t⁴ Workshop Report[*]:

Quality Assurance of Metabolomics

Mounir Bouhifd[1], **Richard Beger**[2], **Thomas Flynn**[3], **Lining Guo**[4], **Georgina Harris**[1], **Helena Hogberg**[1], **Rima Kaddurah-Daouk**[5], **Hennicke Kamp**[6], **Andre Kleensang**[1], **Alexandra Maertens**[1], **Shelly Odwin-DaCosta**[1], **David Pamies**[1], **Donald Robertson**[7], **Lena Smirnova**[1], **Jinchun Sun**[2], **Liang Zhao**[1], and **Thomas Hartung**[1,8]

[1]Johns Hopkins University, Bloomberg School of Public Health, Center for Alternatives to Animal Testing, Baltimore, MD, USA

[2]US Food and Drug Administration, National Center for Toxicological Research, Division of Systems Biology, Jefferson, AR, USA

[3]US Food and Drug Administration, Center for Food Safety and Applied Nutrition, Laurel, MD, USA

[4]Metabolon, Inc., Durham, NC, USA

[5]Department of Psychiatry and Behavioral Sciences, Duke Institute for Brain Sciences, Duke University, Durham, NC, USA

[6]Experimental Toxicology and Ecology, BASF SE, Ludwigshafen am Rhein, Germany

[7]Bristol-Myers Squibb Pharmaceutical Co., Princeton, NJ, USA

[8]CAAT-Europe, University of Konstanz, Germany

Summary

Metabolomics promises a holistic phenotypic characterization of biological responses to toxicants. This technology is based on advanced chemical analytical tools with reasonable throughput, including mass-spectroscopy and NMR. Quality assurance, however – from experimental design, sample preparation, metabolite identification, to bioinformatics data-mining – is urgently needed to assure both quality of metabolomics data and reproducibility of biological models. In contrast to microarray-based transcriptomics, where consensus on quality assurance and reporting standards has been fostered over the last two decades, quality assurance of metabolomics is only now emerging. Regulatory use in safety sciences, and even proper scientific use of these technologies,

---

[*]A report of t⁴ – the transatlantic think tank for toxicology, a collaboration of the toxicologically oriented chairs in Baltimore, Konstanz and Utrecht sponsored by the Doerenkamp-Zbinden Foundation; participants do not represent their institutions and do not necessarily endorse all recommendations made.

Correspondence to: Thomas Hartung, MD PhD, Center for Alternatives to Animal Testing, Johns Hopkins Bloomberg School of Public Health, 615 North Wolfe Street, W7032, Baltimore, MD 21205, USA, thartun1@jhu.edu.

Conflict of interest

The views presented in this article do not necessarily reflect those of the Food and Drug Administration.

demand quality assurance. In an effort to promote this discussion, an expert workshop discussed the quality assurance needs of metabolomics.

The goals for this workshop were 1) to consider the challenges associated with metabolomics as an emerging science, with an emphasis on its application in toxicology and 2) to identify the key issues to be addressed in order to establish and implement quality assurance procedures in metabolomics-based toxicology. Consensus has still to be achieved regarding best practices to make sure sound, useful, and relevant information is derived from these new tools.

1 Introduction

Recent developments in safety testing regulations have initiated global changes in risk assessment. Emerging techniques like omics technologies could make toxicity testing more efficient in terms of time, cost, mechanistic understanding, and relevance to humans. Many challenges, however, need to be addressed to ensure robust and informative results sufficient for solid decision-making. Even though some omics technologies have been used for more than a decade, there is still ongoing discussion about the reproducibility of experiments and the comparability of results at different sites and on different platforms.

In Baltimore, Maryland in November 2013, the Johns Hopkins Center for Alternatives to Animal Testing (CAAT) organized a "Quality Assurance of Metabolomics" workshop with members of the NIH Research Project "Human Toxome" consortium (Bouhifd et al., 2014, 2015) together with invited experts from academia, industry, and regulatory agencies. This report highlights aspects of the presentations and discussions that took place at the workshop. It should be noted that this is not a consensus report, i.e., not every aspect of the report represents the view of all coauthors or their organizations.

Recent publications from the National Research Council, the US EPA's (Environmental Protection Agency) computational toxicology research programs, along with the European REACH (Registration, Evaluation, Authorisation and Restriction of Chemicals) and other cosmetics legislation, are among the drivers of the current landscape changes in risk assessment and toxicity testing (Hartung, 2010, 2011). At the center of this unique advance is the conviction that emerging sciences and techniques, such as omics technologies, high-throughput screening, and computational toxicology, could make toxicity testing more efficient in terms of time, cost, animal use, and relevance to human mechanisms (Leist et al., 2008; Hartung, 2009). This conceptual framework offers many opportunities for modern toxicology, but many challenges need to be addressed to ensure sufficiently robust and informative results. The omics technologies, in particular, contribute to our understanding of toxicity mechanisms and, although some have been extensively used for more than a decade (e.g., microarrays), the reproducibility of experiments and the comparability of results at different sites and on different platforms is still subject to ongoing debate. Consensus has yet to be achieved concerning best practices in many critical topics, such as the experimental

design and protocols for sample preparation and handling, data processing, statistical analysis, and interpretation. One major challenge is how to ensure that sound, useful and relevant information is derived from these new tools. Quality assurance is the first response. The diversity of the technological platforms, complexity of biological systems, and variety of analytical and computational methods make it critical to adopt measures and procedures for ensuring the quality of the data.

