Organisation for Economic Co-operation and Development

ENV/CBC/MONO(2023)41

Unclassified

English - Or. English

23 November 2023

ENVIRONMENT DIRECTORATE
CHEMICALS AND BIOTECHNOLOGY COMMITTEE

OECD Omics Reporting Framework (OORF): Guidance on reporting elements for the regulatory use of omics data from laboratory-based toxicology studies

Series on Testing and Assessment
No. 390

The Guidance Document is accompanied by an Excel template for reporting omics studies that is available at the following link: https://www.oecd.org/chemicalsafety/testing/omics-reporting-framework-reporting-template-2023.xlsx

JT03532487

OECD Environment, Health and Safety Publications

SERIES ON TESTING AND ASSESSMENT

NO. 390

OECD Omics Reporting Framework (OORF): Guidance on reporting elements for the regulatory use of omics data from laboratory-based toxicology studies

INTER-ORGANIZATION PROGRAMME FOR THE SOUND MANAGEMENT OF CHEMICALS

A cooperative agreement among **FAO, ILO, UNDP, UNEP, UNIDO, UNITAR, WHO, World Bank and OECD**

Environment Directorate

ORGANISATION FOR ECONOMIC COOPERATION AND DEVELOPMENT

Paris 2023

About the OECD

The Organisation for Economic Co-operation and Development (OECD) is an intergovernmental organisation in which representatives of 38 industrialised countries in North and South America, Europe and the Asia and Pacific region, as well as the European Commission, meet to co-ordinate and harmonise policies, discuss issues of mutual concern, and work together to respond to international problems. Most of the OECD's work is carried out by more than 200 specialised committees and working groups composed of member country delegates. Observers from several countries with special status at the OECD, and from interested international organisations, attend many of the OECD's workshops and other meetings. Committees and working groups are served by the OECD Secretariat, located in Paris, France, which is organised into directorates and divisions.

The Environment, Health and Safety Division publishes free-of-charge documents in twelve different series: **Testing and Assessment**; **Good Laboratory Practice and Compliance Monitoring**; **Pesticides**; **Biocides**; **Risk Management**; **Harmonisation of Regulatory Oversight in Biotechnology**; **Safety of Novel Foods and Feeds; Chemical Accidents; Pollutant Release and Transfer Registers; Emission Scenario Documents; Safety of Manufactured Nanomaterials;** and **Adverse Outcome Pathways.** More information about the Environment, Health and Safety Programme and EHS publications is available on the OECD's World Wide Web site (www.oecd.org/chemicalsafety/).

This publication was developed in the IOMC context. The contents do not necessarily reflect the views or stated policies of individual IOMC Participating Organizations.

The Inter-Organisation Programme for the Sound Management of Chemicals (IOMC) was established in 1995 following recommendations made by the 1992 UN Conference on Environment and Development to strengthen co-operation and increase international co-ordination in the field of chemical safety. The Participating Organisations are FAO, ILO, UNDP, UNEP, UNIDO, UNITAR, WHO, World Bank and OECD. The purpose of the IOMC is to promote co-ordination of the policies and activities pursued by the Participating Organisations, jointly or separately, to achieve the sound management of chemicals in relation to human health and the environment.

This publication is available electronically, at no charge.

- **Also published in the Series on Testing and Assessment: link**

For this and many other Environment, Health and Safety publications, consult the OECD's World Wide Web site (www.oecd.org/chemicalsafety/)

or contact:

OECD Environment Directorate,
Environment, Health and Safety Division
2 rue André-Pascal
75775 Paris Cedex 16
France

E-mail: ehscont@oecd.org

OECD Environment, Health and Safety Publications

Foreword

OECD Member Countries are interested in increasing the use of omics technologies in regulatory toxicology to advance chemical risk assessment. In this context, the OECD has developed this Guidance Document and accompanying Excel template for reporting omics studies, which is referred to as the OECD Omics Reporting Framework (OORF). The OORF is intended to facilitate data sharing for omics toxicology experiments, increase the transparency of omics data processing approaches used, enable quality assessment of the study experiment and data, and promote reproducibility.

While the OORF guidance provides information on reporting for omics studies, it is not intended to provide technical guidance on best practices for omics study execution, data processing or analysis. The primary goal is to provide a clear and consistent framework for reporting each element of an omics study intended for use in regulatory toxicology, from study design through to data analysis. This is viewed as critical for regulatory use of omics studies.

The OORF is being tested for efficacy through trials with expert users of different types of omics technology each omics technology, bioinformaticians, and regulatory stakeholders. These trials help to ensure that the data provider and end-user understand each of the reporting elements, are able to use the OORF to effectively share information, and ensure reproducibility of an omics experiment.

The development of the OORF was led by delegates of the OECD's Extended Advisory Group in Molecular Screening and Toxicogenomics (EAGMST), with leadership from Canada (Health Canada and the University of Ottawa), the United Kingdom (University of Birmingham), and the United States (Environmental Protection Agency). The working group members spanned government, academia and industry and are listed in Table 1. The group includes expert members involved in developing each module of the OORF as well as external stakeholders who participated in trials of the modules.

Initial drafts of the OORF were reviewed on two occasions and approved by members of EAGMST. This report is published under the responsibility of the OECD Chemicals and Biotechnology Committee.

Table 1. List of contributors

Name	Affiliation	Representing
Joshua Harrill- **Transcriptomics co-lead of OORF**	US Environmental Protection Agency	US
Mark Viant- **Metabolomics lead of OORF**	University of Birmingham	UK
Carole Yauk- **Transcriptomics co-lead of OORF**	University of Ottawa	CA
Matthew Meier- **Trials co-lead**	Health Canada	CA
Scott Auerbach- **BMD lead**	National Institute of Environmental Health Sciences	US
Lyle Burgoon- **DAM lead**	US Army Engineer Research and Development Center	US
Florian Caiment- **RNA-Seq lead**	Maastricht University	NL
Raffaella Corvi- **TERM lead**	European Commission, Joint Research Centre	EU
Tim Gant- **Project conception contributor**	Public Health England	UK
Jason O-Brien- **qRT-PCR lead**	Environment and Climate Change Canada	CA
Vikrant Vijay- **Microarray lead**	US Food and Drug Administration	US
Rusty Thomas- **EAGMST co-chair**	US Environmental Protection Agency	US
Maurice Whelan- **EAGMST co-chair**	European Commission, Joint Research Centre	EU
Rick Beger	US Food and Drug Administration	US
Mounir Bouhifd	European Chemicals Agency	EU

Ivana Campia	European Commission, Joint Research Centre	EU
Donatella Carpi	European Commission, Joint Research Centre	EU
Tao Chen	US Food and Drug Administration	US
Brian Chorley	US Environmental Protection Agency	US
John Colbourne	University of Birmingham	UK
Claire O'Donovan	European Molecular Biology Laboratory	EU
Laurent Debrauwer	Paul Sabatier University, Toulouse	FR
Timothy Ebbels	Imperial College London	UK
Drew Ekman	US Environmental Protection Agency	US
Frank Faulhammer	BASF	BIAC
Laura Gribaldo	European Commission, Joint Research Centre	EU
Gina Hilton	PETA Science Consortium International	ICAPO
Stephanie Jones	Environment and Climate Change Canada	CA
Aniko Kende	Syngenta	BIAC
Thomas Lawson	Michabo Health Science Ltd	UK
Sofia Batista-Leite	European Commission, Joint Research Centre	EU
Pim Leonards	Vrije Universiteit Amsterdam	NL
Mirjam Luijten	Dutch National Institute for Public Health and the Environment	NL

Alberto Martin	European Chemicals Agency	EU
Laura Moussa	US Food and Drug Administration	US
Serge Rudaz	University of Geneva, Swiss Centre for Applied Human Toxicology	CH
Oliver Schmitz	BASF	BIAC
Tomasz Sobanski	European Chemicals Agency	EU
Volker Strauss	BASF	BIAC
Monica Vaccari	Agency for Prevention, Environment and Energy (Arpae)	IT
Ralf Weber	University of Birmingham	UK
Antony Williams	US Environmental Protection Agency	US
Andrew Williams	Health Canada	CA

Table of contents

Introduction to OECD Omics Reporting Framework Guidance Document

Omics methodologies are increasingly applied in research and regulatory toxicology to provide a greater understanding of mechanisms of toxicity. The ability of omics to measure molecular initiating events and downstream molecular phenotypes that are predictive of toxicity paves the way for applications in hazard (or adverse outcome) identification, mode of action analysis, chemical grouping to inform read-across and characterising potency via omic points-of-departure (Krewski et al., 2019). However, while mature technologies are readily available and applied for research purposes, application in regulatory decision-making has been relatively limited to date.

Significant barriers to routine application of omics in regulatory decision-making include: 1) lack of transparency for data processing methods used to convert raw data into an interpretable list of observations; and 2) lack of standardisation in reporting to ensure that omics data, associated metadata and the methodologies used to generate results are available for review by stakeholders, including regulators. To promote regulatory uptake, a comprehensive reporting framework was proposed to thoroughly document the components of an omics study (Sauer et al., 2017). Such a framework would increase the transparency for data processing methods used to convert raw omics data into an interpretable list of observations and ensure that all of the required data, associated metadata and analytical processes are readily available for review by end-users in the regulatory community (Buesen et al., 2017; Bridges et al., 2017; Gant et al., 2017; Kauffmann et al., 2017).

