In vitro assessment of hepatotoxicity by metabolomics : a review

In vitro assessment of hepatotoxicity by metabolomics: a review

Matthias Cuykx[1,2*], Robim M Rodrigues[2], Kris Laukens[3,4], Tamara Vanhaecke[2$], Adrian Covaci[1*$]

[1]- Toxicological centre, University of Antwerp, Universiteitsplein 1, 2610 Wilrijk, Belgium

[2]- *In Vitro* Toxicology and Dermato-Cosmetology, Vrije Universiteit Brussel, Laarbeeklaan 103, 1090 Jette, Belgium

[3]- Department of Mathematics & Computer Science, University of Antwerp, Middelheimlaan 1, 2020 Antwerp, Belgium

[4]- Biomedical informatics network Antwerpen (biomina), University of Antwerp, Middelheimlaan 1, 2020 Antwerp, Belgium

* - corresponding authors: Matthias.Cuykx@uantwerpen.be; Adrian.Covaci@uantwerpen.be

fax: +32 3 265 27 22

[$]equal last authors

Abstract

Omics technologies, and in particular metabolomics, have received increasing attention during the assessment of hepatotoxicity *in vitro*. However, at present, a consensus on good metabolomics practices has yet to be reached. Therefore, in this review, a range of experimental approaches, applied methodologies and data processing workflows are compared and critically evaluated. Experimental designs among the studies are similar, reporting the use of primary hepatocytes or hepatic cell lines as the most frequently used cell sources. Experiments are usually conducted in short time frames (<48 h) at sub-toxic dosages. Applied sample preparations are protein precipitation or Bligh-and-Dyer extraction. Most analytical platforms rely on chromatographic separations with mass spectrometric detection using high resolution instruments. Untargeted metabolomics was typically used to allow the simultaneous detection of several classes of the metabolome, including endogenous metabolites that are not initially linked to toxicity. This non-biased detection platform is a valuable tool for generating hypothesis-based mechanistic research. The most frequently reported metabolites that are altered under toxicological impulses are alanine, lactate and proline, which are often correlated. Other unspecific biomarkers of hepatotoxicity *in vitro* are the downregulation of choline, glutathione, and 3-phosphoglycerate. Disruptions on the Krebs cycle are associated with increased glutamate, tryptophan and valine. Phospholipid alterations are described in steatosis, lipo-apoptosis and oxidative stress. Although there is a growing trend towards quality control, data analysis procedures do often not follow good contemporary metabolomics practices, which include feature filtering, false-discovery rate correction and reporting the confidence of metabolite annotation. The currently annotated biomarkers can be used to identify hepatotoxicity in general and provide, to a certain extent, a tool for mechanistic distinction.

Keywords: metabolomics, *in vitro*, liver, hepatotoxicity, drug induced liver injury (DILI)

43 **List of abbreviations**:

2D	Two-dimensional
3D	Three-dimensional
3-PG	3 phospho-glycerate
3R	Replace/Reduce/Refine- principle for animal testing
AA	Amino acids
AM	Accurate mass
AMP	Adenosine monophosphate
ANOVA	Analysis of variance
AOP	Adverse outcome pathway
ATP	Adenosine triphosphate
BPA	Bisphenol A
CE	Capillary electrophoresis
DG	Diacyl glycerol
DILI	Drug-induced liver injury
DMBA	Dimethylbenz(a)anthracene
DNA	Deoxyribosyl nucleic acids
EtOH	Ethanol
FDR	False discovery rate
FT-ICR	Fourier-transform ion cyclotron resonance mass spectrometry
GC	Gas chromatography
GSH	Reduced glutathione
GSSG	Oxidised glutathione
HBCD	Hexabromocyclododecane
HCA	Hierarchical clustering analysis
HCV	Hepatitis C virus
HILIC	Hydrophilic liquid interaction chromatography
HR	High resolution
IC	Inhibitory concentration
IP	Ion pairing
LC	Liquid chromatography
LIT	Linear ion trap mass spectrometer
LPC	Lysophosphatidylcholine
MeOH	Methanol
MIE	Molecular Initiating Event
MOA	Mode of Action
MS	Mass spectrometry
MSI	Metabolomics Standards Initiative
MTT	3-(4,5-dimethylthiazol-2-yl)- 2,5-diphenyltetrazolium bromide
MV	Missing values
MWU	Mann-Whitney U
N/A	Not applicable
N/R	Not reported
NADPH	Nicotinamide adenine dinucleotide phosphate
NMR	Nucleic magnetic resonance
NOAEL	No observed adverse effect level
NRU	Neutral red uptake
PAH	Poly-aromatic hydrocarbons
PBS	Phosphate buffer saline
PC	Phosphatidyl choline
PCA	Principal component analysis

PE	Phosphatidyl ethanolamine
PEP	Phospho-enol pyruvate
PG	Prostaglandins
PLS	Partial least squares analysis
PQN	Probabilistic quotient normalisation
Q	Quadrupole mass spectrometer
QC	Quality control
QMTLS	Quantum mechanical total line shape fitting
RNA	Ribosyl nucleic acids
ROS	Reactive oxygen species
RP	Reversed phase
SAM	s-adenosyl methionine
TCA	Tricarboxylic acid cycle
TCDD	Tetrachlorodibenzo-p-dioxin
TG	Triacyl glycerol
TOF	Time-of flight mass spectrometer
UPLC	Ultra-performant liquid chromatography
VIP	Variable importance of projection

Highlights:

- Metabolomics is expanding into *in vitro* hepatotoxicity research.
- Primary hepatocytes and hepatic cell lines are the main *in vitro* systems used.
- Sample treatments rely on protein precipitation and Bligh and Dyer extraction.
- Most toxicity markers are linked to dysregulation of energy and lipid homeostasis.
- Data reporting is inconsistent and metabolite identification lacks transparency.

1. Introduction

Since the dawn of the new millennium, the field of Toxicology is reorienting itself, investigating toxicological problems using new approaches. After initial observations like lethal dose determination, differences in body weight and macroscopic and histopathological "end-point" alterations in the organism, new principles aim for a more mechanistic interpretation of inflicted hazards[1–4]. The mechanisms involved in adverse reactions to xenobiotics are complex, often involving multiple steps and/or organ systems. Understanding these processes rather than describing their final outcome is an important trend in the assessment of potential toxicological hazards.

The liver is a multifunctional organ, its high metabolic rate and biotransformation capacity make it very vulnerable for both metabolically induced, as well as drug-induced liver injury (DILI)[5,6]. Non-alcoholic fatty liver disease (NAFLD), which is often seen as a manifestation of the metabolic syndrome, is of growing concern and has a prevalence of up to 25 % in Western coutries[7]. Although DILI contribution to the NAFLD population is limited, many drugs have an important impact on steatosis progression[8]. DILI poses a very important threat and is of primary concern in safety testing of new chemical entities, since it is the most frequently reported specific target of toxicity[9]. The occurrence of DILI in drug development results in high attrition rates and even post-market drug withdrawal[10]. Therefore, good *in vitro* methodologies could help identifying potential hepatotoxic compounds early during the drug development process, reducing health hazards and improving cost-efficiencies for the industry [2,11,12].

The rise of -*omics* techniques during the last decades, such as genomics (1990's), transcriptomics and proteomics (2000's) and metabolomics (2010's), allows an unprecedented systematic screening of different genes, proteins and metabolites. These approaches generate large amounts of data that represent the complex regulations involved in the endogenous metabolism[4,13]. The integration of biochemical profiles provides a perspective on the distortions in the cell, and allows to generate hypotheses regarding the Modes of Action (MOA) involved in the toxicological outcome[3].

Metabolomics is defined as the study of the biochemical profile at the small molecule level in an organism[14]. Because the metabolome is the most downstream level in the biomolecular organisation of a system, metabolomics fingerprints are very dynamic and alterations may be induced even by tiny initial external triggers[4,13]. The metabolome fingerprint provides a rich dataset that not only could help to characterise the toxicological outcome, but also the potential mechanisms involved[4].

Although metabolomics was initially mainly applied to *in vivo* studies to search for valuable mechanistic biomarkers, its applications are expanding towards *in vitro* methodologies[15]. Especially for compound safety evaluation, a shift can be observed from *in vivo* to *in vitro* toxicology. Indeed, the use of animals in research is a topic of debate in many countries[16]. Ethical concerns, inter-species differences, high

costs, and time investments of such *in vivo* experiments are important reasons to shift from animal methods to *in vitro* and *in silico* alternatives[16–20]. Other advantages of *in vitro* models are their reduced complexity, allowing specific focus on single mechanisms. A general low variability within each model also reduces noise and biases, while improving throughput [4,20].

This review focuses on *in vitro* investigations of hepatotoxicity using experimental metabolomics set-ups. The earliest articles date from 2011, the metabolomic research in *in vitro* hepatotoxicity is gaining momentum starting from 2014, with 17 of the 28 papers published after 2015. With the growing interest in metabolomics as a novel approach for the *in vitro* investigation of hepatotoxicity, this is an ideal moment to review the current methods, comparing them with good metabolomics practices, and to reveal which metabolites are consistently altered among the different articles and could therefore serve as potential markers of hepatotoxicity.

2. Study designs

2.1 Hepatic *in vitro* systems

Careful design of the experiment is vital for a successful metabolomics application. Due to the inherent characteristics of each cell type and culture platform, there are similarities and differences in comparison to the *in vivo* situation. The selection of the cell type is vital to the outcome of the experiment and its implications should carefully be considered when designing the experiment. Two comparative articles show the different metabolic capacity of various cells. Kim et al.[21] investigated the difference between human adult and fetal primary hepatocytes, observing differences in the amino acid metabolism, the glycolysis and the citric acid cycle, and in the urea cycle. Their interpretations state that the fetal hepatocytes rely on glycolytic activity because of mitochondrial inactivation, in contrast to mature cells which show gluconeogenetic activity while they rely on lipogenic beta-oxidation[21]. Sugiyama et al.[22] compared steatogenic principle of HCV-infected Huh-7 cells, resulting in increased inflammatory intermediates, sugar and amino-acid uptake and inhibited cholesterol synthesis.