Metabolomics, an interdisciplinary science that combines analytical chemistry, biochemistry, statistics, and bioinformatics, is one of the most promising omics tools in the post-genome era. It is primarily the comparative analysis of the endogenous metabolites present in any biological system at a given physiological state. Metabolomics also includes aspects of patho-biochemistry, systems biology, and molecular diagnostics when applied to toxicology (Griffiths et al., 2010). Its approaches have been applied in clinical settings and have been increasingly expanded to other fields (such as toxicology), because they have the ability to provide information that allows to better understand the mechanisms of toxicity (Craig et al., 2006; Heijne et al., 2005; Ruepp et al., 2002; Schnackenberg et al., 2006, 2009; Montoya et al., 2014). From an analytical perspective, the goal of metabolomics in toxicology studies is to "*achieve a comprehensive measurement of the metabolome and how it changes in response to stressors, with biological payoff being an illumination of the relationship between the perturbations and affected biochemical pathways*" (Robertson and Lindon, 2005). Toxicological applications have been detailed in many publications (Ramirez et al., 2013; Bouhifd et al., 2013; Robertson, 2005). In early 2000, metabolomics was suggested for the first time as a new technique for rapid toxicity screening (Robertson and Bulera, 2000), was used in academic research (to predict liver and kidney toxicity *in vivo*) Lindon et al., 2005)), and also in industry to elucidate toxicological modes of action allowing for early safety decisions and lowering the cost through reduced animal studies (van Ravenzwaay et al., 2012). *In vitro* applications are emerging and have been driven by two major factors: 1) the call for a better understanding of biochemical changes induced by a toxic insult in a defined and controllable experimental system and 2) the increasing requirement to move towards the use of human-relevant, non-animal alternatives (Ramirez et al., 2013). *In vitro* measurements of intracellular metabolites have allowed for organ-specific *in vitro* toxicity testing, e.g., neurotoxicity (van Vliet et al., 2008), renal toxicity (Ellis et al., 2011), hepatotoxicity (Ruiz-Aracama et al., 2011), mitochondrial toxicity (Balcke et al., 2011), and lung toxicity (Vulimiri et al., 2009).

Undoubtedly, the promise of metabolomics in various scientific disciplines, including *in vitro* toxicology, is recognized. Nevertheless, many obstacles must be addressed before the discipline can achieve its full potential. Besides the challenges inherent to any toxicological study, we discussed the issues specific to metabolomics with an emphasis on *in vitro* applications. These included quality assurance practices in academia and regulatory agencies and also aspects of conducting metabolomics studies in industrial settings.

2 Quality assurance in toxicology studies

Quality assurance is fundamental to all good scientific practice. The maintenance of high standards is essential for ensuring the reproducibility, reliability, credibility, acceptance, and

proper application of the results generated. The challenges and limitations of models and test methods in toxicology have been recognized and discussed (Hartung, 2009, 2011, 2013). Currently, toxicological risk assessments rely mainly on *in vivo* animal experimentation that is often expensive and follows test guidelines that are usually a few decades old. (However, by adding omics measurements to such studies, the information content and therefore scientific quality of such *in vivo* studies can be significantly increased). The throughput is low, preventing many substances from being adequately assessed (Grandjean and Landrigan, 2006; Judson et al., 2009). Selecting a test species that will best predict the human response is also challenging. On the other hand, *in vitro* toxicology studies depend significantly on cell models that differ in many aspects from normal physiology, making them difficult to reproduce in culture (Hartung, 2007a). Besides these intrinsic challenges, the discipline suffers from a lack of standards in methods and model standardization, and inefficient documentation and reporting.

Despite these problems, guidance has been developed that acknowledges the inherent variation of *in vitro* test systems and promotes standardization. Good Cell Culture Practice (GCCP) sets the minimum standards for any *in vitro* work involving cell and tissue cultures (Hartung et al., 2002). It aims to reduce uncertainty in the development and application of *in vitro* procedures by encouraging the establishment of principles for greater international harmonization, standardization, and rational implementation of laboratory practices, quality control systems, safety procedures, and reporting (Coecke et al., 2005). This guidance is comparable to the OECD Principles of Good Laboratory Practice (GLP) (OECD, 2004) (the two have actually cross-fertilized each other), which cannot normally be fully implemented in basic research because of cost and lack of flexibility. The requirement that all personnel need to be fully trained before executing tests, in particular, cannot be met in an academic setting, where much of the work is done by students. However, through some simple actions based on GLP principles, a higher level of quality can be achieved even in academic research.