In response to this proposed need, the OECD Extended Advisory Group in Molecular Screening and Toxicogenomics (EAGMST) adopted a project in 2017 to develop guidance for reporting omics data types. This work was conceived in ECETOC workshops from 2015 (Buessen et al., 2017) supported by ECETOC and CEFIC/LRI for the MEtabolomics standaRds Initiative in Toxicology (MERIT) project and the Optimal Data Analysis Framework (ODAF). The OECD EAGMST project was divided into two working groups developing reporting frameworks for transcriptomics and metabolomics, respectively; it was decided to pursue proteomics in future work. Elements of the OECD reporting frameworks are based on previously established frameworks for toxicology study annotation for data sharing and regulatory application (e.g. Fostel et al., 2007; McConnell et al., 2014; Schneider et al., 2019; Segal et al., 2015), as well as previously established frameworks for annotation of data from transcriptomic studies (e.g. Brazma et al., 2001; Conesa et al., 2016; Parkinson et al., 2005; Waters et al., 2003). These frameworks focus primarily upon annotation of data (raw and normalised), samples, sample to data relationships and technology-specific feature annotation. The framework developed in this OECD project includes all of these elements, but also provides a means to document the computational steps used to analyse the omics data and generate downstream results that may be of use in a regulatory decision-making context. The metabolomics working group took advantage of the extensive work within the MERIT project (Viant et al. 2019). MERIT produced an initial set of best practice guidelines and minimal reporting standards for the acquisition, processing and statistical analysis of untargeted metabolomics and targeted metabolite data in the context of regulatory toxicology.

The outcome of this OECD project is a single, integrated, modular reporting framework known as the OECD Omics Reporting Framework (OORF). At present, certain modules within the OORF are specific to transcriptomic and metabolomic technologies. The modular structure readily enables integration of proteomics or additional technologies in future iterations of the OORF when these technologies reach a state of maturity for use in regulation. The OORF focuses specifically on reporting omics studies in toxicology and is not intended to recommend best practices. Its primary purpose is to address the aforementioned barriers to the adoption of omics

data in regulatory toxicology and to foster and encourage international acceptance and use of omics data, for example in the context of the OECD's programme on chemical safety (https://www.oecd.org/chemicalsafety/). Reporting using the OORF will provide the essential information needed to evaluate study designs, data quality and applicability to regulatory decision-making processes and maximise the likelihood that the analytical results of an omics experiment can be reproduced.

Objectives

This OORF Guidance Document is intended to describe the information that should be reported when an omics technology is applied in the context of regulatory decision-making, to enable assessment of the quality of the study from its design through the collection, analysis, and reporting of data. The guidance does not extend to the interpretation of these data. Adherence to such a reporting framework is also anticipated to maximise the likelihood that the results can be reproduced and potentially reused in the form of a knowledge base.

The information to be reported includes the experimental design, quality assurance and quality control (QA/QC), sampling of biological specimens, sample processing and extraction of biological molecules, data acquisition and processing, annotation and/or identification of the molecules, and statistical analyses specific to the regulatory application. The specific elements used in a report submitted to a regulator will depend on the context of use of the omics assay.

Scope

The OORF is a tool for documenting the details of laboratory-based toxicology studies that apply omics technologies: i.e. assays that measure the abundance of many molecular endpoints simultaneously and thus provide highly multiplexed outputs. The OORF is appropriate for use in documenting experiments involving the use of either *in vivo* or *in vitro* laboratory models. It is intended to facilitate the comprehensive and transparent documentation of an omics study including the experimental design, sample processing procedures, data collection, data normalisation and downstream computational analyses, the results of which could be used in regulatory decision-making contexts.

The OORF addresses the needs of two main types of end-users: regulators and researchers. The information captured by the OORF can be used by regulators in assessing the quality of data generated in an omics study and evaluating its suitability for use in regulatory decision-making. In addition, the information captured by the OORF provides researchers with the technical details needed to reproduce either the experimental or data analytical phase of a study, or both. The information in the OORF should be of sufficient detail for end-users to assess critical aspects of the experiment in each of the aforementioned areas to support regulatory decision-making processes.

Importantly, the OORF is constructed around a modular structure that facilitates the updating of individual technologies and allows the development of additional modules for new technologies. Each omics study should be reported using the following four types of modules: Study Summary Reporting Module (SSRM), Toxicology Experiment Reporting Module (TERM), Data Acquisition & Processing Reporting Module(s) (DAPRM), and Data Analysis Reporting Module(s) (DARM), which are structured as shown in Figure 1.

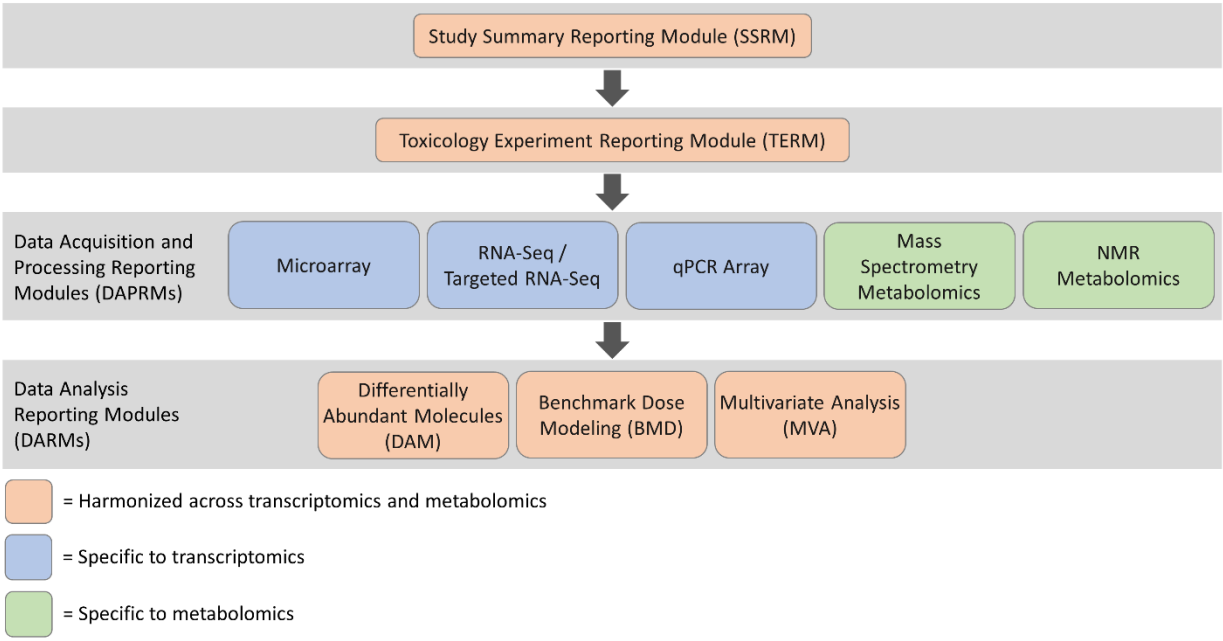

Figure 1. Modular structure of the OECD Omics Reporting Framework (OORF). To complete the OORF, scientists select the reporting modules that are relevant to their study and report the information that would be required by an end-user to fully comprehend and replicate the analyses. The four types of modules are: Study Summary Reporting Module (SSRM) to provide a high-level overview of the whole study; Toxicology Experiment Reporting Module (TERM) describing the *in vivo* or *in vitro* toxicology study; Data Acquisition and Processing Reporting Modules (DAPRM) describing the omics assays, data acquisition and processing; and Data Analysis Reporting Modules (DARM) describing the statistical analysis of the omics data. Orange modules are harmonised across omic technology types (e.g. transcriptomics, metabolomics, etc.). Blue modules are specific to transcriptomics, and green modules are specific to metabolomics. Additional DAPRMs and DARMs can be developed as needed to address new technologies and analytical methods of interest.

To report an omics study, scientists select the relevant reporting modules (and hence reporting elements), minimally consisting of an SSRM, TERM, one DAPRM and one DARM.

- **Study Summary Reporting Module (SSRM)**

  This module describes a subset of mandatory reporting elements in order to provide a high-level overview of a regulatory toxicology experiment involving an omics technology. There is one SSRM per study.

- **Toxicology Experiment Reporting Module (TERM)**

  This module serves to capture and report the key descriptors of the *in vivo* or *in vitro* toxicology study from which samples are derived for the omics analysis. There is one TERM per study.

- **Data Acquisition & Processing Reporting Modules (DAPRMs)**

  These modules serve to capture and report descriptions of the omics assays, data acquisition and associated data processing prior to statistical analysis (see Figure 1). There

is one or more DAPRM per study (a minimum of one is mandatory) dependent upon the number of omics assays applied to the samples from the study.