The HepG2 cell line is the most popular cell model, being applied in ±35 % of the studies[2,15,23–30], followed by primary hepatocytes (human (10 %)[21,31], rodent (10 %)[32–34], fish (10 %)[35–37]). The list is completed with L02 (10 %)[38–40], HepaRG (10 %)[41–43], Huh-7 (10 %)[22,44], murine hepatoma Hepa1c1c7[45] and V79-fibroblasts[13] (Table 1). Primary human hepatocytes are considered to be the golden standard for *in vitro* hepatotoxicity studies, but they have a high inter-donor variability, low availability and limited lifespan and metabolic stability *in vitro*[6,20]. Rodent hepatic cell cultures offer a cheap alternative to scarce primary human hepatocytes, but they have the same scientific drawbacks as their human

counterparts. Furthermore, the use of rodent hepatocytes does not account for inter-species differences[20]. Søfteland and Olsvik exposed fish hepatocytes to pesticides, bisphenol A, polyaromated hydrocarbons and genistein in the context of agricultural applications[35–37]. Although inter-species differences between fish and human hepatic metabolism are even more pronounced than between rat and human primary hepatocytes, information resulting from *in vitro* experiments with animal primary hepatocytes can still be useful when common metabolic pathways are investigated.

Immortalised cell lines, such as the HepG2, L02, Huh-7 and HepaRG cell lines, are alternative human first-screening tools because of their low variabity[20]. However, HepG2 and Huh-7 have a poor biotransformation capacity and are not stable in long-term 2D cultures[20,46]. HepaRG cells have biotransformation capacity and are genetically more stable than the other hepatic cell lines[47,48]. Their ability to differentiate both in hepatocyte-like and biliary cells allow to detect cholestatic effects[2,49,50]. The down-side of this cell line is the fact that only undifferentiated cells proliferate and hepatic differentiation is necessary, thereby introducing additional variation that in turn reduces the reproducibility of metabolomics measurements[2].

In vitro metabolomics experiments need a relative large amount of cell material when compared to most trivial cellular assays. Cell cultures are usually cultured in 6 well plates (35 %); but larger surfaces, such as T75 flasks and petri dishes (30 %) are also popular. An alternative is the 12-well plate, which is used to economise on culture material (21 %). 2D monolayer cultures are the standard procedures and facilitate cell scraping as harvesting method. Adjustments to improve cell harvesting exist, such as the use of carrier glasses[51], chamber slides[52] and membrane carriers[13,53], but they are not widely applied[2,13,43]. Advanced culture platforms, such as biochips and 3D cultures, can be incorporated to improve the *in vitro – in vivo* simulation, but their applications in metabolomics are still limited[20,23,31]. Vorrink et al.[31] investigated the stability of primary human hepatocytes in 3D spheroid cultures and observed stable levels of ornithine, cholic acids, acyl carnitines and citric acid as well as improved reduced and oxidised glutathione (GSH/GSSG) ratios.

Cell supernatants are analysed by Snouber et al.[23], Wang et al.[33], Ramirez et al.[2], Vorrink et al.[31] and Toyoda et al.[34]. The main advantage of sampling cell supernatant is the non-invasive procedure, which gives the opportunity to investigate the evolution of metabolome of the same cell-unit at different time-points. However, the supernatant does not necessarily represent all intra-cellular alterations, and extrapolations to intra-cellular mechanisms should be considered carefully. Ramirez et al.[2] observed that the investigated toxicological MOAs invoked more specific changes in the intracellular metabolome than in the extracellular medium.

2.2 Experimental design

A comprehensive overview of all experimental designs is given in Table 1. The standard *in vitro* metabolomics setting involves the study of a single compound which is administered in different concentrations and time points. García-Cañaveras et al.[26,54] and Ramirez et al.[2] reported multiple experiments including different hepatotoxicants with the ultimate goal to build classifier models able to distinguish hepatoxic compounds according to different modes of action. To increase the biochemical impact of steatotic agents on the lipid metabolism, García-Cañaveras et al.[55] applied lipid loading before exposure, resulting in more distinct metabolic profiles. Søfteland and Olsvik[37] applied a full factorial exposure design of polycyclic aromatic hydrocarbons (PAH) and pesticides to determine synergistic hepatotoxicity in fish. Brown et al. [41], Seeßle et al.[44] and Meissen et al.[29] investigated dietary impacts on the hepatic metabolome by exposure to palmitate/oleate and glucose/fructose, respectively.

Exposure times

The time-frame in *in vitro* metabolomics experiments is short, with 24 h (48 %) and 48 h (44 %) accounting for the majority of time-points. Shorter time frames have also been considered (24 %), often in combination with multiple time-point investigations to observe time-related differences, with the final end-point at 24 h or 48 h[13,25,29,30,45]. A longer time frame (72 h) is considered in a minority of the applications (12 %) to evaluate sub-chronic effects or to increase the impact of the MOA on the metabolome, resulting in more pronounced differences[30,42,43].

Testing concentrations

Determination of the concentrations to be used is performed by viability assays, which is reported in 85 % of the applications concerning xenobiotic toxicity. The most popular assay (88 %) applies a bioconversion of 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) or water soluble tetrazolium salts (WST-1 and CCK-8) through the mitochondrial activity and NADPH flux as a projection of cell viability[56,57]. However, these bioconversion assays are biased towards lower concentrations since these conversions are linked to mitochondrial metabolism pathways, which can be specifically inhibited by the tested compound, thereby providing a biased indication of cell viability. The neutral Red Uptake assay, which is based on lysosomal storage of the cationic dye, and cell adhesion/impedance assays are alternative method often applied when the energy metabolism is suspected to be part of the toxicant's MOA[58,59].

The determined maximal concentrations to be applied are at sub-cytotoxic levels, usually determined as the Inhibitory Concentration (IC) 20, i.e. the concentration where the exposed concentration leads

to 20 % less viable cells. Although low IC-values are considered to be sub-cytotoxic, the toxicological insult can be considered to be significant. As a result, early-onset alterations can be missed in such investigations. Toxic concentrations, such as the IC30 by Søfteland et al.[37] or Chatterjee et al.[27] and the IC50 by Balcke et al.[13] or Yu et al.[28], have such a strong impact that only alterations related to general toxicity can be observed, impeding mechanistic fingerprinting. Combining a high and low dose exposure is recommended to investigate the trends of multiple markers of toxicity through the toxicological process; a second exposure at lower dosages (e.g. one tenth of the reported IC-concentration) is often applied to investigate metabolic alterations before any cell death occurs.

Number of replicates

The number of replicates is often limited, with 48 % of the experiments reporting more than 6 replicates for a defined exposure window. Increasing the number of replicates allows for a more accurate reflection of the true distribution of the metabolites within a tested class, improving the statistical power of the experiment. Also the use of cell lines instead of primary cells improves statistical power as they originate from the same donor and hence are less prone to biological variability[20,60]. Even with the reduced biological variation, enough replicates should still be considered to account for the analytical/technical variation. According to Sumner et al.[61] and Martano et al.[51], at least 3 replicates and preferably >5 replicates are required for reliable *in vitro* metabolomics applications.

Randomisation

Four publications explicitly mention a randomisation of their samples during processing and/or LC-MS analysis[15,26,42,43]. Although the effect of population bias (e.g. gender, age, diet etc.) is not prevalent in *in vitro* research using cell lines or pooled primary cells, biases can still be introduced through processing order (e.g. seeding, degradation during sample processing, the loss of sensitivity because of time-dependent source contamination). The randomisation of all samples during exposure, sample preparation and LC-MS injection is an important step in untargeted metabolomics to remove potential correlations between exposure groups and non-controlled experimental factors and to prevent technical/instrumental biases[62,63]. Randomisation can be performed blindly over the entire dataset or can be constrained to ensure an equal distribution of the exposure categories over the entire handling process[64].

2.3 Combination with other assays

Metabolomics can be applied as a stand-alone approach to obtain information regarding metabolic pathways involved in the MOA of the toxicological insult. Yet, the combination of metabolomics with

other assays generates additional information, which improves the biochemical interpretation. Functional assays (23 %) and stainings (12 %) can directly link histological and biochemical outcomes with the obtained information from the metabolomics experiment[22,23,27,34,39,41,44].

The combination of metabolomics with other omics to provide a systems biology perspective is a powerful approach, and the technique is often applied to *in vitro* hepatotoxicity. Metabolomics is mostly combined with transcriptomics (30 %), followed by genomics (15 %) and proteomics (4 %)[21,22,45,24,25,27,30,31,35,37,44]. A recent report from Rodrigues et al. combined transcriptomics, proteomics and metabolomics *in vitro* data to identify drug-induced cellular responses involved in cholestasis and validate a previously developed AOP for drug-induced cholestasis[50]. Alterations in both the transcriptome and the metabolome are a good indication of the pathways involved in the toxicological mechanism. Combining the outcomes of both omics levels with pathway analysis approaches, the metabolic accumulations and depletions can be interpreted together with the cellular responses to these differences in the intracellular homeostasis. This strategy assists in revealing the different metabolic pathways involved in the MOA, improving the final biological interpretation.

3. Analytical methods

Sample preparation procedures and data-acquisition methods are crucial parts of the analytical workflow. Many methods have been developed for different metabolomics applications in order to extract and separate the metabolome as efficiently as possible[60,65–67]. The extraction of intracellular metabolites has been widely investigated, comparing culture carriers[51–53] and extraction solvents[65,67–69]. The chemical space of metabolites is vast and there is consensus that one optimal method for all metabolites does not exist[52,70]. Although generic analytical methods are often applied, the development of a protocol according to state-of-the-art recommendations yields higher results concerning metabolic coverage, accuracy and precision than the application of a standard non-tailored protocol[70,71].

The discussion of the method section of all articles is divided in sections describing the pre-analytical (4.1) and instrumental phase (4.2) followed by the data-processing and statistical analysis (4.3) and identification strategies (4.4). A comprehensive overview is shown in Table 2.

3.1 Quenching and extraction

All protocols perform a cold quenching method. Liquid nitrogen is applied in 30 % of the cases in literature, followed by methanol addition (19 %). Eukaryotic cells only have a phospholipid cell membrane that can easily be disrupted during cold solvent extraction. Hence, harsh conditions such as acidic and high temperature extractions are not necessary. Snap freezing and cold extractions are

indeed often applied methods since they prevent heat-related degradation, while they denature proteins, inhibiting enzymatic catalysis[69,70,72].