Quality assurance of *in vitro* methods could be further reinforced by the principles of validation. The term 'validation' is used differently in different contexts. All fields of science and engineering technically validate methods with regard to the internal performance parameters of a method. Formal validation was introduced for the acceptance of regulatory test methods to help agencies decide on the implementation of new tools, especially those replacing animal experiments. Predictivity, usually in comparison to the traditional (animal) test method, is also validated in addition to the internal performance characteristics (reliability). Guidelines were primarily developed by three organizations: the Organisation for Economic Cooperation and Development (OECD) (OECD, 2005), the European Centre for the Validation of Alternative Methods (ECVAM) (Hartung et al., 2004), and the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM). Criteria to be addressed in a validation exercise include: test definition (including purpose, need, and scientific basis), relevance of the test method, repeatability and reproducibility, inter-laboratory transferability, predictive capacity, and applicability domain. Questions arise about whether the validation process, as it has been formalized over the last two decades, might meet the challenges of emerging methods and technologies (such as omics), especially in toxicity testing (Hartung, 2007b; Leist et al., 2012).

A pragmatic approach would adapt some of the principles and criteria listed above for ensuring some degree of quality that could lay the ground for a quality assurance system in toxicometabolomics studies. In other words, toxicometabolomics needs quality-controlled model systems, and simply adding an omics endpoint does not make a test better.

3 Toxicometabolomics quality assurance

The comprehensive analysis of small molecules and their changes in response to stressors is a challenging exercise. The success of a toxicometabolomics study often depends on multiple experimental, analytical, and computational steps. A typical workflow used in metabolomics studies is outlined in Table 1. This process involves many steps, starting from the actual study design, which depends on the adopted metabolomic approach. Indeed, many approaches are currently used in metabolomics studies ranging from fingerprinting to non-targeted profiling to targeted analysis (Robertson et al., 2011). The study design involves the selection of the test system (e.g., animal model, *in vitro* cell culture), the type of the stressor, and the route of exposure. The choice of the biological matrix is also important; typical matrices analyzed include blood, serum, and urine, as well as intra- and extra-cellular extracts. The extraction method has to be specifically developed and optimized for each matrix before sample preparation and analysis. Once the metabolite data are generated, they are handled in order to prepare and reduce analytical instrument raw data (e.g., MS chromatograms) to data matrices for further analysis. This typically involves the execution of a series of tasks ranging from low-level processing (background correction, feature detection, normalization, alignment, etc.) to higher level processing consisting of various tools and methods for interpretation and visualization of the pre-processed data.

In his presentation during the workshop, Dr Donald Robertson stated, "*In the past fifteen years, I have been involved in approximately 500 metabolomic studies. Of those studies that failed, more than 90% of the failures could be attributed to errors in study design, study conduct, sample collection, or sample preparation. Relatively few failed due to analytical reasons.*" Furthermore, the American Society for Mass Spectrometry (ASMS) survey of about 600 participants at its 2009 conference (American Society for Mass Spectrometry, 2009) clearly showed that the analytical element was considered of less concern than the interpretation of metabolomics data and its biological significance. We will summarize below the main elements of a metabolomics study in toxicology and related quality measures.

4 Study design

The suitable design of scientific studies is the first necessary condition to ensure robust and trustworthy conclusions. By definition, experimental design is the process of planning data gathering in order to meet predefined objectives and answer the research question of interest as clearly as possible. Experimental design takes into account specific considerations for the experiment type (e.g., treatment vs. control), experimental variables (e.g., dose response, time dependence), experiment controls, and acceptance criteria. The number of replicates considered in the study is a critical determinant of the quality of the experiment. There is no "magic" number relative to the number of replicates needed, since this will depend on the

multiple sources of variability in the experiment. The Metabolomics Standards Initiative (MSI) (Sumner et al., 2007) suggests a minimum of three to five replicates, with a preference of biological replication (i.e., repetitive analyses of different samples obtained under the same experimental conditions) over technical replication (repetitive analyses of the same sample). It is a good practice, however, to conduct a preliminary pilot study to evaluate the data variation under the specific conditions and to perform a power analysis to guide the determination of the optimal number of replicates. Traditional power analysis would calculate the number of replicates based on the expected effect strength, the significance level aimed for, and the variability of the measurement in the model. This is not as easy for metabolomics, as effect strengths are typically small, multiple parallel measurements have to be accounted for (false discovery rate corrections), and the very different variability for different metabolites impair such calculations. Therefore, extensive evaluation of (control) variability over time is needed to a) ensure reproducibility and robustness, and therefore reliability of the measurements and b) enable the identification of biologically significant results. For the latter, statistically significant changes occurring in an experiment can be compared to the historical control data and variability of the respective metabolite (van Ravenzwaay et al., in press). In addition, the use of quality control (QC) samples is recommended and has been increasingly adopted (Dunn et al., 2012). These are usually representative of all the samples being analyzed in the study to represent a "mean" of all analyzed metabolites (Gika et al., 2007). A QC sample in the context of metabolomics could be obtained by pooling all samples in the study or by using additional control groups and pooling the samples derived from these control groups. Aliquots of a unique "pooled" QC sample, applied for an entire study at regular intervals, can help determine variations of all processes involved in terms of data acquisition (e.g., retention time and abundance) and also in data pre-processing (e.g., feature extraction). Furthermore, blank samples, which are analyte-free and prepared exactly as the test samples, give an idea of the overall levels of contamination and carryover. An additional quality measure in the experimental design is randomization of the sample analysis sequence. This procedure minimizes the bias introduced when preparing and analyzing replicate samples jointly. According to our own experience, QC samples account for about 30% and up to 50% of the total number of injections in an LC-MS run.