- **Data Analysis Reporting Modules (DARMs)**

  These modules serve to capture and report descriptions of the statistical analysis that is undertaken in the omics study, for example for the purposes of discovering differentially abundant molecules (e.g. differentially expressed genes in a transcriptomic study). The modules were designed to accommodate various types of omics data and can thus be applied to multi-omic data sets. There is one or more DARM per study (a minimum of one is mandatory) dependent upon the type(s) of data analysis applied to the datasets.

The reporting fields included in this guidance were determined by the subject matter experts participating in the development of each module. The reporting fields are annotated as either "recommended" or "optional" with the former classification viewed as being relatively more important for promoting transparency and repeatability of an omics-based toxicology study. Although only select reporting elements are noted as "recommended", it should be understood that a more comprehensive submission will enhance confidence and potential use in regulatory applications.

Also note that outputs from an 'upstream' reporting module can potentially be used an input for a 'downstream' reporting module. An example is a normalized and filtered gene counts output from the RNA-Seq DAPRM as the input for the differentially abundant molecule (DAM) DARM.

This narrative Guidance Document should be used in parallel with the minimal reporting guidelines presented in tabular spreadsheet format, one per module, to facilitate ease of reporting.

For some reporting fields it is appropriate for the OORF user to provide files (or links to files) that are relevant for understanding and/or reproducing an omics study. In the tabular reporting spreadsheets, these are referred to as 'data objects' - i.e. any machine-readable input, output or metadata file. Examples include tables of omics sample identifiers and associated metadata essential for analysis or interpretation of the omics dataset, lists of meaningful contrasts, omics platform annotations, etc. A file manifest tab is included in the tabular reporting spreadsheets, the purpose of which is to list all data objects relevant to a study.

The key principles of use of the OORF are summarized in Box 1.

Box 1: Principles of use of the OORF

The purpose of the OORF is to provide a framework for the standardisation of reporting of omics data generation and analysis for toxicology experiments. It serves to ensure that all the information necessary to understand and evaluate the quality of an omics experiment and its analysis are available.

The OORF:

- Does **NOT** stipulate the methods of data generation or analysis to be used in a regulatory toxicology study;
- Is a guidance (a set of recommendations) on reporting omics information and is **NOT** an OECD Test Guideline;
- Describes the technical components of an omics experiment that should be reported to achieve transparency and repeatability;
- Is modular and intended to be used in this way; this provides flexibility to integrate with other reporting frameworks as well as to be extended to cover additional technologies or data analysis methods;
- Can be used in a narrative reporting style if desired; the tabular reporting format is strongly recommended for the DAPRMs and DARMs;
- Should be used to report per study – i.e., a single regulatory study, whether it be on one or more test items, using one or more omics technologies, should be entered in a single OORF.

References

Brazma, A., Hingamp, P., Quackenbush, J., Sherlock, G., Spellman, P., Stoeckert, C., . . . Vingron, M. (2001). Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. Nat Genet, 29(4), 365-371. doi:10.1038/ng1201-365

Bridges, J., Sauer, U. G., Buesen, R., Deferme, L., Tollefsen, K. E., Tralau, T., . . . Pemberton, M. (2017). Framework for the quantitative weight-of-evidence analysis of 'omics data for regulatory purposes. Regul Toxicol Pharmacol, 91 Suppl 1, S46-S60. doi:10.1016/j.yrtph.2017.10.010

Buesen, R., Chorley, B. N., da Silva Lima, B., Daston, G., Deferme, L., Ebbels, T., . . . Poole, A. (2017). Applying 'omics technologies in chemicals risk assessment: Report of an ECETOC workshop. Regul Toxicol Pharmacol, 91 Suppl 1, S3-S13. doi:10.1016/j.yrtph.2017.09.002

Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., . . . Mortazavi, A. (2016). A survey of best practices for RNA-seq data analysis. Genome Biol, 17, 13. doi:10.1186/s13059-016-0881-8

Fostel, J. M., Burgoon, L., Zwickl, C., Lord, P., Corton, J. C., Bushel, P. R., . . . Tennant, R. (2007). Toward a checklist for exchange and interpretation of data from a toxicology study. Toxicol Sci, 99(1), 26-34. doi:10.1093/toxsci/kfm090

Gant, T. W., Sauer, U. G., Zhang, S. D., Chorley, B. N., Hackermuller, J., Perdichizzi, S., . . . Poole, A. (2017). A generic Transcriptomics Reporting Framework (TRF) for

'omics data processing and analysis. Regul Toxicol Pharmacol, 91 Suppl 1, S36-S45. doi:10.1016/j.yrtph.2017.11.001

Kauffmann, H. M., Kamp, H., Fuchs, R., Chorley, B. N., Deferme, L., Ebbels, T., . . . van Ravenzwaay, B. (2017). Framework for the quality assurance of 'omics technologies considering GLP requirements. Regul Toxicol Pharmacol, 91 Suppl 1, S27-S35. doi:10.1016/j.yrtph.2017.10.007

Krewski D, Andersen ME, Tyshenko MG, Krishnan K, Hartung T, Boekelheide K, Wambaugh JF, Jones D, Whelan M, Thomas R, Yauk C, Barton-Maclaren T, Cote I. (2019) Toxicity testing in the 21st century: progress in the past decade and future perspectives. Arch Toxicol. 2020 Jan;94(1):1-58. doi: 10.1007/s00204-019-02613-4.

McConnell, E. R., Bell, S. M., Cote, I., Wang, R. L., Perkins, E. J., Garcia-Reyero, N., . . . Burgoon, L. D. (2014). Systematic Omics Analysis Review (SOAR) tool to support risk assessment. PLoS One, 9(12), e110379. doi:10.1371/journal.pone.0110379

Parkinson, H., Sarkans, U., Shojatalab, M., Abeygunawardena, N., Contrino, S., Coulson, R., . . . Brazma, A. (2005). ArrayExpress--a public repository for microarray gene expression data at the EBI. Nucleic Acids Res, 33(Database issue), D553-555. doi:10.1093/nar/gki056

Sauer, U. G., Deferme, L., Gribaldo, L., Hackermuller, J., Tralau, T., van Ravenzwaay, B., . . . Gant, T. W. (2017). The challenge of the application of 'omics technologies in chemicals risk assessment: Background and outlook. Regul Toxicol Pharmacol, 91 Suppl 1, S14-S26. doi:10.1016/j.yrtph.2017.09.020

Schneider, K., Schwarz, M., Burkholder, I., Kopp-Schneider, A., Edler, L., Kinsner-Ovaskainen, A., . . . Hoffmann, S. (2009). "ToxRTool", a new tool to assess the reliability of toxicological data. Toxicol Lett, 189(2), 138-144. doi:10.1016/j.toxlet.2009.05.013

Segal, D., Makris, L., Kraft, A., Bale, A., Fox, J., Gilbert, M., … Crofton, K. (2015). Evaluation of the ToxRTool's ability to rate the reliability of toxicological data for human health hazard assessment. Regul Toxicol Pharmacol, 72(1), 94-101. doi:10.1016/j.yrtph.2015.03.005

Viant, M., Ebbels, T., Beger, R., Ekman, D., Epps, D., Kemp, H., … Weber, R. (2019). Use cases, best practice and reporting standards for metabolomics in regulatory toxicology. Nat Commun, 10(1): 3041. doi:10.1038/s41467-019-10900-y

Waters, M., Boorman, G., Bushel, P., Cunningham, M., Irwin, R., Merrick, … Tennant, R. (2003). Systems toxicology and the Chemical Effects in Biological Systems (CEBS) knowledge base. EHP Toxicogenomics, 111(1T):15-28.

1. Study Summary Reporting Module (SSRM)

This module describes a subset of mandatory reporting elements in order to provide a high-level summary of the application of omics to a regulatory toxicology study. With the exception of the reporting element in Section 1 - *Study identifiers*, see below, each section is derived from reporting elements described elsewhere within this Guidance Document. Specifically, this includes summarising the TERM, any Data Acquisition & Processing Reporting Module(s) (DAPRMs) used and any Data Analysis Reporting Module(s) (DARMs) used.

1.1. Study identifiers

REPORT:

1.1.1. Abstract

Provide a narrative abstract describing the toxicology study. Identify the test item(s) and include details of the experimental models, general design (e.g. dose-response, time series, etc.), omic technology used and omics data analysis that were performed. Also provide a brief description of the results of regulatory interest.

1.1.2. Study Identifier

Provide a unique identifier for the study.

1.1.3. Standardised toxicology dataset

If applicable, provide a link to a standardised toxicology dataset associated with this study. An example is linkage to an OECD test dataset (e.g. IUCLID ESR).

1.1.4. Complete omics dataset

If applicable, provide a link to the complete omics dataset associated with this study. An example is linkage to a MetaboLights, ArrayExpress or NCBI GEO dataset accession number.

1.2. Study Rationale

REPORT:

1.2.1. Background Information

Provide necessary background information for the end user to understand the rationale for why the study was undertaken, including the regulatory question(s).