Metabolite extraction can be generally divided in the "protein precipitation" method (41 %) and the "Bligh-and-Dyer" based methods (52 %), each of them having their own advantages and disadvantages.

The "protein precipitation" method generally adds a polar organic solvent (e.g. methanol) to precipitate proteins, followed by evaporation of the supernatant and reconstitution of the metabolites in an appropriate solvent for the instrumental analysis. These methods are straightforward, introduce less analytical variation and improve throughput. However, the lack of selectivity can introduce downstream problems, such as solubility issues (lipids vs polar) and instrumental issues, such as ion suppression[73]. Protein precipitation methods should therefore only be considered for untargeted one-platform experiments.

The variations on the Bligh-and-Dyer method use water, methanol or ethanol, and a non-polar solvent such as chloroform, dichloromethane or methyl-tert-butyl ether[68,74,75]. Liquid-liquid extraction provides a rough separation between polar and non-polar metabolites[74,75], but it does not introduce much variation and removes potential interfering, high-abundant metabolites, such as phospholipids. This extraction is nonetheless non-specific, which makes it very interesting for semi-targeted platforms.

Application methods using Bligh-and-Dyer methods are often combined with multi-platform or semi-targeted/targeted approaches. Balcke et al.[13], Kim et al.[21], Toyoda et al.[34] and Liu et al.[38] removed interfering phospholipids, increasing sensitivity and resolution for the polar metabolite analysis using Capillary-Electrophoresis- MS and NMR applications. Ruiz-Aracama et al.[15], Søfteland et al.[37], Van den Eede et al.[42], Van den Hof et al.[30], García-Cañaveras et al.[26], Ramirez et al.[2] and Cuykx et al.[43] obtained two fractions which can be analysed using dedicated methods for both fractions, reducing interferences within one analysis and increasing coverage and signal reproducibility over the entire platform.

3.2 Analytical platforms

Chromatography hyphenated to mass spectrometry is the most applied method, being mentioned in 70 % of the literature. Versatility, high sensitivity and broad metabolic coverage are the main reasons why these analytical instruments are often applied in metabolomics. Reversed Phase (RP) methods using C18 stationary phases in combination with mobile phases consisting of water (A) and methanol or acetonitrile (B), with additional formic acid, are often described because of the non-specific

retention mechanism, making it a preferred choice for non-expert research groups. Although this method does separate many metabolites, the single platform is not-optimal, lacking retention for polar metabolites and lacking resolution for many non-polar metabolites[26,66,76,77]. Hydrophilic-liquid-interaction-chromatography (HILIC), capillary electrophoresis and ion-pairing-reversed phase chromatography are solutions often described in metabolomic applications to increase the separation of polar metabolites[78–83], but they are not widely implemented in *in vitro* hepatic research[13,21,22,26,34,66]. GC-MS applications can be considered for lipidomics, especially when the desired analyses focus on Fatty-acid compositions[2,29,35,36,41,44].

In untargeted applications, the eluting mobile phase is coupled to high-resolution/accurate mass spectrometers (HR/AM-MS) to improve separation of metabolites in a second (m/z) dimension. Time-of-flight (TOF) instruments have the best trade-off between scan-time and resolution (10 000-40 000), making them ideal instruments to be applied for UPLC applications[15,21,22,26,28,29,32,34,39,42]. Ion Trap/orbitrap instruments have a lower scan-speed, but provide higher resolution (>100 000) which is an advantage for high molecular weight (glycomics/lipidomics) metabolomics fields[77]. Targeted approaches do not rely on accurate mass and m/z resolution; triple quadrupole configurations, more robust platforms for quantitation, are therefore often applied for targeted metabolomics[13,30,31,33,44,84].

Although Nuclear Magnetic Resonance (NMR) was initially more employed, the current LC-MS systems provide exceptional improvements regarding sensitivity, resolution and repeatability[14,70,77,85]. Nonetheless, NMR still plays an important, complementary role in metabolomics research since it is fast, reproducible and sensitive for polar compounds. This complementarity is acknowledged and applied by Ruiz-Aracama et al.[15], Søfteland et al.[37] and Van den Hof et al.[30]. The quest for good resolution is also present in NMR applications, with all groups using >400 MHz instrumentation.

3.3 Data processing

Metabolomics platforms generate large and complex data. As in other omics disciplines, there is a typical discrepancy between a relatively low number of samples (and replicates) and a large number of variables or features[63]. In the downstream data processing, there is an extensive range of algorithms that can be used, each with an even larger number of parameters that need to be set. The software implementations range from vendor software, which uses often proprietary methods, to academic (often open-source) platforms, which provide more methodological transparency. Both have their advantages and limitations: vendor software provides ready-to-use solutions that typically allow to perform a complete analysis in the same software environment. However, besides the lack of transparency, this comes at the cost of limited flexibility. Open-source tools provide better transparency and can, with sufficient programming expertise, be further adapted to the needs of the

application. However, the learning-curve is steeper and the choice of initial parameters requires optimisation.

In both cases, every data processing step, from peak picking, to missing value imputations, filter steps, scaling and normalisation, has an important impact on the outcome of the experiment[86,87]. Properly reporting all the steps and parameters in the data processing workflow is essential and full transparency determines the reproducibility of the experiment. Standards regarding reporting have been established by the community during several years and their implementation is now encouraged by the metabolomics community and publishers[61–63,71,77,88,89].

Different peak picking algorithms exist to convert the raw data into variables representing the measured metabolites which are further grouped to create a data-matrix for statistical analysis. This conversion is mostly performed through the vendor software (61 %), which is a straightforward choice when minimal interaction is preferred[2,13,43,21,23,25,27,31,34,35,42]. Biological data are often skewed and, unless absolute quantification is performed, quantitative data are usually log-transformed, normalised and mean-centred. To correct for unequal sample amounts or systematic quantitative biases, normalisation procedures are applied[71,90,91]. In the literature, popular techniques were based on protein quantity (30 %) or on data-driven transformation (44 %). Normalisation based on protein content is considered a pre-analytical normalisation strategy to account for differences in the amount of cell material extracted, anticipating for eventual cell death related to the exposed groups. Silva et al.[92] determined the optimal normalisation to be based on DNA material, since protein determination is less reliable for normalisation of each sample within the dataset. However, normalising by cellular material has the risk of introducing additional bias. Examples of normalisation bias include the co-extraction of dead cells in 3D cultures and the distortion of the DNA/RNA/protein expression by the xenobiotic's MOA, e.g. through the induction of Cytochrome P450 families. Alternative data-transforming approaches omit the need for additional experiments as they are entirely data-driven. Signal correction is then either dependent on the direct signal acquired from the instrument, such as the total ion count or phospholipid signal, or it is based on the data to be used in the multivariate analysis[90,91,93]. According to Ejigu et al.[90], data-driven normalisation standards are preferred over internal standard normalisation in untargeted approaches since the behaviour of the internal standard is not identical or similar to all metabolites.

Data filters (based on e.g. variance, missing values) are mentioned by several studies[24,28,35,37,42,43]. The removal of unreliable features because of high variance or many missing values is a form of quality assurance (QA) which reduces the noise and sparsity of the final dataset, preventing biases in the further statistical analysis[87,89]. The filtering of the dataset has benefits for subsequent univariate and

multivariate analysis, as it reduces errors and increases the contribution of the exposure to the overall variance in the dataset, respectively[62,63,89].

A commonly applied additional filtering step consists of the subtraction of blank samples [2,26,31,39,42,43]. Blank subtraction is an easy approach to remove features that are related only to the platform and sample preparation. Such features do not represent the metabolome of the hepatic cultures and could otherwise introduce unnecessary variables and complexity, obscuring the biological results.

Since absolute validation is impossible in untargeted analyses, QC is essential in untargeted metabolomics applications and should always be implemented in any metabolomics experiment[62,94]. QC is often performed by pooling aliquots of all samples, generating a QC-pool representing the average of the dataset[62]. This QC-approach can be implemented in multiple levels of the analytical workflow, ranging from system equilibration to analytical variability assessment and identification[62,66]. This form of quality control has been reported by 20 % of the reviewed literature [2,26,28,37,42,43]

The detailed description of all processing steps and QC implementations was limited prior to 2015. With the growing awareness of the need for good metabolomics practices, the scientific excellence of metabolomics in *in vitro* hepatoxicity assays is expected to continue to increase.

3.4 Statistical interpretations

Most statistical interpretations perform a classification based on multivariate analysis and apply univariate tests to select the significant biomarkers.

Univariate tests, such as t-tests and ANOVA, are straightforward tests to compare the intensities of a single variable between different groups. The outcome and interpretation of this statistical approach is very clear, but the implementation of univariate tests for hundreds of variables inflates the risk of false positive features[63,95]. Corrections for these false discoveries can be applied (such as Bonferroni and Benjamini-Hochberg correction) and according to Kim et al.[95], the best compromise is achieved with the Benjamini-Hochberg correction. The Benjamini-Hochberg correction is less conservative than the standard Bonferroni correction, which is considered too strict since many variables are not independent of each other[95]. Univariate tests are applied in the majority of the articles (86 %), but only 40 % of them explicitly report the implementation of a multiple testing correction.

Principal component analysis (PCA) is the most used unsupervised multivariate technique. PCA projects the maximum variance of a multi-dimensional space in principal components and summarizes the dataset in a limited number of components. Being unsupervised, this technique does not explicitly account for class-based separations and is therefore mostly used as an exploratory technique. In the

reviewed literature, PCA is often applied to interpret if the dependence of all variables projects a biologically relevant pattern, e.g. to distinguish different levels of toxicity.

A commonly used supervised variant of PCA is the (orthogonal) Partial Least Squares (PLS) analysis, which describes the most variance in order to distinguish the experimental classes, unravelling the metabolic patterns most important for the classification. The metabolites that have a large impact on the projection are selected using either S-plots or Variable Importance of Projection (VIP)-values. Supervised multivariate approaches are prone to overfitting to irrelevant or noisy features. Cross-validation procedures use a fraction of the data-set as an independent test-set to estimate the performance of the model and are essential to evaluate the discriminative capacity of the revealed markers[63,88].