5 Sampling and extraction

Differences in methods to collect, prepare, store, and otherwise handle samples are important sources of bias in life sciences in general and have been a major problem in biomarker detection as, for example, noted by Teahan et al. (2006), where allegedly promising results were sometimes difficult to reproduce and validate. Diverse biological systems, such as microorganisms, plants, biofluids or mammalian cells, are studied in metabolomics, making it challenging to devise a unique method or guideline for sample collection and preparation.

There are rather general good practices and examples, such as those included in the NCI best practices for biospecimen resources (National Cancer Institute, 2011). Although the guideline is primarily intended for human specimens, it provides technical and operational best practices to ensure levels of consistency and standardization. It also identifies a variety

of factors that may affect biospecimen quality and thus research results. Recommendations include, where possible, the use of validated methods, training of technical staff, inclusion of appropriate quality control and reference samples, randomization, and standardized methods for documenting. Guidelines for the use of biofluids in proteomics studies (Rai et al., 2005) are also applicable to metabolomics. They evaluate a number of pre-analytical variables that can potentially impact the outcome. These include, among others, the sample type, the collection system, the processing methods, and storage parameters. During the workshop, Dr Hennicke Kamp reported that standardization of all steps, from collecting the sample from a biological system through sample preparation and metabolite extraction is essential to obtaining robust and reliable metabolome data. The large diversity of physico-chemical properties of the metabolome poses an additional challenge. Chemical compounds analyzed differ in molecular weight, polarity, boiling and melting point, functional groups, etc. Moreover, these compounds are present in concentrations that span orders of magnitude within the same sample (Maier et al., 2010).

Metabolomics involves, therefore, the analysis of a heterogeneous chemical space and across a broad dynamic range, which makes considerations for standardization of protocols challenging. An efficient method would allow the adequate recovery of the largest number of metabolites from samples while preventing the exclusion of compounds due to their physical or chemical properties (Winder et al., 2008). While it is obvious that no unique analytical method can fulfill these requirements, consistent quenching, extraction protocols, as well as adequate sample storage would limit variability in metabolite extraction and analysis (Zhou et al., 2012). Documentation in the form of Standard Operating Procedures (SOPs), optimized for the specific metabolomic application, should be detailed enough to allow an unambiguous and reproducible execution of the procedure (Bouhifd et al., 2013).

6 Metabolomics data complexity

A fundamental characteristic of metabolomics is the huge diversity of chemicals involved. Unlike a genome, which involves only four bases, and the proteome with its twenty amino acids, the metabolome consists of at least a few thousand chemicals (Wishart, 2011). The various chemical and physical properties of these molecules would require a combination of analytical technologies to obtain good coverage. Historically, the analytical tools of choice in metabolomics have been NMR and MS. The latter is combined with a chromatographic separation technique such as liquid chromatography (LC) or gas chromatography (GC). The characteristics, advantages, limitations, and differences between the technologies and platforms have been extensively described in several review articles and will not be addressed here (Kaddurah-Daouk et al., 2008; Robertson, 2005; Dunn and Ellis, 2005).

Despite the recent technology advances, no single analytical platform is a perfect tool for metabolomics, with all having advantages and limitations, although LC-MS now appears to be the preferred technology in many studies. Besides the biological variability described earlier, metabolomics data can suffer from analytical variability. This includes mainly drifts in retention times, altered instrument sensitivity, and – very rarely – drifts in measured mass to charge ratio (m/z) values. Although the technologies involved are complex, it is well accepted in the community that the analytical process is not the main limitation.

One of the biggest challenges in metabolomics remains metabolite identification. An accurate identification of the chemicals involved in any particular study is necessary to derive meaningful biological information. It is now a very common practice to generate metabolomic datasets comprising thousands of "features," but their identification is certainly not straightforward. A mass spectrometry measurement typically results in a list of entities represented by mass-to-charge (m/z) ratio, retention time (RT), and intensity. These parameters might be informative but do not provide a direct chemical annotation of the entity in question. One needs first to convert the raw analytical data to metabolites (namely chemicals).