1.2.2. Objectives

Clearly define the objectives of the omics study toward informing the regulatory question. Describe whether the omic study results are intended to be interpreted in isolation or in combination with results from other (non-omics) studies.

1.2.3. Test Item

a. Test item name

b. Test item CAS (if number exists)

c. Test item SMILES

d. Test item IUPAC name

e. Test item additional information

f. Test item sources of identifiers

1.2.4. Test System Characteristics (In Vivo)

a. Species

b. Strain

c. Sex

d. Age

1.2.5. Test System Characteristics (In Vitro)

a. Cell type

b. Species of origin

1.2.6. Study Design

a. Dose levels (including units)

b. Dose intervals (including units)

c. Number of biological replicates per treatment condition

d. Timetable – sample collection

1.2.7. Treatment Conditions

a. Route of administration

b. Identity of vehicle

1.2.8. Sample Collection

a. Type of biological sample for omics

Examples include cells, cell media, etc. for in vitro and biofluid, cells, tissues, organ, organism, etc. for in vivo.

b. Sample preservation method

1.3. Platform-Specific Data Acquisition and Processing Reporting Module (DAPRM)

In this section, specify the Data Acquisition and Processing Reporting Module(s) used to document the omics experiment. Also provide brief narrative descriptions of the omics technology used in the experiment, the rationale for using that technology, the sample preparation method and the use of quality assurance and quality control (QA/QC) methods used during data acquisition or processing.

REPORT:

1.3.1. DAPRM Module(s)

1.3.2. Omics technology, manufacturer, and model

1.3.3. Description and rationale for data acquisition approach

1.3.4. Sample preparation method

1.3.5. Demonstration of quality of analysis (i.e. QA/QC used)

1.4. Data Analysis Reporting Module (DARM)

In this section, specify the Data Analysis Reporting Module(s) used to document the analysis of omics data that were performed. Also provide a brief narrative description and accompanying rationale for the data analysis approach(es) that were used.

REPORT:

1.4.1. DARM Module(s)

1.4.2. Description and rationale for data analysis approach

.

2. Toxicology Experiment Reporting Module (TERM)

The TERM serves to capture and report the key descriptors of the *in vivo* or *in vitro* toxicology study from which samples are derived for omics analyses. One TERM should be reported per study, irrespective of whether one or more omics approaches are applied or whether one or more types of data analysis are performed.

2.1. Study Rationale

A clear and concise report of the study rationale is necessary to understand the suitability of the experimental design for the regulatory question being addressed, including the selection of experimental model, sex, target tissue, dosing regimen, etc. These fundamental aspects of the experimental design are clearly dependent on the study rationale.

Omics has increasingly been applied to chemical toxicity and disease studies, revealing new insights into mode(s) and mechanism(s) of action, disease markers, and toxicity signatures in human and environmental health. A variety of applications to support decision-making in a regulatory context have been demonstrated, but to date have not been widely implemented in the regulatory community (van Ravenzwaay et al., 2016; ECETOC, 2010a, 2010b, 2013). Recommended applications include: 1) discovery of mode of action and key events; 2) chemical grouping and read-across; 3) supporting weight of evidence approaches to identify hazard and risk; 4) tiered assessment screens; 5) cross-species extrapolation of toxicity pathways; and 6) deriving points of departure (Cote et al., 2016; Krewski et al., 2020; Thomas et al., 2013; Viant et al., 2019). Not surprisingly, the value of the results collected from studies conducted for these and other test applications will depend largely on the appropriate use of experimental designs and analytical approaches.

Experimental designs for each of the applications described above will differ. For example, if the identification of differentially abundant molecules (i.e. statistical testing) is required for a mode of action analysis, appropriate sample sizes per experimental group are required. In contrast, establishing similarities in omic profiles to support chemical groupings for read-across through unsupervised clustering approaches may be done with a smaller number of biological replicates, but may require either the availability or production of a database of omic profiles against which comparisons can be made. The use of benchmark-dose analysis to identify a point of departure benefits from a larger number of dose groups than typically used in toxicological studies but this can be offset by smaller sample sizes per group. In addition, the purpose of the study may govern the analytical platform used (e.g. use of whole transcriptome versus targeted transcriptomic approaches; use of mass spectrometry for the detection of metabolites at low concentrations), which will impact the choices made for downstream analyses.

Overall, a clear rationale describing why an omics study was undertaken is required to assess the suitability of the experimental design and its intended use for regulatory decision making. This

section is intended to be a narrative description of the reasoning for the study and its design. Details of the study parameters are to be provided in later sections.

REPORT:

2.1.1. Background information

Provide necessary background information for the end-user to understand the rationale for why the study was undertaken, including the regulatory question(s).

2.1.2. Objectives

Clearly define the objectives of the study toward informing the regulatory question. Describe whether the study results are intended to be interpreted in isolation or in combination with results from other studies.

2.1.3. Test guideline compliance

If appropriate, please refer to which OECD Test Guidelines have been followed in the performance of the method.

2.1.4. Mechanistic understanding

Briefly describe any prior toxicological, mode of action, or mechanistic information that is useful to understanding the study rationale (e.g. established mechanism of action and its relationship to the toxicological effect of interest).

2.1.5. Model selection

a. For *in vivo* studies, briefly explain how and why the selected animal species and model being used can address the scientific objectives and, where appropriate, the study's relevance to human biology. Provide a rationale for tissue or organ selection for the study.
b. Provide a rationale for the species and strain used.
c. For *in vitro* studies, briefly describe the biological relevance of the test system used in relation to the tissue/organ/species of interest.

2.1.6. Dose / concentration level and interval selection

Provide a brief rationale for the selection of the employed dose (*in vivo*) or test concentration (*in vitro*) levels and intervals. For example, selection for *in vivo* experimentation may be based on known toxicological effects or molecular changes documented for the test item identified in prior studies, allowing for “read-across” between omics data generated and other in-life findings, or clinical pathological changes and pathological observations. Note that dose interval information should include the time of exposure during the day. Similarly, in the case of *in vitro* experiments, if a relationship is sought to *in vivo* exposures, test item concentrations may be chosen using a quantitative *in vitro*/*in vivo* extrapolation (IVIVE) rationale; a rationale for the top concentration selected should be included.

2.1.7. Route of administration

Where relevant, provide a rationale for the choice of route of administration, referring to objectives of the study, potential route of human exposure, the physical and chemical characteristics of the test item and the relevance for the evaluated endpoint.

2.1.8. Time point selection

Provide a brief rationale for the exposure durations and sampling time points. Transcriptome, proteome, and metabolome profiles are dependent on both the duration of exposure and the time of sample collection. Responses reflect cellular adaptation over time, with early time points reflecting molecular initiating and early key events, and later time points reflecting pathological changes or adaptation. For *in vivo* studies, the interval between the final dosing and sample procurement should be specified. The same is true for sampling-time post-treatment for *in vitro* studies.

2.1.9. Samples and replicates

Provide a clear rationale for the choice of:

a. Biological replicate number, based on the scientific question posed and statistical power calculations predicting adequate coverage of biological variability.
b. Number of technical and analytical replicates, based on accepted and/or published standards for the assay and compliance with statistical power calculations.

2.1.10. Limitations

To facilitate regulatory evaluation, when appropriate, indicate the study limitations that could affect the outcome or interpretation of the results. These can include technical or mechanistic limitations in relation to known modes of action. For example, if the experiment includes poorly described test systems that would be a source of uncontrollable variability (e.g. a limitation of cell systems may be a lack of information about metabolic capacity), such information should be available to the evaluator. Likewise, some of the test items used might have physicochemical properties (lipophilicity, volatility, etc.) that might lead to a cell exposure that is different from the expected exposure (through interaction with plastic or proteins in culture plates and medium) or that produce large and confounding signals in omics outputs (e.g. NMR or MS spectra). In the case of *in vivo* studies, discussion of limitations should include any potential source of bias of the animal model or imprecision associated with the result.

2.2. Test and Control Items

According to the *OECD Series on Principles of Good Laboratory Practice and Compliance Monitoring*, a test item is defined as the subject of a study, and is also associated with “test compound”, “test substance”, “test article”, or other similar terms to describe the item being tested (OECD 2018a). Studies submitted for analysis to regulatory agencies should in the spirit of good laboratory practice (GLP) report test item transportation, receipt, identification, labelling (see

Section 2.7 - *Sample Identification Codes*), sampling, handling, storage and characterisation (OECD 2018b). Information regarding the test item characterisation is needed to inform potential route of exposure, as well as physicochemical properties that might influence the study (i.e. solubility, volatility, etc.).

Regulatory scientists must have all test items, vehicle, and control identification and characterisation information to accurately interpret omic study results. The following information should be reported for all test items: A) test substance, B) vehicle, and C) controls, including: test item name, mixture formulation composition, preparation of test item, physicochemical properties, chemical stability (OECD 2018c), commercial source and substance-specific identifiers. Additional information for nanomaterial test items should also be provided according to the 2016 OECD Workshop Report on *Physical-Chemical Parameters: Measurements and Methods Relevant for the Regulation of Nanomaterials* (OECD 2016a).