3.5 Identification

Annotation of the significant signals in untargeted metabolomics datasets is the current bottleneck in metabolomics research. NMR identification relies on small differences in proton coupling, related to the chemical structure of the metabolite. Each proton in a given molecule experiences a specific influence to the magnetic field depending of the electron-shielding, generating a specific NMR-spectrum. High-resolution instrumentation allows an accurate measurement of the shift of each proton in the spectrum, reducing the number of potential candidates for each measured proton-shift. The information of common metabolites is available in in-house libraries or in public metabolomics databases, such as the Human Metabolome Database[96].

Mass spectrometric strategies imply MS-only and ion fragmentation strategies (MS/MS) to identify the accurate m/z of precursor and fragments ions related to the formula and the structure of the metabolite, respectively. The manual annotation of all signals is a tedious task, which is often supported by the use of databases, such as Lipidmaps, Metlin and the Human Metabolome Database in combination with in-house built MS libraries[96–98]. *In silico* fragmentation software assist the annotation of MS/MS based strategies by predicting the potential fragments of the parent structures[99–102].

Any identification in metabolomics has to take into account a level of uncertainty. Wrong identifications yield erroneous biological interpretations and this issue is particularly problematic since chemical isomers can represent entirely different biological molecules. Based on consensus reached by the Metabolomics standards initiative (MSI), different proposals to provide the level of certainty of the annotation exist, but they are infrequently implemented[103,104]. In general, the identification can be categorised in different levels with increasing confidence, ranging from complete unknowns to

tentative annotations (group-based annotation), matches against a database and confirmation with standards.

The scale of the MSI is oriented towards unambiguous identification, since it does include enantiomers characterisation (level 0), but it does not distinguish between a chemical formula and an accurate mass only (both level 4)[105]. The scale of Schymanski does provide an extra level for molecular formulas, which is a higher level of confidence in identification than m/z alone[103]. The scales proposed by the MSI[105] and Schymanski et al.[103] are merged into Figure 1 to reflect the entire range of identification confidence.

The level of confidence should not necessarily be implemented as a restriction (e.g. removal of level 4 and 5-metabolites) in the subsequent analyses, the scale rather serves as an additional tool to consider the confidence during further downstream interpretation (e.g. level 1 and 2 annotated metabolites are more reliable when different outcomes are possible during interpretation). Annotation of lipid species is a particular challenge since the permutation of fatty acyls on the backbone of different lipid classes generates a large collection of lipid species, and often no commercial standards and/or validated experimental MS/MS data are available for every lipid species[97,106]. As a result, lipid annotation confidence is theoretically lower than the average confidence in small molecule annotation.

The used identification strategies have been reported by 77 % of the reviewed literature, of which 13 % include a level of confidence[26,42,43]. Providing the strategies and confidence of the identified structures is a challenge yet to be implied in the standard workflow, as it opens the opportunity to improve the integration of the outcomes between different research groups.

3.6 Pathway analysis

Integration of the identified markers of toxicity into biologically relevant modes of action is a challenging aspect of the metabolomics workflow. Strategies for interpretation can be divided in two main categories, namely pathway-based methodologies or unbiased (network)based methodologies [87,107].

Pathway-based interpretations rely on existing databases, such as the Kyoto Encyclopedia for Genes and Genomes (KEGG) [108].The discovered markers of toxicity are projected on metabolic pathways, such as the glycolysis, the Krebs cycle and amino acid metabolism[13,22,23,28,33,41,43]. Similar to the transcriptomics[109] and proteomics[110] field, statistical enrichment analyses are used to test whether certain pathways are statistically enriched or depleted in the study. Popular pathway analysis software packages include Ingenuity Pathway analysis, MetaboAnalyst and IMPaLA[21,27,114,30,35,39,45,67,111–113].

Network-based approaches link covarying metabolites, creating a network of metabolite clusters[87]. An extra advantage of data-driven pathway analyses is the omission of identification restriction of non-identified metabolites; which improves the potential coverage of the metabolome through inclusion of for example enzymatic transformation products, which are not reported in standard libraries[107]. Network analyses are less frequently applied in the reviewed literature, but provide a valuable tool to infer new linkages between existing pathways[29,87,107]. The combination of data-driven network analysis supported by the biochemical knowledge represented in known pathways is a promising tool to construct new pathways and link existing pathways through new biochemical linkages, but it is yet to be applied in *in vitro* hepatic metabolomics research.

4. Metabolites as markers of hepatotoxicity

Most articles report the metabolic changes observed, describing the general function or pathway of the metabolite involved. Since the observed metabolic alterations do not always imply a toxic outcome (e.g. they are part of an adaptive response), the description of significant biomarkers can be fitted in potential pathways to argument potential toxicity. This is illustrated by the results of Søfteland, who noticed a higher number of metabolic perturbations with pesticides rather than with PAHs, although PAHs were toxic at lower concentrations[37]. This description is a good example that the number of metabolomic alterations should not be considered from a NOAEL-perspective, but that it should be interpreted as pathways involved in the potential toxicological mode of action[4]. García-Cañaveras et al.[26] and Ramirez et al.[2] classified different toxicants according to their mode of action, observing specific responses for each different end-point of toxicity. Their approach proves the rationale of different metabolomic fingerprints in perspective to the modes of toxicity involved.

Differentiating between the different modes of action of hepatotoxicity reported by all articles is not straightforward: many articles focus on subsets of the metabolome. This results in under-representation of other common markers of toxicity, e.g. focussing on the lipidome for steatosis might not reveal the alterations in the glycolytic intermediates. Additional concerns related to the integration of the results are potential differences in response to a toxicant due to cell line bias. An ideal example is the difference between the HepG2 and the HepaRG cell line: the first does not possess biotransformation capacity, obscuring the importance of bio-activation in the potential pathways of toxicity. The latter does have an advanced CYP450-dependent metabolism, and can therefore include additional mechanisms in its pathways of toxicity[115].

Combining the information of literature reveals common metabolic patterns in hepatotoxicity. Systematically altered metabolites can represent general toxicity, or they can be related to a specific mode of action of toxicity, discriminating between different adverse outcomes such as cholestasis and steatosis.

Of all metabolites reported as markers of toxicity, alanine was the most prevalent marker identified, being altered in over 40 % of the studies. The regulation is not consistent, upregulations being described as a catabolism response, while downregulations are explained as a wasting and authophagic process[15,23,25–27,30,34,36,38].

Lactate is also frequently observed as a toxicological marker (40 %) and is often correlated with the alanine levels [15,23,25,34,36]. A potential explanation can be the removal of pyruvate when the cell's metabolism is altered towards the glycolysis, the excess of pyruvate is further metabolised in lactate, with a minor shift towards alanine through transamination. Also proline disruptions (30 %) correlate with lactate levels[15,23,34,36,38].

3-Phospho glycerate is mostly downregulated (26 %), being a general marker of toxicity. Its central position in the glycolysis might explain why it is considered to be less specific. Citrate, a key metabolite in the TCA cycle, is also often reported (21 %), but the alterations are not consistent between the different studies.

Glutathione in reduced and oxidized state is often described as an altered metabolite in 40 % of the toxicological insults[15,25,26,35,36,41]. A decreased GSH/GSSG ratio is hallmark for oxidative stress, but it is also observed in other processes, making it not a specific biomarker for hepatotoxicity. Oxidative stress can be a downstream effect of other toxicological molecular initiating events, for example through uncoupling of the TCA cycle.

Other unspecific markers of hepatotoxicity are choline (21 %) downregulation and taurine disruptions (30 %). For the latter, no specific up-or downregulation could be observed.

The upregulation of glutamate is an important notice in many toxicological insults, being associated with disruptions in the TCA cycle[23,28,35,38]. Potential suggestions are protein catabolism to feed the amino acids in degradation pathways for energy productions. This might be supported by the observed upregulation of leucine, serine, and threonine (21 %). In this setting, tryptophan and valine are also usually up-regulated (26 %), especially observed during exposure to oxidative uncouplers[23,26,35,45].

Phosphatidylcholines (PC) and phosphatidylethanolamines (PE) are often correlated in steatosis and phospholipidosis[26,36,42,44]. Different behaviour of phospholipids by their chain length is observed by Kalkhof et al.[45], Olsvik et al.[36] and Cuykx et al.[43], related to recruitment of unsaturated FAs for

inflammatory intermediates[36,43,45]. PCs are inversely correlated with Lyso-PCs (LPC), dependent on the synthesis of PCs for the membrane function, or the catalysis of PCs to induce lipo-apoptosis. Yu et al. [28] have observed a correlation of acyl carnitine and LPC with inverse correlation of PC supposing either a recruitment of LPC from PCs or an inhibition of PC-formation. LPC is often upregulated which is reported to induce lipo-apoptosis [26,28,29,42,44]. However, downregulations can also relate to toxicity patterns when LPCs are combined with acyls to form PCs during oxidative stress[39].

Acylated carnitines are related with steatogenic processes, except by Shi and Meissen, which show a downregulation[26,28,29,35,41]. Sphingomyelines and ceramides are also upregulated in steatotic processes, their increases are often associated with oxidative stress[41,43,116–119].

Markers which were not often detected (< 20 %) but have similar disruptions among different articles are prostaglandins (up), glycine (up), AMP (down) UDP-hexose (down) PEP (down) PI (up) SAM (up) lysine (down) and creatine (up).

Figure 2 summarizes all observations into a comprehensive pathway. Many of the reported markers of toxicity are related to each other, being part of the glucose-metabolism or the lipid-metabolism, linked by the mitochondrial respiration. This observation can be explained by a) the direct impact of a toxicant on the energy homeostasis, and b) the reflection of end-stage toxicity, during which the final cell metabolism is the down-stream effect of the initial Molecular Initiating Event (MIE). While the former is a diagnostic identification of the pathways involved, the latter obscures the initial MIE, but it can provide indirect information of the pathways of toxicity, depending on the severity of disruption for every biochemical pathway.