The accuracy and confidence in this conversion (identification, in other words) vary widely because of the complexity of the process and its dependence on the analytical platform and robustness of the methods applied, as well as the databases and resources used (Creek et al., 2014). Indeed, this process should discriminate not only between metabolites with different masses, but also those with the same nominal mass but different molecular formula and monoisotopic mass, and also metabolites with the same nominal and monoisotopic masses but different chemical structures. In addition, a single metabolite can form multiple different ion types (in the case of electrospray ionization, for example) such as sodium and potassium adducts, along with the standard protonated form (Dunn et al., 2013). Diverse strategies have been adopted with different levels of confidence. These confidence levels were divided into four categories by a dedicated working group of the Metabolomics Standards Initiative (MSI) and the following definitions were proposed (Sumner et al., 2007):

1. Confidently identified compounds where at least two orthogonal properties (e.g., m/z, RT, fragmentation mass spectrum) of the candidate metabolite are verified with an authentic reference standard under the same analytical conditions;

2. Putatively annotated compounds where the physicochemical properties are compared to chemical library without reference to authentic standards;

3. Putatively characterized compound classes based upon characteristic physicochemical properties of a chemical class of compounds (e.g., lipids), or by spectral similarity to known compounds of a chemical class, and;

4. Unknown compounds which are unidentified and unclassified metabolites that can still be differentiated using spectral data.

Although the exact basis for what constitutes valid metabolite identification could be debated, a major contribution of the MSI is the detailed formulation of the reporting needs of the identification procedure and its performance. Different strategies could be adopted, but metabolites are typically characterized on the basis of accurate mass, retention time, and tandem mass spectrometry (MS/MS) data. First, m/z values are searched in metabolite databases (peak annotation). When a hit is returned within the expected error of the mass spectrometer, the annotation is still putative and sometimes needs manual curation. To increase the level of confidence as described above, an authentic reference standard is used, and retention time and/or MS/MS data is generated in the same analytical conditions and compared to that from the biological sample (Patti et al., 2012). Following this description,

two important elements emerge regarding quality assurance – namely, metabolite databases and reference standards.

Databases (DBs) are sources of chemical information in the form of web-based or locally hosted services, and fulfill many objectives. They are of different types and contain diverse information such as metabolic pathway information, compound-specific information, spectral information, disease/physiology information, or organism-specific metabolomic information (Wishart et al., 2009). Although these resources are increasingly helpful, some limitations still exist. DBs have different coverages, and although they might be partly complementary, missing metabolite identifiers and ambiguous names for metabolites affect the comparison (Stobbe et al., 2011). Considerable manual intervention and curation is required to unify the DBs, prompting a need for standardization of the metabolite names and identifiers. This effort is challenging since it needs not only expert knowledge but also some degree of automation.

7 Metabolomics in a systems biology context

Systems biology has been defined as the "*study of the mechanisms underlying complex biological processes as integrated systems of many diverse, interacting components. It involves (1) a collection of large sets of experimental data (by high-throughput technologies and/or by mining the literature of molecular biology and biochemistry); (2) proposal of mathematical models that might account for at least some significant aspects of this data set; (3) accurate computer solution of the mathematical equations to obtain numerical predictions; and (4) assessment of the quality of the model by comparing numerical simulations with the experimental data*" (Ferrario et al., 2014). Metabolomics is a key tool for producing such large datasets and lends itself especially to the characterization of phenotypic changes in a systems toxicology approach (Hartung et al., 2012). The major advantage of toxicology compared to other disciplines is that we have the "disease agent" at hand, i.e., we can experimentally induce and monitor pathogenesis and are not restricted to comparison of healthy versus diseased tissue and biofluids. The mathematical models of metabolism, however, will have to reflect the dynamics of the networked systems and will likely depend on the measurement of metabolite fluxes. This topic will open up further aspects of quality assurance beyond what was discussed in this workshop. The Human Toxome project is pioneering some of this (Bouhifd et al., 2015) and not surprisingly prompted the need for discussion about quality assurance via this workshop.

8 Future directions: Need for collaborative activities

The discussions of this workshop showed that a combination of expert consensus – for example, reporting standards and good practices – and experimental assessments (e.g., ring trials between laboratories) is needed. The importance of both analytical and biological validation was emphasized. Validation in a broad sense demonstrates suitability for an intended purpose or "fitness for purpose." To simplify, we may distinguish two main components: reliability (robustness/quality/confidence) and relevance (usefulness/biological utility). These two elements, if satisfied, will ensure reproducible and meaningful research.

Such discussions took place for transcriptomics a decade ago (for example, the first transatlantic consensus workshop on validation of transcriptomics (Corvi et al., 2006)).

The present workshop was prompted by the ongoing Human Toxome project (Bouhifd et al., 2013, 2014). This project aims for the identification of pathways of toxicity (Kleensang et al., 2014) by using a multi-omics approach (Hartung and McBride, 2011). For the purpose of this project, it will be necessary to assess especially whether intracellular metabolomics, i.e., the metabolomic analyses of cell extracts, are sufficiently robust to allow reliable and consistent identification of specific changes in newly produced or altered metabolites in response to a known toxicant stimulus. This is a prerequisite for unambiguous identification of the underlying molecular mechanisms or pathways of toxicity. The overall strategy would consist of generating standardized biological samples and assessing within-run, within-lab, and between-lab reproducibility of metabolomics analysis. Unfortunately, such quality assurance studies lack appeal for most funding bodies, but they have the potential not only to move forward a given project but to further the proper use of an entire technology.