REPORT:

2.2.1. Test item name

2.2.2. If test item is a mixture, report formulation composition

- a. Identify substances that make up the mixture
- b. Relative proportions of substances (if known)

2.2.3. Preparation of test item (composition)

- a. Concentration of test item
- b. Concentration of diluent(s)
- c. Identification of impurities

2.2.4. Physicochemical properties

- a. Appearance/physical state/colour
- b. Molecular weight
- c. Melting point/freezing point
- d. Boiling point
- e. pH
- f. Viscosity
- g. Density
- h. Vapour pressure
- i. Partition coefficient (octanol/water)
- j. Water solubility
- k. Fat solubility of solid and liquid substances
- l. Particle size distribution/fibre length and diameter distribution (if applicable)

m. Additional physicochemical information (i.e. agglomeration, porosity, etc.) (if applicable)

2.2.5. Chemical stability

a. Stability in organic solvents and identity of relevant degradation products

b. Stability: thermal, sunlight, metals (if relevant)

c. Stability: dissociation constant (if relevant)

2.2.6. Commercial source

a. Vendor

b. Manufacture ID

c. Lot (Batch) number

d. Purity

e. Salt form

f. Expiration date

g. Storage conditions

2.2.7. Test item-specific identifiers

a. CAS

b. SMILES

c. IUPAC name

d. Additional information where available (e.g. InChIKey, InChI string, Distributed Structure-Searchable Toxicity (DSSTox) substance identifier (DTXSID), etc.)

e. Sources of identifiers

2.3. Test System Characteristics

Similar to traditional toxicity testing, it is critical that regulatory scientists applying omics data for risk assessment be provided with comprehensive information regarding the characteristics of the test system from which the data are derived. Test system refers to the biological system that is exposed to the test items to obtain experimental data. There are numerous examples in the literature demonstrating differential susceptibility of different species, strains within a species and sexes to chemical toxicity. Likewise, *in vitro* test systems derived from different species, tissues or even individuals vary in terms of relative sensitivity to toxicant exposure. The end-user must be equipped with detailed and accurate information regarding the test species or *in vitro* test system used to generate the data in order to critically evaluate the results and accurately compare the results across studies and data types.

With respect to *in vivo* toxicology studies, researchers should include relevant taxonomic information (i.e. species and strain), sex, age (at onset of dosing and at study termination) and commercial source of all animals included in a study. If determination of sex was not included in

the study design (such as in the case of some types of alternative species studies), or pooled samples from multiple animals were examined, then this should be explicitly described by the researcher. Researchers should also include detailed information on the housing conditions for all animals in a study including number of animals housed per cage, type of bedding, type of food, type of water provided, food and water accessibility (i.e. ad libitum or defined quantities), light/dark cycle, relative humidity, and other housing conditions the researcher may deem relevant for study interpretation. In general, information following the ARRIVE guidelines should be included (Kilkenny, Browne et al. 2010).

With respect to *in vitro* toxicology studies, researchers should include relevant information on culture type including species, strain (if applicable), sex of the organism, and organ or tissue from which the cells were derived. Researchers should include detailed information on culture conditions used to conduct the study as applicable, including complete media formulations, culturing vessel, growth substrate, passage number, donor lot, source (including commercial vendor or academic source), incubator conditions and proof of cell line authentication if available (OECD 2018d). In studies using complex, multicellular culture models (e.g. 3D cell models, organoids, organ-on-chip, etc.), the researchers should report what types of cells the cultures are expected to contain, cite relevant literature characterising the model system and describe any other relevant characteristic that might not be listed here.

REPORT:

2.3.1. General characteristics of the test system or subject (*In Vivo*):

a. Animal species
b. Strain
c. Sex
d. Age during study
e. Developmental stage
f. Individual weights/lengths at start
g. Supplier
h. Any interventions that were conducted before or during the experiment
i. Quality criteria before use
j. Health status and acclimation prior to study start
k. Randomisation of animals to groups
1. Other

2.3.2. General characteristics of the test system or subject (*In Vitro*):

a. Cell type (cell line or primary cells, tumour cells, etc…)
b. Origin (animal/organ/tissue)
c. Cell passage number (of frozen stock and passage number at the start of the treatment).
d. Differentiation stage
e. Absence of mycoplasma

f. Metabolic competence
g. Supplier
h. Quality criteria before use
i. Other

2.3.3. Housing and husbandry (*In Vivo*):

a. Type of facility (e.g. specific pathogen free [SPF])
b. Type of cage or housing
c. Bedding material
d. Number of cage companions
e. Tank shape (for fish) and its material
f. Breeding programme
g. Light/dark cycle
h. Temperature
i. Quality of water (e.g. for aquatic toxicity tests)
j, Type of food (ingredients in food as detailed as possible, including vendor and lot number)
k. Access to food and water
l. Environmental enrichment
m. Methods for fertilisation/collection of eggs, if applicable
n. Other

2.3.4. Cell culture conditions (*In Vitro*):

a. Incubation characteristics
    i. Temperature
    ii. CO_2/O_2 conditions
    iii. Humidity
    iv. Other
b. State of the cells before use
    i. Viability (including test used)
    ii. Quality control
    iii. Morphology
    iv. Recommended confluency of use
    v. Other
c. Culture media (in case of multiple)
    i. Use of serum (with details such as species of origin, age, sex, etc.)

ii. Use of antibiotics

d. Growth support substrates (if applicable)

i. Use of feeder cells

ii. Use of matrixes or scaffolds

2.4. Study Design

Study designs are based on the assessment of all available information for factors that have the potential to influence study results. A detailed description of all elements and parameters included in the study design will increase transparency and confidence in the data, and as a result, will have greater utility in regulatory assessment. The following guidance for reporting study design using *in vitro* and *in vivo* systems to generate omics data is based on previously published OECD Guidance Documents for the respective areas (OECD 2014; OECD 2017). Much of the reporting guidance is based on the application of good laboratory practice (GLP) principles and good *in vitro* method practice (GIVIMP) according to an internationally accepted definition, ensuring mutual acceptance of data (MAD) across OECD countries (OECD 1998; OECD 2018d). Although GLP-compliant study protocols have been developed for most areas of regulatory risk assessment, there are no absolute requirements for their application in the generation of omics data submitted for regulatory purposes (EPA 2009; FDA 2015). In general, study designs should provide experimental detail, standard operating procedure (SOP) information, and statistical design information in equivalence to the sentiment of GLP study design, but not necessarily requiring all aspects of traceability etc., which are generally required for GLP auditing. However, a study running under GLP is preferred, as it is the standard for regulatory studies following OECD Test Guidelines. In addition to these recommendations, those published recently by the MERIT project (Viant et al., 2019) should be followed to the fullest extent possible for metabolomic studies. Finally, the report should detail any 3R (reduction, refinement and replacement) arguments underlying the study design; e.g. choice between *in vivo* and *in vitro* tests systems, statistical powering (see also Section 2.1 - *Study Rationale*, and Section 2.3 - *Test System Characteristics*) (EC 2010).

REPORT:

2.4.1. 3Rs considerations

a. Briefly describe how the study addresses the 3R principles.

2.4.2. Dose selection

a. Dose levels: Indicate each of the dose levels/concentrations used in the study (including units) and identify the matched vehicle/solvent controls to be used.

b. Dose interval: Indicate the dose interval (e.g. acute single or chronic dosing) and exposure duration (e.g. hours, days, weeks, or months) used in the study, including units.

2.4.3. Description of the test method instruments, equipment and reagents

Provide a full description of the instruments and equipment used for the collection and processing of samples for the analysis, with details concerning:

a. Commercial source: Provide the suppliers/manufacturers of instrumentation, other laboratory equipment and reagents relevant to the study.

b. Manufacturer's instrument model identification.

c. Manufacturer's reagent and kit information.

d. Any special safety/handling requirements.

2.4.4. Types of treatments

The study design report must include a description of the type(s) of treatment including:

a. Controls: Defined as experimental samples derived from animals or cells treated with their respective dose formulation, in the absence of test item. All control types should be reported (i.e. positive control, negative control, vehicle control, blank, etc.) following the criteria described in Section 2 - *Test and Control Items*.

b. Pre-treatments: Where necessary, a description of pre-treatments involving metabolic activation, for example of specific cytochromes(s) P450, should be provided.

c. Acclimation: A brief description of acclimation should be reported to include the length of the acclimation period, health status of the test system, and environmental conditions. Relevant quarantine conditions should be described, where applicable.

d. Types of replicates: The report should clearly define the number of biological replicates (samples derived from individual animals or cell samples), utilised for each dose level. This should be clearly delineable from any technical replicates generated, (the sample processed more than once) and/or analytical replicates (the same sample analysed more than once) (Blainey, Krzywinski et al. 2014). If relevant, clearly define what constitutes a biological replicate for the in vitro study being reported.

2.4.5. Numbers of animals/samples per treatment

a. Biological replicates: Describe the number of biological replicates in each treatment condition.

b. Technical/analytical replicates: Describe the number of technical and analytical replicates.

