the Orbitrap Tribrid Series Mass Spectrometer 80000-97515 Revision A. 2018;

463. Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Research. 1999 Jan 1;27(1):29–34.

464. Pence HE, Williams A. ChemSpider: An Online Chemical Information Resource. Journal of Chemical Education. 2010 Nov 1;87(11):1123–4.

465. Rosli MAF, Mediani A, Azizan KA, Baharum SN, Goh HH. UPLC-TOF-MS/MS-Based Metabolomics Analysis Reveals Species-Specific Metabolite Compositions in Pitchers of Nepenthes ampullaria, Nepenthes rafflesiana, and Their Hybrid Nepenthes × hookeriana. Frontiers in Plant Science. 2021 Apr 21;12.

466. Dunphy LJ, Grimes KL, Wase N, Kolling GL, Papin JA. Untargeted Metabolomics Reveals Species-Specific Metabolite Production and Shared Nutrient Consumption by Pseudomonas aeruginosa and Staphylococcus aureus. mSystems. 2021 Jun 29;6(3).

467. Qi J, Li K, Shi Y, Li Y, Dong L, Liu L, et al. Cross-Species Comparison of Metabolomics to Decipher the Metabolic Diversity in Ten Fruits. Metabolites 2021, Vol 11, Page 164. 2021 Mar 12;11(3):164.

468. Barupal DK, Fiehn O. Generating the Blood Exposome Database Using a Comprehensive Text Mining and Database Fusion Approach. Environmental Health Perspectives. 2019 Sep 1;127(9).

469. Wernisch S, Pennathur S. Evaluation of coverage, retention patterns, and selectivity of seven liquid chromatographic methods for metabolomics. Analytical and Bioanalytical Chemistry. 2016 Sep 1;408(22):6079–91.

470. Walsby-Tickle J, Gannon J, Hvinden I, Bardella C, Abboud MI, Nazeer A, et al. Anion-exchange chromatography mass spectrometry provides extensive coverage of primary metabolic pathways revealing altered metabolism in IDH1 mutant cells. Communications Biology 2020 3:1. 2020 May 20;3(1):1–12.

471. Mairinger T, Causon TJ, Hann S. The potential of ion mobility–mass spectrometry for non-targeted metabolomics. Current Opinion in Chemical Biology. 2018 Feb 1;42:9–15.

472. Zhou Z, Tu J, Zhu ZJ. Advancing the large-scale CCS database for metabolomics and lipidomics at the machine-learning era. Current Opinion in Chemical Biology. 2018 Feb 1;42:34–41.

473. Causon TJ, Si-Hung L, Newton K, Kurulugama RT, Fjeldsted J, Hann S. Fundamental study of ion trapping and multiplexing using drift tube-ion mobility time-of-flight mass spectrometry for non-targeted metabolomics. Analytical and Bioanalytical Chemistry. 2019 Sep 1;411(24):6265–74.

474. Paglia G, Smith AJ, Astarita G. Ion mobility mass spectrometry in the omics era: Challenges and opportunities for metabolomics and lipidomics. Mass Spectrometry Reviews. 2021;

475. Haug K, Salek RM, Conesa P, Hastings J, De Matos P, Rijnbeek M, et al. MetaboLights—an open-access general-purpose repository for metabolomics studies and associated meta-data. Nucleic Acids Research. 2013 Jan 1;41(D1):D781–6.

476. Sud M, Fahy E, Cotter D, Azam K, Vadivelu I, Burant C, et al. Metabolomics Workbench: An international repository for metabolomics data and metadata, metabolite standards, protocols, tutorials and training, and analysis tools. Nucleic Acids Research. 2016 Jan 4;44(D1):D463–70.

477. McLean C, Kujawinski EB. AutoTuner: High fidelity, robust, and rapid parameter selection for metabolomics data processing. bioRxiv. 2019 Oct 21;812370.

478. 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 Jan 1;16(1):1–12.