Acknowledgments

References

Balcke GU, Kolle SN, Kamp H, et al. Linking energy metabolism to dysfunctions in mitochondrial respiration – a metabolomics in vitro approach. Toxicol Lett. 2011; 203:200–209. http://dx.doi.org/10.1016/j.toxlet.2011.03.013. [PubMed: 21402135]

Bouhifd M, Hartung T, Hogberg HT, et al. Review: Toxicometabolomics. J Appl Toxicol. 2013; 33:1365–1383. http://dx.doi.org/10.1002/Jat.2874. [PubMed: 23722930]

Bouhifd M, Hogberg HT, Kleensang A, et al. Mapping the human toxome by systems toxicology. Basic Clin Pharmacol Toxicol. 2014; 115:24–31. http://dx.doi.org/10.1111/bcpt.12198. [PubMed: 24443875]

Bouhifd M, Andersen ME, Baghdikian C, et al. The human toxome project. ALTEX. 2015; 32:112–124. http://dx.doi.org/10.14573/altex.1502091. [PubMed: 25742299]

Coecke S, Balls M, Bowe G, et al. Guidance on good cell culture practice. A report of the second ECVAM task force on good cell culture practice. Altern Lab Anim. 2005; 33:261–287. [PubMed: 16180980]

Corvi R, Ahr HJ, Albertini S, et al. Meeting report: Validation of toxicogenomics-based test systems: ECVAM-ICCVAM/NICEATM considerations for regulatory use. Environ Health Perspect. 2006; 114:420–429. http://dx.doi.org/10.1289/ehp.8247. [PubMed: 16507466]

Craig A, Sidaway J, Holmes E, et al. Systems toxicology: Integrated genomic, proteomic and metabonomic analysis of methapyrilene induced hepatotoxicity in the rat. J Proteome Res. 2006; 5:1586–1601. http://dx.doi.org/10.1021/pr0503376. [PubMed: 16823966]

Creek DJ, Dunn WB, Fiehn O, et al. Metabolite identification: Are you sure? And how do your peers gauge your confidence? Metabolomics. 2014; 10:350–353. http://dx.doi.org/10.1007/s11306-014-0656-8.

Dunn WB, Ellis DI. Metabolomics: Current analytical platforms and methodologies. Trac-Trends In Analytical Chemistry. 2005; 24:285–294. http://dx.doi.org/10.1016/j.trac.2004.11.021.

Dunn WB, Wilson ID, Nicholls AW, et al. The importance of experimental design and QC samples in large-scale and MS-driven untargeted metabolomic studies of humans. Bioanalysis. 2012; 4:2249–2264. http://dx.doi.org/10.4155/bio.12.204. [PubMed: 23046267]

Dunn WB, Erban A, Weber RJM, et al. Mass appeal: Metabolite identification in mass spectrometry-focused untargeted metabolomics. Metabolomics. 2013; 9:S44–S66. http://dx.doi.org/10.1007/s11306-012-0434-4.

Ellis JK, Athersuch TJ, Cavill R, et al. Metabolic response to low-level toxicant exposure in a novel renal tubule epithelial cell system. Mol Biosyst. 2011; 7:247–257. http://dx.doi.org/10.1039/c0mb00146e. [PubMed: 21103459]

Ferrario D, Brustio R, Hartung T. Glossary of reference terms for alternative test methods and their validation. ALTEX. 2014; 31:319–335. http://dx.doi.org/10.14573/altex.1403311. [PubMed: 24819604]

Gika HG, Theodoridis GA, Wingate JE, et al. Within-day reproducibility of an HPLC-MS-based method for metabonomic analysis: Application to human urine. J Proteome Res. 2007; 6:3291–3303. http://dx.doi.org/10.1021/pr070183p. [PubMed: 17625818]

Grandjean P, Landrigan PJ. Developmental neurotoxicity of industrial chemicals. Lancet. 2006; 368:2167–2178. http://dx.doi.org/10.1016/S0140-6736(06)69665-7. [PubMed: 17174709]

Griffiths WJ, Koal T, Wang Y, et al. Targeted metabolomics for biomarker discovery. Angew Chem Int Ed Engl. 2010; 49:5426–5445. http://dx.doi.org/10.1002/anie.200905579. [PubMed: 20629054]

Hartung T, Balls M, Bardouille C, et al. Good Cell Culture Practice. ECVAM Good Cell Culture Practice Task Force Report 1. Altern Lab Anim. 2002; 30:407–414. [PubMed: 12234246]

Hartung T. Food for thought ... on cell culture. ALTEX. 2007a; 24:143–152. http://www.altex.ch/All-issues/Issue.50.html?iid=87&aid=1. [PubMed: 17926373]

Hartung T. Food for thought ... on validation. ALTEX. 2007b; 24:67–80. http://www.altex.ch/All-issues/Issue.50.html?iid=86&aid=3.