2.4.6. Statistical design

a. Exposure design: Describe the various statistical approaches (Chow 2014) used in the study design to prevent exposure bias (e.g. randomised block, latin square, incomplete block, etc.)

b. Sampling schemes: Describe sampling schemes used to prevent sample collection bias, and to ensure proper sample labelling post-collection, using methods (e.g. sequential, stratified, systematic, randomised, ranked set, etc.)

c. Sample blinding: Describe sample blinding approaches following the sample collection, used to prevent experimental bias in downstream sample processing. This should include a unique identification (as explained in Section 2.7 - *Sample Identification Codes*) that does not represent the sample or treatment type.

2.4.7. Observations/examinations during treatment

Where appropriate, include details of other experimental observations used in the experiments generating the samples, including:

a. In-life cage-side or clinical observations, feed consumption, water consumption and body weight in for *in vivo* experiments

b. Toxicokinetics

c. Histopathology and organ weight

d. Clinical pathology in *in vivo* experiments

e. Reason animals were removed from the study

f. Cytological analyses in *in vivo* experiments

g. Cytobiological examinations in *in vitro* experiments (such as cell morphology and cytotoxicity testing)

h. Other biochemical and/or molecular biological analyses

2.5. Treatment Conditions

In the context of the Adverse Outcome Pathway (AOP) framework (http://www.oecd.org/chemicalsafety/testing/projects-adverse-outcome-pathways.htm), reasoned, well-defined exposures through relevant routes of administration result in changes in transcript, protein and metabolite levels and pathways representative of key events possibly related to the final adverse outcome. The objective of the omic study informs the selection of the treatment conditions, which, in turn, impact the final outcome. Omic analyses may produce different results depending on the route of administration, the dose levels and the time and schedule of the exposure. Thus, a thorough description of treatment conditions is necessary for interpreting omic study results. For both *in vivo* and *in vitro* studies, it is understood that the amount of chemical that reaches the target will affect the outcome. If available, provide information on tissue dosimetry either measured or modelled.

The present guidance for reporting treatment conditions in omic studies is based mainly on previously published OECD documents (OECD 2014; OECD 2017; OECD 2018d; OECD 2018a) and standard harmonised templates for reporting of information derived from *in vivo* or *in vitro* studies for the risk assessment of chemicals (OECD 2016b; OECD 2018e). In addition, the ARRIVE guidelines for reporting animal research were taken into account (Kilkenny, Browne et

al. 2010). Following GLP-like requirements for accurate, comprehensive reporting will help in the evaluation of the relevance of data deriving from omic studies.

REPORT:

2.5.1. Route of administration (*In vivo*): For *in vivo* studies, indicate the selected route of administration for the test item. Examples are oral (e.g. gavage or diet), dermal, inhalation, parenteral, implantation, etc. If applicable, for implantation experiments, state the exposure regime (e.g. static, semi-static, flow through). Also state other information that may be required to understand the route of exposure for the test item.

2.5.2. Route of administration (*In vitro*): For *in vitr*o studies, describe how the test item was administered to the *in vitro* system. Examples are direct addition of the test item to cultures, substitution of culture medium, exposure at air-liquid interface, etc. Also state other information that may be required to understand how the test time item was administered to the *in vitro* system.

2.5.3. Housing condition modifications (*In vivo*): Describe any modifications of the standard culture/housing conditions occurring before and during the test item exposure (refer to 'Housing, husbandry, and culture conditions' in Section 2.3 - *Test System Characteristics*). Examples include fasting period, use of anaesthesia and/or analgesia or other modifications to standard housing conditions that may have occurred before or during exposure to test items.

2.5.4. Culture condition modifications (*In vitro*): Describe any modification of the culture conditions occurring before and during the test item exposure (refer to 'Cell Culture Conditions' in Section 2.3 - *Test System Characteristics*). Examples include switching to serum-free or serum-depleted medium, use of items for limiting media evaporation or other modifications to standard culture conditions that may have occurred before or during exposure to test items.

2.5.5. Test item preparation

Describe all the steps leading to the test item preparation for administration to the test system and any modifications to the original procedure, e.g. problems with chemical solubility:

a. Dilution in a vehicle

b. Preparation steps (warming, grinding, etc.)

c. Separation steps (centrifugation, decantation, filtration, etc.)

d. Extraction steps (for specific test items, such as medical devices)

e. Storage conditions

f. Stability during storage

g. Expiration date

h. Whether nominal or measured concentrations were used, if applicable

i. Analytic controls on measured concentrations, if applicable

j. Dosing solution homogeneity

2.5.6. Test Item Stability/Reactivity

Describe the procedures used to assess:

a. Stability of the test substance under test conditions

b. Solubility and stability in the solvent/vehicle

c. Reactivity of the test substance with the solvent/vehicle or the cell culture medium, if applicable

d. Impact of separation and extraction steps on integrity, homogeneity, concentration and stability of the prepared test item

2.5.7. Test Item Preparation for *In Vivo* Studies

For *in vivo* studies, provide details about test item preparation:

a. Procedures for test substance formulation/diet preparation

b. Procedures for generation of test atmosphere and chamber description

c. Achieved concentration

d. Stability and homogeneity of the preparation

e. Test item intake for dietary or drinking water studies. Conversion from diet/drinking water or inhalation test substance concentration (ppm) to the actual dose (mg/kg body weight/day), if applicable.

2.5.8. Vehicle description and delivery volume

If the test chemical is dissolved or suspended in a suitable vehicle, provide all the relevant information on:

a. Identity of the vehicle

b. The delivery volume

c. The final concentration of the vehicle in the test item preparation.

d. For *in vivo* studies, The maximum volume of liquid that has been administered by gavage or injection. The use of volume exceeding the suggested volume should be justified. In terms of reporting, refer to Section 2.2 - *Test and Control Items*).

2.5.9. Exposure schedule

a. Frequency of test item administration (e.g. once daily).

b. Time of day of dosing (*in vivo*)

c. Time of dosing after seeding (*in vitro*)

d. Indicate the recovery period (in days, weeks, months, if any) after the last exposure to the test substance and sample collection.

2.5.10. Housing/culture condition deviations during treatment

Describe any undesired deviations from the housing/culture conditions established in the study plan that occurred during the treatment and/or the observation time and their possible impact on the study results.

a. Temperature

b. Humidity

c. CO_2 %

d. pH

e. Availability and quality of nutrients

f. Other

2.6. Study Exit & Sample Collection

Omics can be conducted using samples from animal or *in vitro* studies to produce a "snapshot" of transcript, protein or metabolite levels and enable analysis of perturbations in biological pathways and processes. Proper sample preparation is a key step in omics studies and adequate care must be taken to ensure sample fidelity. Several factors with regard to preparation for sampling need to be considered because of their potential for causing alterations in the transcriptome, proteome, and metabolome, which may confound biological interpretation of the data. This applies to both *in vivo* and *in vitro* studies, although the type of biological sample may affect the selection of the subsequent handling steps.

Animal studies should always be conducted in strict accordance with ethical principles and regulations. When terminating an animal study, euthanasia must be performed using appropriate techniques and equipment to ensure death is induced in a manner that is as painless and stress-free as possible. Consequently, anaesthetics are commonly used in these procedures. Sometimes, analgesic drugs are administered during and/or after surgical procedures in laboratory animals. Several studies have demonstrated that anesthesia and euthanasia may impact omics results (HK 2004; Overmyer, Thonusin et al. 2015; Nakatsu, Igarashi et al. 2017). In addition, for *in vitro* studies, the methods used for harvesting the samples may influence the data produced (Ramirez et al. 2018). A detailed report on the methodologies used for collecting and storing specimens will allow reviewers to appropriately evaluate the quality of the omic studies.

Additionally, biotic and abiotic information at the time of harvesting should be collected to allow for assessment of sample fidelity. This includes the conditions under which samples are stored until further processing. As molecules in cells are sensitive to environmental conditions, both harvesting and storage procedures should be reported as accurately as possible.

REPORT:

2.6.1. Type of biological sample collected (*in vitro*: single cell, cell culture, 2D or 3D culture, single or multi type cell culture; *in vivo*: biofluid, cells, tissue, organ, organism *in toto*)

2.6.2. Study Exit (*In vivo*) (if applicable)

a. Anaesthetic used: substance (e.g. isoflurane, ether), dosage, route of administration

b. Analgesic used: substance, dosage, route of administration

c. Method of euthanasia (e.g. carbon dioxide asphyxiation, exsanguination)

d. Phenotypic characteristics (e.g. body weight, organ weight)

e. Methods used for collection of biological sample(s) (e.g. dissection, isolation of tissues or organs)

2.6.3. Study Exit *(In vitro)* (if applicable)

a. Collection of biological material: method used (e.g. detergent), substance, concentration, duration

b. Cell density at time of harvesting

c. Growth phase/stage (e.g. cell cycle phase, if available)

d. Number of culture passage

e. Morphology

f. Methods used for collection of biological sample(s) (e.g. removal of media, wash (see below), quench (see below), scrape into sampling vial, etc.)