5. Standardization of analyses/future perspectives

To successfully increase the throughput of metabolomics applications for *in vitro* assessment of hepatotoxicity, standardisation is a critical point in order to ensure optimal comparison between studies[4,61,120]. The standardisation of untargeted metabolomics protocols is a difficult task, since the chosen parameters are often based on the focus of a subset of the metabolome, which is dependent on the design of experiment and on the equipment used. The optimal cell type and culture conditions for the planned experiment can be incompatible with the conventional set-up and might require alternative approaches. Furthermore, different metabolomics platforms have their own advantages in the detection of several subsets of the metabolome, improving the separation and/or detection of the selected metabolites. Rather than standardising protocols, reporting the full details of the experimental conditions will have a greater impact and value on the scientific field as it improves future comparisons in literature.

Important is a standardisation of transparency and QC implementations, a condition often mentioned in metabolomics recommendations, but not yet widely reported[61,71]. This is especially imperative for the experimental design and data-analytical processing steps, such as blank subtractions, missing value imputations and scaling procedures, which have an important impact on the final outcome of the experiment[61,71,121]. Full description of the applied workflow, mentioning every detail with its impact on the dataset (even of processing steps not applied), can and should be included in supplementary information, assisting the reader to accurately define strengths and benefits of each *in vitro* metabolomics experiment.

As a final recommendation, the implementation of QC is a vital part of good metabolomics practices. Although QC is currently not fully implemented, which has a negative impact for further meta-analyses, this issue can be readily solved by easy implementations in further experiments. The addition of QC samples, such as the QC-pool and blank samples, offers a way to assess the variance of the dataset and reduce the complexity and the noise of the data to ultimately improve the final quality of the experiment. Such procedures should therefore be implemented and evaluated in every metabolomics application[4,62,120]. As a guidance, Table 3 provides a checklist considering key-events in the metabolomics workflow, warning for important pit-falls and providing examples of good metabolomics practices.

Metabolomics is only beginning to be a platform to investigate *in vitro* hepatotoxicity. The growing number of publications shows the increasing importance of phenotype fingerprinting for mechanistic interpretations of hepatotoxicity. By keeping up with the recommendations of metabolomics consortia, combining good experimental design with transparency, the overall quality of *in vitro* assessments for hepatotoxicity improves toward a maturation of *in vitro* applications and will be a valuable additional tool providing important insights for the research field.

Acknowledgments

Matthias Cuykx and Dr. Robim M Rodrigues have been funded by individual fellowships of Research Foundation Flanders (FWO, fellowships 12H2216N and 11Z3318N). Further funding has been available from the Vrije Universiteit Brussel (VUB, Brussels, Belgium) and the University of Antwerp (UA, Antwerp, Belgium).

Conflict of Interests

The authors declare to have no conflicts of interests.

References

1. Stierum R, Heijne W, Kienhuis A, Van Ommen B, and Groten J. Toxicol Appl Pharmacol. 2005;179–88.
2. Ramirez T, Strigun A, Verlohner A, Huener HA, Peter E, Herold M, et al. Arch Toxicol. 2017;839–906.
3. Vinken M. Toxicology . 2013;158–65.
4. Ramirez T, Daneshian M, Kamp H, Bois FY, Clench MR, Coen M, et al. ALTEX. 2013;209–25.
5. Rang H Ritter J, Flower R, Henderson G. et al. DM. . 7th ed. 2012.
6. Willebrords J, Pereira IVA, Maes M, Crespo Yanguas S, Colle I, Van Den Bossche B, et al. Prog Lipid Res . 2015;106–25.
7. Younossi ZM, Koenig AB, Abdelatif D, Fazel Y, Henry L, and Wymer M. Hepatology. 2016;73–84.
8. Dash A, Figler RA, Sanyal AJ, and Wamhoff BR. Expert Opin Drug Metab Toxicol . 2017;193–204.
9. Holt MP, and Ju C. AAPS J . 2006;E48-54.
10. Chen M, Suzuki A, Borlak J, Andrade RJ, and Lucena MI. J Hepatol . 2015;503–14.
11. Usui T, Mise M, Hashizume T, Yabuki M, and Komuro S. Pharmacology. 2009;2383–92.
12. Van Summeren A, Renes J, Lizarraga D, Bouwman FG, Noben JP, van Delft JH, et al. OMICS . 2013;71–83.
13. Balcke GU, Kolle SN, Kamp H, Bethan B, Looser R, Wagner S, et al. Toxicol Lett . 2011;200–9.
14. Nicholson JK, Lindon JC, and Holmes E. Xenobiotica. 1999;1181–9.
15. Ruiz-Aracama A, Peijnenburg A, Kleinjans J, Jennen D, van Delft J, Hellfrisch C, et al. BMC Genomics . 2011;251.
16. European Medicines Agency. 2016;JEG-3Rs/450091/2012.
17. ECHA. 2011;ECHA-11-R-004.2.
18. Williams ES, Panko J, and Paustenbach DJ. Crit Rev Toxicol . 2009;553–75.
19. Rovida C HT. ALTEX. 2009;187–208.
20. Godoy P, Hewitt NJ, Albrecht U, Andersen ME, Ansari N, Bhattacharya S, et al. Arch Toxicol . 2013;1315–530.
21. Kim S-R, Kubo T, Kuroda Y, Hojyo M, Matsuo T, Miyajima A, et al. J Toxicol Sci . 2014;717–23.
22. Sugiyama K, Ebinuma H, Nakamoto N, Sakasegawa N, Murakami Y, Chu PS, et al. PLoS One. 2014;e94460:1-18.
23. Snouber LC, Bunescu A, Naudot M, Legallais C, Brochot C, Dumas ME, et al. Toxicol Sci. 2013;8–20.
24. Zhang J, Williams TD, Abdallah MA-EE, Harrad S, Chipman JK, and Viant MR. Toxicol Vitr . 2015;2116–23.
25. Li Z, Zheng L, Shi J, Zhang G, Lu L, Zhu L, et al. Evidence-Based Complement Altern Med . 2015;1–11.
26. García- Cañaveras JC, Castell J V., Donato MT, and Lahoz A. Sci Rep . 2016;27239.
27. Chatterjee N, Yang J, Yoon D, Kim S, Joo SW, and Choi J. Biomaterials . 2017;167–80.
28. Yu J, Zhang H, Li Y, Sun S, Gao J, Zhong Y, et al. Biomed Chromatogr. 2017;1–10.
29. Meissen JK, Hirahatake KM, Adams SH, and Fiehn O. Metabolomics. 2015;707–21.
30. Van den Hof WFPM, Ruiz-Aracama A, Van Summeren A, Jennen DGJ, Gaj S, Coonen MLJ, et al. Toxicol Vitr . 2015;489–501.
31. Vorrink SU, Ullah S, Schmidt S, Nandania J, Velagapudi V, Beck O, et al. FASEB J. 2017;2696–708.
32. Carretero A, Leon Z, Garcia-Canaveras JC, Zaragoza A, Gomez-Lechon MJ, Donato MT, et al. Anal Bioanal Chem. 2014;5465–76.
33. Wang Y, Zhu W, Qiu J, Wang X, Zhang P, Yan J, et al. J Hazard Mater . 2015;9–16.
34. Toyoda Y, Kashikura K, Soga T, and Tagawa Y. J Toxicol Sci. 2017;445–54.
35. Olsvik PA, Skjærven KH, and Søfteland L. Chemosphere . 2017;730–43.

36. Olsvik PA, Berntssen MHG, and Søfteland L. Toxicol Vitr . 2017;1–14.
37. Søfteland L, Kirwan JA, Hori TSF, Størseth TR, Sommer U, Berntssen MHG, et al. Food Chem Toxicol. 2014;157–74.
38. Liu S, Wang W, Zhou X, Gu R, and Ding Z. Environ Toxicol Pharmacol . 2014;150–7.
39. Shi J, Zhou J, Ma H, Guo H, Ni Z, Duan J, et al. Anal Bioanal Chem. 2016;1413–24.
40. Li F, Wang L, Ji C, Wu H, Zhao J, and Tang J. Chemosphere . 2017;88–96.
41. Brown M V, Compton SA, Milburn M V, Lawton KA, and Cheatham B. Obesity. 2013;E561–70.
42. Van den Eede N, Cuykx M, Rodrigues R, Laukens K, Neels H, Covaci A, et al. Toxicol Vitr . 2015;2045–54.
43. Cuykx M, Claes L, Rodrigues RM, Vanhaecke T, and Covaci A. Toxicol Lett. 2018;
44. Seeßle J, Liebisch G, Schmitz G, Stremmel W, and Chamulitrat W. Biochim Biophys Acta - Mol Cell Biol Lipids . 2015;549–65.
45. Kalkhof S, Dautel F, Loguercio S, Baumann S, Trump S, Jungnickel H, et al. J Proteome Res . 2015;164–82.
46. Aninat C, Piton A, Glaise D, Le Charpentier T, Langouet S, Morel F, et al. Drug Metab Dispos . 2006;75–83.
47. Antherieu S, Chesne C, Li R, Guguen-Guillouzo C, and Guillouzo A. Toxicol Vitr . 2012;1278–85.
48. Lubberstedt M, Muller-Vieira U, Mayer M, Biemel KM, Knospel F, Knobeloch D, et al. J Pharmacol Toxicol Methods . 2011;59–68.
49. Guillouzo A, Corlu A, Aninat C, Glaise D, Morel F, and Guguen-Guillouzo C. Chem Biol Interact. 2007;66–73.
50. Rodrigues RM, Kollipara L, Chaudhari U, Sachinidis A, Zahedi RP, Sickmann A, et al. Arch Toxicol . 2018;1–14.
51. Martano G, Delmotte N, Kiefer P, Christen P, Kentner D, Bumann D, et al. Nat Protoc . 2014;1–11.
52. Cuykx M, Mortelé O, Rodrigues RMRM, Vanhaecke T, and Covaci A. Anal Methods. 2017;3704–12.
53. Bordag N, Rennefahrt U, Nachtigall J, Gonzalez Maldonado S, Rezska R, Ramirez-Hernandez T, et al. Metabolomics . 2016;1–4.
54. García-Cañaveras JC, Jiménez N, Gómez-lechón MJ, and Castell J V. Electrophoresis. 2015;1–24.
55. García-Cañaveras JC, Peris-Díaz MD, Alcoriza-Balaguer MI, Cerdán-Calero M, Donato MT, and Lahoz A. Electrophoresis. 2017;2331–40.
56. Sladowski D, Steer SJ, Clothier RH, and Balls M. J Immunol Methods. 1993;203–7.
57. Ishiyama M, Miyazono Y, Sasamoto K, Ohkura Y, and Ueno K. Talanta. 1997;1299–305.
58. Zhang S-Z, Lipsky MM, Trump BF, and Hsu I-C. Cell Biol Toxicol. 1990;219–34.
59. Zhang X, Li F, Nordin AN, Tarbell J, and Voiculescu I. Sens Bio-Sensing Res . 2015;112–21.
60. Madji Hounoum B, Blasco H, Emond P, and Mavel S. TrAC - Trends Anal Chem . 2016;118–28.
61. Sumner LW, Amberg A, Barrett D, Beale MH, Beger R, Daykin CA, et al. Metabolomics . 2007;211–21.
62. Dunn WB, Wilson ID, Nicholls AW, and Broadhurst D. Bioanalysis . 2012;2249–64.
63. Broadhurst DI, and Kell DB. Metabolomics. 2006;171–96.
64. Jonsson P, Wuolikainen A, Thysell E, Chorell E, Stattin P, Wikström P, et al. Metabolomics . 2015;1667–78.
65. Lorenz M a., Burant CF, and Kennedy RT. Anal Chem . 2011;3406–14.
66. Cuykx M, Negreira N, Beirnaert C, Van den Eede N, Rodrigues R, Vanhaecke T, et al. J Chromatogr A . 2017;168–78.
67. García-Cañaveras JC, López S, Castell JV, Donato MT, and Lahoz A. Anal Bioanal Chem . 2016;1217–30.
68. Matyash V, Liebisch G, Kurzchalia T V, Shevchenko A, and Schwudke D. J Lipid Res . 2008;1137–46.
69. Bi H, Krausz KW, Manna SK, Li F, Johnson CH, and Gonzalez FJ. Anal Bioanal Chem. 2013;5279–