Hartung T. Toxicology for the twenty-first century. Nature. 2009; 460:208–212. http://dx.doi.org/10.1038/460208a. [PubMed: 19587762]

Hartung T. Lessons learned from alternative methods and their validation for a new toxicology in the 21st century. J Toxicol Environ Health B Crit Rev. 2010; 13:277–290. http://dx.doi.org/10.1080/10937404.2010.483945. [PubMed: 20574902]

Hartung T. From alternative methods to a new toxicology. Eur J Pharm Biopharm. 2011; 77:338–349. http://dx.doi.org/10.1016/j.ejpb.2010.12.027. [PubMed: 21195172]

Hartung T, McBride M. Food for thought ... on mapping the human toxome. ALTEX. 2011; 28:83–93. http://dx.doi.org/10.14573/altex.2011.2.083. [PubMed: 21625825]

Hartung T, van Vliet E, Jaworska J, et al. Systems toxicology. ALTEX. 2012; 29:119–128. http://dx.doi.org/10.14573/altex.2012.2.119. [PubMed: 22562485]

Hartung T. Look back in anger – what clinical studies tell us about preclinical work. ALTEX. 2013; 30:275–291. http://dx.doi.org/10.14573/altex.2013.3.275. [PubMed: 23861075]

Heijne WH, Kienhuis AS, van Ommen B, et al. Systems toxicology: Applications of toxicogenomics, transcriptomics, proteomics and metabolomics in toxicology. Expert Rev Proteomics. 2005; 2:767–780. http://dx.doi.org/10.1586/14789450.2.5.767. [PubMed: 16209655]

Judson R, Richard A, Dix DJ, et al. The toxicity data landscape for environmental chemicals. Environ Health Perspect. 2009; 117:685–695. http://dx.doi.org/10.1289/ehp.0800168. [PubMed: 19479008]

Kaddurah-Daouk R, Kristal BS, Weinshilboum RM. Metabolomics: A global biochemical approach to drug response and disease. Annu Rev Pharmacol Toxicol. 2008; 48:653–683. http://dx.doi.org/10.1146/annurev.pharmtox.48.113006.094715. [PubMed: 18184107]

Kleensang A, Maertens A, Rosenberg M, et al. t4 workshop report: Pathways of toxicity. ALTEX. 2014; 31:53–61. http://dx.doi.org/10.14573/altex.1309261. [PubMed: 24127042]

Leist M, Hartung T, Nicotera P. The dawning of a new age of toxicology. ALTEX. 2008; 25:103–114. [PubMed: 18551234]

Leist M, Hasiwa N, Daneshian M, et al. Validation and quality control of replacement alternatives – current status and future challenges. Toxicol Res. 2012; 1:8–22. http://dx.doi.org/10.1039/C2tx20011b.

Lindon JC, Keun HC, Ebbels TM, et al. The Consortium for Metabonomic Toxicology (COMET): Aims, activities and achievements. Pharmacogenomics. 2005; 6:691–699. http://dx.doi.org/10.2217/14622416.6.7.691. [PubMed: 16207146]

Maier TS, Kuhn J, Muller C. Proposal for field sampling of plants and processing in the lab for environmental metabolic fingerprinting. Plant Methods. 2010; 6 http://dx.doi.org/10.1186/1746-4811-6-6.

Montoya GA, Strauss V, Fabian E, et al. Mechanistic analysis of metabolomics patterns in rat plasma during administration of direct thyroid hormone synthesis inhibitors or compounds increasing thyroid hormone clearance. Toxicol Lett. 2014; 225:240–251. http://dx.doi.org/10.1016/j.toxlet.2013.12.010. [PubMed: 24370789]

OECD. Advisory document of the working group on GLP – the application of the principles of GLP to in vitro studies. Series on Principles of Good Laboratory Practice and Compliance Monitoring. 2004; 14:1–18.

OECD. Environmental Health and Safety Monograph Series on Testing and Assessment No. 34. OECD Publishing; 2005. Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment.

Patti GJ, Yanes O, Siuzdak G. Metabolomics: The apogee of the omics trilogy. Nat Rev Mol Cell Biol. 2012; 13:263–269. http://dx.doi.org/10.1038/Nrm3314. [PubMed: 22436749]

Rai AJ, Gelfand CA, Haywood BC, et al. HUPO Plasma Proteome Project specimen collection and handling: Towards the standardization of parameters for plasma proteome samples. Proteomics. 2005; 5:3262–3277. http://dx.doi.org/10.1002/pmic.200401245. [PubMed: 16052621]

Ramirez T, Daneshian M, Kamp H, et al. Metabolomics in toxicology and preclinical research. ALTEX. 2013; 30:209–225. http://dx.doi.org/10.14573/altex.2013.2.209. [PubMed: 23665807]

Robertson DG, Bulera SJ. High-throughput toxicology: Practical considerations. Curr Opin Drug Discov Devel. 2000; 3:42–47.