2.6.4. Sampling vial

a. Type of vial or tube

b. Chemicals within the sample tube (EDTA, heparin, etc.)

c. Chemicals added to preserve sample(s) (nitrogen, argon, etc.)

2.6.5. Washing

The primary purpose of washing a sample is to remove contamination. For example, prior to the extraction of adherent cells to study the intracellular metabolome, it is important to remove (via washing) metabolites present in the cell media. Washing is particularly important for an untargeted LC-MS assay as it is a sensitive analytical method.

a. Wash solvent(s)

b. Washing procedure (including temperature)

2.6.6. Quenching (i.e. procedure to stabilise sample)

The primary purpose of quenching is to preserve metabolite levels in the isolated sample as similarly to their levels in the living system.

a. Quench solvent

b. Quenching procedure (including temperature)

2.6.7. Pooling (or aliquoting) of samples

a. Describe any sample pooling procedures.

2.6.8. Sample storage and transport

Transport and storage conditions prior to sample extraction are important factors in the reliability of measurements. Storage temperature, time and the number of freeze-thaw cycles can all affect the stability of transcripts, proteins, and metabolites.

a. Post sample collection handling, prior to sample extraction

b. Storage temperature and duration

c. Transportation method (e.g. between experimental facilities)

d. Number of freeze-thaw cycles

2.6.9. Time Table

Detail the timetables used to perform the study protocol with respect to:

a. Treatments

b. Sample collections

c. Time since last dose administered

d. Time to sample extractions

2.7. Sample Identification Codes

Sample management is a critical component of regulatory and non-regulatory experimentation which should be carefully planned. To aid in laboratory organisation and management, a laboratory information management system (LIMS) can be used to consolidate laboratory tasks, such as: sample management, laboratory work-flows and protocols, documentation, management of laboratory stocks and solutions, and clinical data (List, Schmidt et al. 2014). Samples used in omics experiments should be given a unique identification code and information stored in a secure LIMS where available.

A standard operating procedure (SOP) should be established to ensure identification, tracking, unbiased testing and data collection records for each sample. The sample identification code generation should be produced in the spirit of good laboratory practice (GLP) in order to maintain

proper records of samples and their associated method of experimentation (OECD 1998). The code identification of each unique sample should be securely linked to test item information, experimental study number, and experimental metadata.

REPORT:

2.7.1. Laboratory information management system (LIMS)

If a LIMS software was used for information management, report the name of the software and the software version.

2.7.2. Method/Schema for Assigning Unique Sample Identifiers

Describe the method or schema used to assign unique identifiers to samples in the study. Examples include consecutive numbering, alphanumeric or (in the case of *in vitro* studies) a combination of plate identifier and well coordinate.

2.7.3. Metadata table:

Provide a 'data object' (e.g. file) containing the unique sample identifiers and any metadata fields required to analyse or interpret the study. Examples of metadata crucial for data analysis and interpretation include, but are not limited to, sample type (e.g. control or treatment group), species, sex, strain, cell type, dose level, exposure duration, etc. Other types of metadata that are not crucial for downstream data analysis interpretation may also be included in this table at the researcher's discretion.

2.8. Supporting Data Streams

Omics studies can be used to address different types of regulatory questions. To be able to fully appreciate an omics study and its resulting data, a clear and concise report is required. The framework described in this guidance ensures that all essential information is captured to allow for this detailed understanding.

However, there may be situations where even a higher level of detail is needed to allow for use of omic data for regulatory decision-making. Moreover, data may be re-used for other regulatory questions or for a similar regulatory question at a later point in time. To benefit optimally from the data generated, additional information should be reported to the extent possible. This information can range from OECD Test Guidelines for a particular animal study to methodological Standard Operating Procedures (SOP) to scientific publications in which analysis of (a subset of) the data has been described. Toxicity or cytotoxicity experiments necessary to establish the appropriate doses/concentrations can also be reported here.

2.9. References

Blainey, P., M. Krzywinski, et al. (2014). "Replication." Nature Methods **11**: 879.

Chow, S. C. (2014). "Adaptive clinical trial design." Annu Rev Med **65**: 405-415.

Cote, I., M. E. Andersen, et al. (2016). "The Next Generation of Risk Assessment Multi-Year Study-Highlights of Findings, Applications to Risk Assessment, and Future Directions." Environ Health Perspect **124**(11): 1671-1682.

EC. (2010). "Directive 2010/63/Eu Of The European Parliament An D Of The Council Of 22 September 2010 On The Protection Of Animals Used For Scientific Purposes." from https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32010L0063.

ECETOC Workshop Report No. 19 (2010a) “Omics in (Eco)toxicology: Case Studies and Risk Assessment”, 52 pages, ISSN 2078-7200-19.

ECETOC Technical Report No. 109 (2010b) “High information content technologies in support of read-across in chemical risk assessment”, 87 pages, ISSN-0773-8072-109.

ECETOC Workshop Report No. 25 (2013) “Omics and Risk Assessment Science”,70 pages, ISSN-2078-7200-25.

EPA. (2009). "An Approach to Using Toxicogenomic Data in U.S. EPA Human Health Risk Assessments: A Dibutyl Phthalate Case ", from https://goo.gl/4KcmUz.

FDA. (2015). "Critical Path Innovation Meetings. Guidance for Industry." from https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM417627.pdf.

Hamadeh, HK., C.A. Afshari. Toxicogenomics: Principles and Applications. New Jersey, John Wiley & Sons, Inc. Hoboken.

Kilkenny, C., W. J. Browne, et al. (2010). "Improving bioscience research reporting: the ARRIVE guidelines for reporting animal research." PLoS Biol **8**(6): 1000412.

Krewski D, Andersen ME, et al. (2020). “Toxicity testing in the 21st century: progress in the past decade and future perspectives.” Arch Toxicol. **94**(1):1-58.

List, M., S. Schmidt, et al. (2014). "Efficient sample tracking with OpenLabFramework." Sci Rep **4**: 4278.

Nakatsu, N., Y. Igarashi, et al. (2017). "Isoflurane is a suitable alternative to ether for anesthetizing rats prior to euthanasia for gene expression analysis." J Toxicol Sci **42**(4): 491-497.

OECD. (1998). "OECD Series On Principles Of Good Laboratory Practice And Compliance Monitoring Number 1." from http://www.oecd.org/chemicalsafety/testing/good-laboratory-practiceglp.htm.

OECD. (2014). "Guidance Document 116 on the Conduct and Design of Chronic Toxicity and Carcinogenicity Studies, Supporting Test Guidelines 451, 452 and 453: Second edition, OECD Series on Testing and Assessment, No. 116." from https://doi.org/10.1787/9789264221475-en.

OECD. (2016a). "OECD Environment, Health and Safety Publications Series on the Safety of Manufactured Nanomaterials Physical-Chemical Parameters: Measurements and Methods Relevent for the Retulation of Nanomaterials No. 63." Retrieved October 09, 2018, from http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?doclanguage=en&cote=env/jm/mono.

OECD. (2016b). "OECD Harmonised Template 201: Intermediate effects. (Version [1.11]-[April 2016]) ", from https://www.oecd.org/ehs/templates/harmonised-templates-intermediate-effects.htm.

OECD. (2017). "Guidance Document for Describing Non-Guideline In Vitro Test Methods, OECD Series on Testing and Assessment, No. 211." from https://doi.org/10.1787/9789264274730-en.

OECD. (2018a). "OECD Environment, Health and Safety Publications Series on Principles of Good Laboratory Practice (GLP) and Compliance Monitoring No. 19." Retrieved October 09, 2018, from http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2018)6&doclanguage=en.

OECD. (2018b) “OECD Guidelines for the Testing of Chemicals, Section 1”. Retrieved Oct. 25, 2022, from https://www.oecd-ilibrary.org/environment/oecd-guidelines-for-the-testing-of-chemicals-section-1-physical-chemical-properties_20745753

OECD. (2018c). "OECD Harmonised Templates 1 to 23-5 and 101 to 113: Physico-chemical properties." Retrieved October 09, 2018, from https://www.oecd.org/ehs/templates/harmonised-templates-physico-chemical-properties.htm.

OECD. (2018d). "Guidance Document on Good In Vitro Method Practices (GIVIMP). Series on Testing and Assessment No. 286. ENV/JM/MONO(2018)19." from http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=ENV/JM/MONO(2018)19&doclanguage=en.

OECD. (2018e). "OECD Harmonised Template 71: Genetic toxicity in vivo (Version [5.]-[August 2018]) ", from https://www.oecd.org/ehs/templates/harmonised-templates-health-effects.htm.

Overmyer, K. A., C. Thonusin, et al. (2015). "Impact of anesthesia and euthanasia on metabolomics of mammalian tissues: studies in a C57BL/6J mouse model." PLoS One **10**(2).

Ramirez, T., A. Strigun, et al. (2018). “Prediction of liver toxicity and mode of action using metabolomics in vitro in HepG2 cells.” Arch Toxicol **92**:893-906.