89.
70. Leon Z, Garcia-Canaveras JC, Donato MT, and Lahoz A. Electrophoresis . 2013;2762–75.
71. Hayton S, Maker GL, Mullaney I, and Trengove RD. Cell Mol Life Sci. 2017;4421–41.
72. Dietmair S, Timmins NE, Gray PP, Nielsen LK, and Krömer JO. Anal Biochem . 2010;155–64.
73. Vuckovic D. Anal Bioanal Chem. 2012;1523–48.
74. Bligh E, and Dyer W. Can J Physiol Pharmacol. 1959;911–7.
75. Wu H, Southam AD, Hines A, and Viant MR. Anal Biochem . 2008;204–12.
76. García-Cañaveras JC, Donato MT, Castell J V., and Lahoz A. J Proteome Res. 2011;4825–34.
77. Dunn WB, Broadhurst D, Begley P, Zelena E, Francis-McIntyre S, Anderson N, et al. Nat Protoc. 2011;1060–83.
78. Fei F, Bowdish DME, and McCarry BE. Anal Bioanal Chem. 2014;3723–33.
79. Zhang T, Creek DJ, Barrett MP, Blackburn G, and Watson DG. Anal Chem . 2012;1994–2001.
80. Jandera P. Anal Chim Acta. 2011;1–25.
81. Knee JM, Rzezniczak TZ, Barsch A, Guo KZ, and Merritt TJS. J Chromatogr B . 2013;63–73.
82. McCloskey D, Gangoiti J a., Palsson BO, and Feist AM. Metabolomics . 2015;1338–50.
83. Acunha T, Simó C, Ibáñez C, Gallardo A, and Cifuentes A. J Chromatogr A . 2015;1–10.
84. Sandra K, and Sandra P. Curr Opin Chem Biol . 2013;847–53.
85. Amathieu R. World J Gastroenterol . 2016;417–26.
86. Yin P, and Xu G. J Chromatogr A . 2014;1–13.
87. Alonso A. Front Bioeng Biotechnol . 2015;1–20.
88. Gromski PS, Xu Y, Kotze HL, Correa E, Ellis DI, Armitage EG, et al. Metabolites . 2014;433–52.
89. Godzien J, Ciborowski M, Angulo S, and Barbas C. Electrophoresis. 2013;2812–26.
90. Ejigu BA, Valkenborg D, Baggerman G, Vanaerschot M, Witters E, Dujardin J-C, et al. OMICS . 2013;473–85.
91. Wu Y, and Li L. J Chromatogr A . 2015;80–95.
92. Silva LP, Lorenzi PL, Purwaha P, Yong V, Hawke DH, and Weinstein JN. Anal Chem. 2013;9536–9542.
93. Dieterle F, Ross A, Schlotterbeck G, and Senn H. Anal Chem. 2006;4281–90.
94. Godzien J, Alonso-Herranz V, Barbas C, and Armitage EG. Metabolomics . 2014;518–28.
95. Kim KI, and van de Wiel MA. BMC Bioinformatics . 2008;114.
96. Wishart DS, Jewison T, Guo AC, Wilson M, Knox C, Liu YF, et al. Nucleic Acids Res. 2013;D801–7.
97. Fahy E, Sud M, Cotter D, and Subramaniam S. Nucleic Acids Res . 2007;W606-12.
98. Smith CA, O'Maille G, Want EJ, Qin C, Trauger SA, Brandon TR, et al. Ther Drug Monit. 2005;747–51.
99. Allen F, Pon A, Wilson M, Greiner R, and Wishart D. Nucleic Acids Res. 2014;W94–9.
100. Wolf S, Schmidt S, Müller-Hannemann M, and Neumann S. BMC Bioinformatics. 2010;148.
101. Kind T, Yang DW, Leon C, Fiehn O, Ma Y, Kind T, et al. Anal Chem. 2014;10724–31.
102. De Vijlder T, Valkenborg D, Lemière F, Romijn EP, Laukens K, and Cuyckens F. Mass Spectrom Rev . 2017;1–23.
103. Schymanski EL, Jeon J, Gulde R, Fenner K, Ruff M, Singer HP, et al. Environ Sci Technol. 2014;2097–8.
104. Dunn WB, Erban A, Weber RJM, Creek DJ, Brown M, Breitling R, et al. Metabolomics. 2013. p. 44–66.
105. Dunn WB, Sumner LW, Griffen J, Creek D, Trengove R, and Boughton B. Metabolomics conference 2017: workshops. 2017.
106. Sud M, Fahy E, Cotter D, Brown A, Dennis E, Glass C, et al. Nucleic Acids Res. 2007;D527-532.
107. Rosato A, Tenori L, Cascante M, De Atauri Carulla PR, Martins dos Santos VAP, and Saccenti E. Metabolomics . 2018;1–20.
108. Okuda S, Yamada T, Hamajima M, Itoh M, Katayama T, Bork P, et al. Nucleic Acids Res . 2008;W423-6.
109. Khatri P, Sirota M, and Butte AJ. PLoS Comput Biol. 2012;

110. Laukens K, Naulaerts S, and Berghe W Vanden. Proteomics. 2015;981–96.
111. Xia J, Psychogios N, Young N, and Wishart DS. Nucleic Acids Res. 2009;652–60.
112. Chong J, Soufan O, Li C, Caraus I, Li S, Bourque G, et al. Nucleic Acids Res. 2018;1–9.
113. Kamburov A, Cavill R, Ebbels TMD, Herwig R, and Keun HC. Bioinformatics. 2011;2917–8.
114. Van den Eede N, Cuykx M, Rodrigues RM, Laukens K, Neels H, Covaci A, et al. Toxicol Vitr . 2015;2045–54.
115. Rodrigues RM, Heymans A, De Boe V, Sachinidis A, Chaudhari U, Govaere O, et al. Toxicol Lett . 2016;50–9.
116. Gorden DL, Myers DS, Ivanova PT, Fahy E, Maurya MR, Gupta S, et al. J Lipid Res . 2015;722–36.
117. Turner N, Kowalski GM, Leslie SJ, Risis S, Yang C, Lee-Young RS, et al. Diabetologia. 2013;1638–48.
118. Alonso C, Fernández-Ramos D, Varela-Rey M, Martínez-Arranz I, Navasa N, Van Liempd SM, et al. Gastroenterology. 2017;1449–1461.e7.
119. Amacher DE. Toxicology . 2011;10–8.
120. Holmes C, McDonald F, Jones M, Ozdemir V, and Graham JE. OMICS. 2010;327–32.
121. Hrydziuszko O, and Viant MR. Metabolomics. 2012;161–74.

738 **Figure 1:** Levels of annotation confidence as proposed by Dunn et al.[105] with the inclusion of an additional level distinguishing between chemical formula an
739 exact mass, as proposed by Schymanski et al.[103]. Adapted from Dunn et al. and Schymanski et al.[103,105].