Robertson DG. Metabonomics in toxicology: A review. Toxicol Sci. 2005; 85:809–822. http://dx.doi.org/10.1093/toxsci/kfi102. [PubMed: 15689416]

Robertson, DG., Lindon, JC. Metabonomics in toxicity assessment. Boca Raton, FL, USA: Taylor & Francis; 2005. http://dx.doi.org/10.1201/b14117

Robertson DG, Watkins PB, Reily MD. Metabolomics in toxicology: Preclinical and clinical applications. Toxicol Sci. 2011; 120(Suppl 1):S146–170. http://dx.doi.org/10.1093/toxsci/kfq358. [PubMed: 21127352]

Ruepp SU, Tonge RP, Shaw J, et al. Genomics and proteomics analysis of acetaminophen toxicity in mouse liver. Toxicol Sci. 2002; 65:135–150. http://dx.doi.org/10.1093/toxsci/65.1.135. [PubMed: 11752693]

Ruiz-Aracama A, Peijnenburg A, Kleinjans J, et al. An untargeted multi-technique metabolomics approach to studying intracellular metabolites of HepG2 cells exposed to 2,3,7,8-tetrachlorodibenzo-p-dioxin. BMC Genomics. 2011; 12:251. http://dx.doi.org/10.1186/1471-2164-12-251. [PubMed: 21599895]

Schnackenberg LK, Jones RC, Thyparambil S, et al. An integrated study of acute effects of valproic acid in the liver using metabonomics, proteomics, and transcriptomics platforms. OMICS. 2006; 10:1–14. http://dx.doi.org/10.1089/omi.2006.10.1. [PubMed: 16584314]

Schnackenberg, LK., Sun, J., Beger, RD. General, Applied and Systems Toxicology. John Wiley & Sons, Ltd; 2009. Metabolomics in Systems Toxicology: Towards Personalized Medicine. http://dx.doi.org/10.1002/9780470744307.gat217

Stobbe MD, Houten SM, Jansen GA, et al. Critical assessment of human metabolic pathway databases: A stepping stone for future integration. BMC Syst Biol. 2011; 5:165. http://dx.doi.org/10.1186/1752-0509-5-165. [PubMed: 21999653]

Sumner LW, Amberg A, Barrett D, et al. Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). Metabolomics. 2007; 3:211–221. http://dx.doi.org/10.1007/s11306-007-0082-2. [PubMed: 24039616]

Teahan O, Gamble S, Holmes E, et al. Impact of analytical bias in metabonomic studies of human blood serum and plasma. Anal Chem. 2006; 78:4307–4318. http://dx.doi.org/10.1021/ac051972y. [PubMed: 16808437]

van Ravenzwaay B, Herold M, Kamp H, et al. Metabolomics: A tool for early detection of toxicological effects and an opportunity for biology based grouping of chemicals – from QSAR to

QBAR. Mutat Res. 2012; 746:144–150. http://dx.doi.org/10.1016/j.mrgentox.2012.01.006. [PubMed: 22305969]

van Ravenzwaay B, Peter E, Krennrich G, et al. The development of a data base for metabolomics – looking back on 10 years of experience. in press.

van Vliet E, Morath S, Eskes C, et al. A novel in vitro metabolomics approach for neurotoxicity testing, proof of principle for methyl mercury chloride and caffeine. Neurotoxicology. 2008; 29:1– 12. http://dx.doi.org/10.1016/j.neuro.2007.09.007. [PubMed: 18023877]

Vulimiri SV, Misra M, Hamm JT, et al. Effects of mainstream cigarette smoke on the global metabolome of human lung epithelial cells. Chem Res Toxicol. 2009; 22:492–503. http://dx.doi.org/10.1021/tx8003246. [PubMed: 19161311]

Winder CL, Dunn WB, Schuler S, et al. Global metabolic profiling of Escherichia coli cultures: An evaluation of methods for quenching and extraction of intracellular metabolites. Anal Chem. 2008; 80:2939–2948. http://dx.doi.org/10.1021/Ac7023409. [PubMed: 18331064]

Wishart DS, Knox C, Guo AC, et al. HMDB: A knowledgebase for the human metabolome. Nucleic Acids Res. 2009; 37:D603–610. http://dx.doi.org/10.1093/nar/gkn810. [PubMed: 18953024]

Wishart DS. Advances in metabolite identification. Bioanalysis. 2011; 3:1769–1782. http://dx.doi.org/10.4155/Bio.11.155. [PubMed: 21827274]

Zhou B, Xiao JF, Tuli L, et al. LC-MS-based metabolomics. Mol Biosyst. 2012; 8:470–481. http://dx.doi.org/10.1039/c1mb05350g. [PubMed: 22041788]

Tab. 1

Typical metabolomics workflow

Study design	Problem formulation Experimental condition definition Toxicant treatment (time, route, formulation, etc.)
Sample preparation	Harvesting/sampling/preparing/storage Metabolite extraction
Data generation	Measurement (e.g., LC-MS) Data processing Feature selection
Confirmation	Metabolite identification Metabolite quantification
Conclusions (biological relevance)	Modeling Data interpretation Hypothesis generation and/or verification Reporting