740

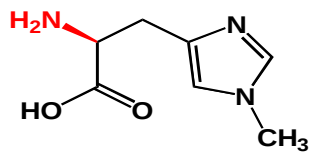
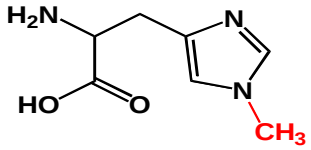
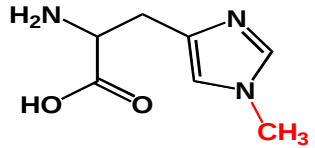
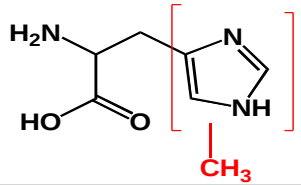
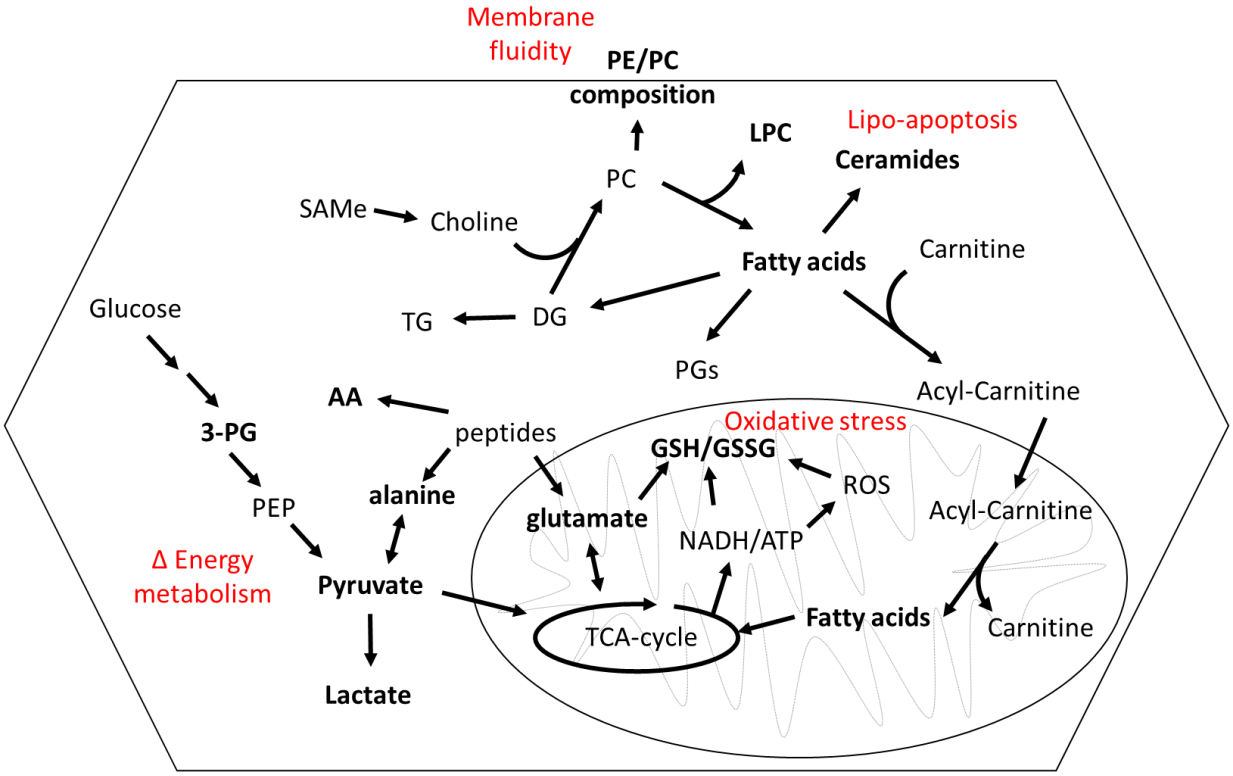
	Level of confidence	Requirements
	**Level 0:** Unambiguous structure	Full structure and stereochemistry of isolated, pure compound
	**Level 1:** Confident structure	Reference standard match using at least two orthogonal techniques MS, MS/MS and RT matching authentic reference standards
	**Level 2:** Probable structure	Match to literature/database of at least two techniques e.g. MS, MS/MS, RT matching libraries
	**Level 3:** Probable structure/class Several candidates possible	Match of at least two techniques to support candidate structure/class e.g. MS, MS/MS, RT (indicative)
C7H11N3O2	**Level 4:** Unequivocal chemical formula, but no class identification	Match of Mass spectrum (mass/isotopes) with proposed formula
169.0851	**Level 5:** Unknown feature	No identification, exact mass only

Figure 2: Proposed connection between different markers of toxicity (black) and the most reported pathways of toxicity (red). 3-PG: 3-phospho glycerate, AA: different amino acids, DG: diacylglycerols, GSH/GSSG: ratio of reduced and oxidised glutathione, LPC: lyso-phosphatidylcholines, PC: phosphatidylcholines, PE: phosphatidylethanolamines, PEP: phosphor-enol pyruvic acid, PG: prostaglandins, ROS: reactive oxygen species, SAMe: S-adenosylmethionine, TCA: tricarboxylic acid, TG: triacylglycerols.



751 **Table 1:** Culture conditions applied during *in vitro* experiments.

752

Article	Year	Cell type	Culture type	Tested compounds	Viability assay	Exposure	Concentration	Concentrations	replicates	ref
Balcke	2011	V79 fibroblast	6 well plate	oxidative uncouplers: KCN, sodium-azide, dinitrofenol, trifluoromethoxyphenylhydrazone	MTT	1 h - 24 h	IC20 & IC50	10 µM (FCPP) to 500 µM (DNP) 1 mM (KCN/NaN_3)	5	[13]
Ruiz-Aracama	2011	HepG2 cell line	T75 flask	TCDD	MTT	48 h	IC20	10 nM	5x5	[15]
Snouber	2012	HepG2 cell line	Biochip: 2 cm^2	Flutamide	N/A	48h	Plasma C	10 µM	3	[23]
Brown	2013	HepaRG cell line	12 well plate	Palmitate and oleate	N/A	48 h	N/R	0.5 mM	3	[41]
Carretero	2014	Primary rat hepatocytes	6 well plate	Cumene hydroperoxide	N/R	24 h	N/R	500 µM	6	[32]
Kalkhof	2014	Murine Hepatoma Hepa1c1c7	Petri dish 15 cm	BenzoApyrene	NRU	7 pt - 24 h	Subacute - IC20	50 nM – 5 µM	3	[45]
Kim	2014	Human adult vs fetus hepatocytes	Equivalent of 2 million cells	Comparison control conditions	N/A	N/R	N/A1	N/A	8	[21]
Liu	2014	L02 cell line	10 cm petri dish	Cisplatin	CCK8	24 h	Hormetic C & inhibitory C	1 nM- 1 µM-1 mM	8	[38]
Søfteland	2014	Primary salmon hepatocytes	6 well plate	PAH/Pesticides	impedance	48 h	IC30	1 µM – 100 µM	5	[37]
Sugiyama	2014	HPI vs huh-7 cell line	Equivalent of 1 million cells	Comparison cell types	N/A	N/R	N/A	N/A	N/R	[22]
Li	2015	HepG2 cell line	T75 flask	Matrine	N/R	3pt 48 h	N/A	0.2 µM- 0.4 µM- 0.7 µM	5-4-3	[25]
Meissen	2015	HepG2 cell line	12 well plate	Glucose vs fructose	N/A	5 pt 48 h	N/A	5 mM	6	[29]
SeeßLe	2015	Huh-7 cell line	12 well plate	Palmitate	N/A	8 h	N/A	300 µM	6	[44]
Van den Eede	2015	HepaRG cell line	12 well plate	Triphenylphosphate Acetaminophen	MTT	72 h	IC10	50 µM – 2 mM	6	[42]
Van den Hof	2015	HepG2 cell line	T75 flask	Cyclosporine A	MTT	24 h -72 h	IC20	3 µM - 20 µM	5	[30]
Wang	2015	Primary rat hepatocytes	12 well plate	Hexaconazole	MTT	24 h	N/R	2 µM	3	[33]
Zhang	2015	HepG2 cell line	6 well plate	HBCD and DMBA	MTT	24 h	1/10 of IC50	4 µM – 10 µM	6	[24]
Garcìa-cañaveras	2016	HepG2 cell line	6 well plate	Different drugs: toxic and non-toxic	MTT	24 h	1/10 IC10 – IC10	µM to Mm range	3	[26]
Shi	2016	L02 cell line	6 well plate	Pekinenal	MTT	48 h	Non-toxic C vs IC20 & IC30	5 µM-10 µM-20 µM	6	[39]
Chatterjee	2017	HepG2 cell line	6 well plate	Nanotubes	ez-cytox	24 h	IC20	2 mg/L – 10 mg/l	N/R	[27]
Olsvik	2017	Primary fish hepatocytes	6 well plate	Pirimiphos methyl	MTT	48 h	Not cytotoxic	0.1 µM - 1 mM	6	[35]
Olsvik	2017	Primary salmon hepatocytes	6 well plate	BPA and genistein	MTT	48 h	Not cytotoxic	100 µM	5	[36]
Ramirez	2017	HepG2 cell line	Petri dish 35 mm	Different hepatotoxicants	WST-1	48 h	I1/3 IC20-IC20	10 µm – 1 mM	8	[2]
Toyoda	2017	Primary mouse hepatocytes	6 well plate	Acetaminophen	WST-1	12 h	IC90	10 mM	4	[34]
vorrink	2017	primary human hepatocytes	12 well vs 3D culture	Comparison control conditions	N/A	N/A	N/A	N/A	9	[31]
Yu	2017	HepG2 cell line	Petri dish 10 cm	Cationic liposomes	CCK8	24 h	IC50	120 µg/mL	6	[28]
Li	2017	L02 cell line	Petri dish 10 cm	Trichloropropyl phosphate	CCK8	24 h	Hormesis C & toxic C	1 µM – 100 µM	8	[40]
Cuykx	2018	HepaRG cell line	Chamber slides 4 cm^2	Valproic acid	NRU	24 h-72 h	IC10 & 1/10 IC10	66.5 µM-665 µM	6	[43]

753

Table 2: Pre-analytical design and instrumental platforms applied for *in vitro* hepatic metabolomics. CE: capillary electrophoresis, DI-FTICR: direct infusion Fourier-transformed ion cyclotron, IP: ion pairing, IT: ion trap, MS: mass spectrometry, NMR: nuclear magnetic resonance, QTRAP: quadrupole ion trap, QQQ: triple quadrupole, RP-LC: reversed phase chromatography, TOF: Time of Flight, UPLC: ultra-performant liquid chromatography.

Quenching	Extraction	Platform	Data-preparation	Normalisation	QC	Data-analysis	ref
Cold ethanol/DCM	DCM-ETOH-water	IP-UPLC-QQQ-MS	Log transformation paretoscale	Protein determination	N/R	PLS-DA	13
Cold isotone water/ultrasonication	Chloroform-water	400 MHz NMR; GC-MS and RP-UPLC-TOF-MS	Log transformation	Phospholipid signals NMR	N/R	ANOVA PCA	15
Not applicable	Protein precipitation	800 MHz NMR	Manual fitting	Cell number	N/R	MWU PLS-DA	23
Methanol-Room temperature	Protein precipitation	GC and RP-LC-IT-MSMS	Log transformation	Protein determination	N/R	Welch ttest /ANOVA + FDR PCA	41
Liquid nitrogen	Protein precipitation	RP-LC TOF	Quantified	N/R	Stock solutions	MWU	32
Methanol/water	Protein precipitation	RP-LC-QTRAP	Quantified	Internal standards	N/R	Wilcoxon + FDR	45
Hypertonic solution	Chloroform-methanol-water	CE-TOFMS	Internal standard ratio	Cell number	N/R	Welch t-test PCA	21
Cold PBS	Chloroform-methanol-water	500 MHz NMR	N/R	Mean centered, total area	N/R	ANOVA PCA & PLSDA	38
Lyophilisation	Chloroform-methanol-water	FTICR and 600 MHz NMR	Filter, MV imputation log transformation	Mean centered, PQN	QC-pool	ANOVA + FDR PCA & PLSDA	37
Methanol	Protein precipitation	two platform CE/LC-TOF-MS	Fold change	N/R	N/R	PCA & HCA	22
Trypsinisation	Chloroform-methanol-water	400 MHz NMR	PCA-outlier removal Pareto scale	N/R	N/R	ANOVA +FDR OPLS	25
Cold PBS/methanol	Protein precipitation	HILIC-LC and GC TOF	Quantified	Average sum of intensities	N/R	differential (univariate)statistics PLS	29
Freeze-thaw cycles	Methylation for GC	GC-Q-MS	Not applied	N/R	N/R	ANOVA+ FDR	44
Liquid nitrogen	Chloroform-methanol-water	RP-UPLC-QTOF	Filter Log transformation	Mean of QC	QC pools and blanks	T-test + FDR	42
Cold isotone water	Chloroform-water	NMR + RP- LC-QQQ-MS	Log transformation	PQN	N/R	T-test	30
Cooled	Protein precipitation	RP-LC-QQQ-MS	Not applied	N/R	N/R	T-test	33
Methanol/water	DCM-methanol-water	DI-FTICR	Filters MV imputation Log transformation	PQN	N/R	ANOVA + FDR PCA	24
Liquid nitrogen	Chloroform-methanol-water	3 LC-QTOF-MS platforms	Log transformation unit variance scaled mean centered	Protein determination	QC pools and blanks	ANOVA +FDR PCA - PLSDA	26
Freeze-thaw cycles	Protein precipitation	RP-UPLC-QTOF-MS	Markerlynx	Base peak intensity	Blanks	T-test PLS-DA	39
Not reported	Protein precipitation	600 MHz NMR	Not applied	Total intensity area	N/R	ANOVA PLS	27
Liquid nitrogen	METABOLON, not mentioned	GC-MS, RP-LC-LIT/FTICR	Log transformation	Protein determination	N/R	ANOVA + FDR	35
Liquid nitrogen	METABOLON, not mentioned	GC-MS, RP-LC-LIT/FTICR	MV imputation normality check outlier removal log transformation	Protein determination	N/R	ANOVA + FDR	36
Liquid nitrogen	DCM-Ethanol-water	GC-MS, RP-LC-QTRAP	Log transformation mean centering unit variance scaled	Pool-baselined, median normalized	QC pool and blanks	T-test correlation PCA	2
Mannitol washing and methanol	Chloroform-methanol-water	CE-TOFMS	Not applied	Mouse albumin	N/R	T-test	34
Organic solvent and liquid nitrogen	Protein precipitation	RP-LC-QQQ and orbitrap-MS	Quantified	N/R	Blanks	Correlation analysis	31

Methanol	Chloroform-methanol	RP-LC-QTOF-MS	Filters	Autoscaling	QC-pool	T-test	[28]
Not reported	Chloroform-methanol-water	500 MHz NMR	Log transformation	Mean centering	N/R	PLS-DA ANOVA	[40]
Liquid nitrogen	Chloroform-methanol-water	4 LC- QTOF-MS platforms	Filter Log transformation outlier removal	PQN, median centering	QC-pool and blanks	PCA & PLS-DA	[43]

757

Table 3: Critical steps in a metabolomics workflow, with inclusion of the most important risks and pitfalls, potential solutions and important reporting conditions.

Workflow		Risks	Pitfalls	Solutions and recommendations	Examples	Report
Study design	Cell selection	Cell type bias	Low biotransformation capacity	Choose appropriately	HepG2 (low capacity) vs primary hepatocytes (high capacity) Fetal vs adult primary hepatocytes; rat vs human primary hepatocytes	Cell culture: Cell-carrier, Cell density, Cell passage number, Cell type, Coating, Replicates
			Inter-species differences	Choose appropriately		
	Culture conditions	Culture bias	Cell stability	Choose appropriate culture conditions	2D vs 3D culture, Validated time-frame, Cell density	Cultivation parameters: Cultivation time between seeding and exposure, Incubator parameters, Medium composition
			Biotransformation capacity		Medium composition (serum, DMSO, antibiotics, …)	
	Replicates	Not enough statistical power	Low number of replicates	Change design	Paired samples, Time-point analysis	Exposure conditions: Dosages, Exposure method, Time-frame, Method for dosage estimation, Randomization of culture
			High biological variance	Increase replicates	Replication experiment	
				Reduce variance	Cell lines vs primary hepatocytes, reproducible seeding and exposure protocols	
	Exposure	Dosage bias	No observed effects	Decide dosages based on dose-response	Use cell-viability assay to assess potential toxic dosages	
			Biased cell viability assay	Choose an independent cell viability method	Neutral red uptake assay when mitochondrial toxicity is suspected.	
			General cell toxicity	Choose sub-cytotoxic concentrations	Imply extra (lower) levels of concentrations, correct cell viability assay	
		Temporal bias	Time-frame too short	Longer time-frames	24 h versus 72 h exposure	
			Time-frame too long	Shorter sampling times	Inclusion of extra time points	
Sample preparation	Cell material	Low concentrations of metabolites	Low amount of cell material	Cultivate on large surfaces Perform targeted analysis	6-well plate or flasks	Washing: Mention if applied or not, Washing solution composition, Washing solution Temperature
			General cell death	Apply lower dosages	Imply extra (lower) levels of concentrations, correct cell viability assay	Cell harvesting: Harvesting method
	Quenching	High variance	Extraction extracellular matrix	Apply washing step	Apply washing step	
		Low concentrations	Metabolite leakage	Validate extraction	Quench remaining metabolism as soon as possible	Quenching: Quenching method, Quenching temperature, Quenching time
		Response bias	Trypsinisation	Cell scraping		
	extraction	Increased variability	Protein precipitation	Choose method in function of application	Bligh and dyer extraction for lipidomics	Extraction: Extraction method, Extraction time, Randomisation, Solvent composition, Temperatures
			Oxidation	Validate protocol	Addition of stabilisers	
		Matrix effects	Extraction variability	Internal standards (targeted methods)		QC: Implementation, Pooling method, Source
						Storage conditions: Temperature Time
						Additional: biological normalisation techniques
Analytical platforms	Platform decision	Sensitivity bias	Low sensitivity of lipids for NMR	Choose in function of subset	HILIC-MS for polar metabolites GC-MS for fatty acid compositions	Platform: Equilibration, Instrumentation, Parameters, Solvent compositions
	Platform use	Time-bias	Changing sensitivity due to contamination	Randomise samples Implement instrument equilibration Internal standards (targeted methods)	Block randomisation Repeated QC-pool injections before sample analysis	Reconstitution: Reconstitution method, Solvent composition, Temperature, Time between reconstitution and injection
Data processing	Feature selection	Algorithm bias	Isotopes Peak splitting Wrong alignments	Isotope and adduct grouping Correct parameters of processing workflow	Estimate system performance (mass accuracy, peak width)	Software: Version Algorithms: Rationale, Parameters
	Feature filtering	Selection bias	Sparce dataset Retaining noisy features False positives No relevant markers	Filter steps based on meta-data	Missing value filtering Blanc subtraction Variance filtering	Filter steps: Filter thresholds, Rationale Normalisation: Applied algorithm, Impact, Parameters, Rationale
		Selectivity bias	False negatives		Base thresholds on results of a repeated injection (e.g. QC)	
	Normalisation	Normalisation bias	Loss of variance	Estimate impact normalisation	Report variance reductions, select most significant variables	

			Normalisation is interfered by MOA	Data-dependent normalisation	PQN-normalisation DNA-normalisation	
		Scaling bias	Bias to most abundant signals Bias to low-abundant bias	Scale	Log transformation Pareto-scale	
QC	Variance estimation	No implementation of QC	No estimation of biases and variance	Implement QC procedures	QC samples Internal standards	QC-samples: Equilibration, Frequency, Type
	Between batches	No implementation of QC	RT-shifts	Implement a mix of standards	External standards Internal standards	Internal standards: Implementation
			Mass accuracy shifts			Corrections: Outlier Removal, Rationale, Report
			Sensitivity shifts			
Statistical interpretations	Univariate	No FDR-correction	False positive results	Implement correction	FDR-correction Bonferroni/Benjamin-Hochberg	Univariate tests: False discovery correction, Output, Rationale, Thresholds, Type
		Low power		Increase replicates Other tests	Replication study Non-parametric tests or Multivariate test	
	Multivariate	Wrong interpretation Classification only	Loss of information	Choose correct test	PCA: exploratory, PLS-DA: regression/classification, other models: classification	Multivariate tests: Algorithm, Cross-validation, Selection technique, Output, Rationale, Thresholds
				Select important variables	Latent structures or Variables of importance	
		No cross-validation	Overfitting and false positive results	Cross-validate	Leave-one-out, Data permutations, Replication study	
Identification	Databases	Identification bias	Limited number of metabolites	Cross-check multiple databases	Metlin LipidMaps, Human metabolome database	Databases: Type, Version
	Untargeted	Identification bias	Isomers Single response of algorithms	Confirmation of algorithm outcome	Retention time MS/MS Isotopes	Identification: Algorithms, Confidence of identification, Identifiers, Methods, Parameters, Techniques, Thresholds
		Wrong identification	Multiple combinations for single m/z	Orthogonal platforms	NMR Report certainty	
Interpretation	Interpretation	Interpretation bias	False positive outcomes Classification only	Validate biological relevance		Interpretation: Pathway Analysis software, Structured pathway
		Structure	Incomprehensible pathways of Toxicity	Validate metabolite relationships	Literature Presence of biomarkers in common pathways	
		Relevance	Significant features are not necessarily markers of toxicity	Validate biological relevance	Literature	

760