Thomas, R. S., M. A. Philbert, et al. (2013). "Incorporating new technologies into toxicity testing and risk assessment: moving from 21st century vision to a data-driven framework." Toxicol Sci **136**(1): 4-18.

van Ravenzwaay, B., Sperber S., et al. (2016). “Metabolomics as read-across tool: A case study with phenoxy herbicides.” Regulatory Toxicology and Pharmacology **81**:288-304.

Viant, M.R., Ebbels, T.M.D. et al. (2019). “Use cases, best practice and reporting standards for metabolomics in regulatory toxicology.” Nat Commun. **10**(1):3041.

WHO. (2009). "Handbook: Good Laboratory Practice (GLP). Quality Practices for Regulated Non-clinical Research and Development - 2nd Ed." Retrieved 10/15/2018, 2018, from http://apps.who.int/medicinedocs/en/d/Js19719en/.

3. Data Acquisition & Processing Reporting Modules

Data Acquisition and Processing Reporting Modules (DAPRMs) capture and report descriptions of the omics assays, data acquisition and associated data processing prior to statistical analysis. These modules are unique to each omics data type. Furthermore, the reporting described is generalisable and adaptable, because (1) it is not the intention of this Guidance Document to prescribe the types of assays that the regulatory toxicology community should use; (2) technologies will continue to evolve and we want the OORF to remain relevant; and (3) omics assays often do not fit well into predefined boxes.

This section of the OORF describes the following reporting modules:

- Section 3.1 - Microarray
- Section 3.2 - RNA-Seq
- Section 3.3 - qRT-PCR
- Section 3.4 - Mass spectrometry metabolomics
- Section 3.5 – NMR spectroscopy metabolomics

3.1. Microarray

Microarrays measure the abundance of a defined set of transcripts in a sample via labelling and hybridisation to an array of complementary probes attached on a solid surface. The capacity of microarrays to simultaneously detect tens of thousands of transcripts has led to important advances across biology, including the identification of genes that are differentially expressed between diseased and healthy tissues, pharmaco/toxicogenomics responses, and defining gene regulation in different species.

This module provides a reporting framework for describing a microarray technology, documenting transcriptomic study experimental design including platform-specific sample processing (such as addition of controls, labelling and hybridisation), details of raw data acquisition and format and how normalisation, data filtering and outlier identification and removal was conducted.

3.1.1. Technology

This section describes the information a scientist requires for determining analytical sensitivity, limits of detection, interference, and precision (reproducibility and repeatability) of the microarray technology used in a transcriptomic experiment.

Documentation of the identity and number of probes measured and detection calls for individual probes is essential for interpretation of a microarray-based transcriptomic study. Microarray probes vary significantly in their hybridisation properties, and arrays are limited to interrogating only those genes for which probes are designed. In addition, a potential limitation of microarray technology is background hybridisation that limits the accuracy of expression measurements, mainly for transcripts present in low abundance.

A “platform” defines the microarray design and requires documentation of the sequence identity tracking information for each feature on the microarray. Probes (or oligoprobes) are short DNA sequences complementary to a region of a specific transcript and are used to estimate transcript abundance through hybridisation.

Whichever platform is used, the underlying mapping of the probes to biological entities (i.e. transcripts/genes/proteins) must be annotated. While probe sequences don’t change, genome assemblies (e.g. chromosomal sequences) and annotation of biological entities are both subject to change over time. Given the iterative improvement of genome annotations, a certain microarray probe that mapped to gene *X* in one instance could be mapped to gene *Y* in another instance because gene *X* has been made obsolete by a genome annotation update, or its exon-intron structure has changed in light of new supporting evidence. Therefore, the accurate reporting of version information, both in terms of microarray platform and the reference genome used for biological interpretation, is essential for understanding the results of a microarray-based transcriptomic study.

Manufacturers are vital for supplying testing laboratories with reliable products and probe sequences/annotation. Ideally, users and manufacturers communicate so that substantial changes to the product are conveyed to users. Finally, the hardware and software packages that generate and process microarray data represent a wide assortment of data styles and formats. Therefore, information on hardware and software versions and configurations used to collect microarray data should be reported to facilitate assessment of data provenance and quality.

REPORT:

3.1.1.1. Type and version of the platform, manufacturer's name (e.g. Affymetrix U133 Plus 2.0 Array) and associated genome build (e.g., GenBank version – GRCh38)

3.1.1.2. The unique identifier (e.g. serial number) and manufacturer

3.1.1.3. Feature type (e.g. spotted oligonucleotide)

3.1.1.4. Feature annotation (e.g. probe IDs)

3.1.1.5. Purpose of feature(s) (e.g. target gene expression, quality control, etc.)

3.1.1.6. Composition of feature(s) (i.e. oligo sequence, ligated product sequence)

3.1.1.7. Control console operating system

3.1.1.8. Other relevant information

3.1.2. Transcriptomics Experimental Design

The microarray experimental design includes defining sample pooling (if applicable), batch processing (if applicable), the types and use of microarray controls, sample processing (how samples were labelled and hybridised to microarrays), and analytical processes applied to assess sample and hybridisation data quality.

3.1.2.1. RNA Processing

The principles of RNA extraction are similar across organisms (Chomczynski and Sacchi 1987; Chomczynski and Sacchi 2006). The key is to avoid incomplete RNA extraction, RNA degradation and introducing contaminants during sample collection, processing, or storage. There are a number of studies on standardising methodologies for biological sample collection and storage, and optimising procedures for RNA isolation and purification (Wilfinger, Mackey et al. 1997; Vomelova, Vanickova et al. 2009) to improve the quality and yield of the RNA. Moreover, procedures for depleting rRNA and genomic DNA contamination, and enhancing mRNA recovery, have been implemented to improve the performance of downstream transcriptomic applications (Bryant and Manning 1998; Zhao, He et al. 2014). The methodologies used for RNA sample collection, processing, quality assessment, and storage will have effects on the final research results and should be reported in detail.

Successful transcriptomic studies depend on accurate RNA quantification and quality control analysis. Electrophoretic methods have been applied to separate the RNA fragments according to size. RNA quality indices, such as ribosomal ratio and RNA integrity number (RIN), have been established to determine RNA quality. Successful RNA analysis also involves proper RNA quantification. Downstream applications rely on precise amounts of RNA to obtain good technical performance and allow reliable comparisons among sample groups. Thus, RNA quantification will directly affect the quality of data and interpretation of the final results. Several methods for RNA quantification have been routinely used, such as ultraviolet absorbance, fluorescent dye-based RNA quantification, and

Bioanalyzer/TapeStation readouts (Grillo and Margolis 1990). The procedures used to quantify and qualify RNA should be reported to enable an appropriate evaluation of the RNA extraction steps and suitability for use in downstream transcriptomic analyses. Quality thresholds used to define samples with high enough RNA integrity for transcriptional analysis should be defined.

REPORT:

a. Type of extracted RNA (e.g. total RNA, mRNA, miRNA)

b. Extraction and purification techniques

c. Procedures for mRNA enrichment (if applicable), or other enrichment procedures

d. Storage conditions

e. Quantification and Qualification of RNA

   i. Tool for RNA Assessment

   ii. RNA quality (e.g. A_{260}/A_{280} ratio, RIN for eukaryotic RNA, PERM number (Chung et al. 2016) for eukaryotic RNA extracted from formalin-fixed paraffin embedded tissues).

   iii. RNA yield (e.g. RNA/gr of tissue, RNA/10^6 cells)

   iv. Performance metrics

3.1.2.2. Sample Pooling Protocol (if applicable)

The design of the transcriptomics study also includes determining if RNA samples need to be pooled/processed together due to insufficient quantity required for performing the microarray experiment.

REPORT:

a. Whether any sample pooling has occurred (samples being combined into one hybridisation) – yes/no

b. Reason for pooling of samples

c. How samples were pooled – quantity and which samples were pooled?

d. Other relevant information

3.1.2.3. Batch processing of samples (if applicable)

When there are many samples within a study, the microarray experiment may have to be conducted in 2 or more batches. For example, a lab may have a capacity of performing a microarray experiment for 32 samples in one run (which can take 2-3 days to complete). If the study has 64 samples, then the researchers have to conduct the experiment in 2 batches. Assigning samples from different treatment groups (and/or different treatment duration, cell types, doses, etc.) to different batches needs careful consideration so as to include an equal number of samples

from each group in order to minimise batch effects and maximise biological differences. Please report the following parameters:

REPORT:

a. Number of batches

b. Number of samples used in each batch of the microarray experiment

c. How the samples were assigned to each batch (i.e. the criteria)

d. Method of sample assignment to different batches (for example, simple randomisation, randomised block design etc.)

3.1.2.4. Linear Amplification, Labelling, cDNA/cRNA Preparation

Prior to hybridisation to microarrays, RNA samples are converted to cDNA and processed through a variety of platform-specific labelling protocols. Each step of the protocol should be described in detail. When RNA quantity is limited, linear pre-amplification can boost the signal, and a label can be incorporated to permit downstream detection after hybridisation. Labelling efficiency can be checked before hybridisation. Before cDNA preparation, DNA is usually removed using DNase. The cDNA production uses random primers, oligo(dT) primers targeting poly(A) tails of mRNA, or specific primers targeting each RNA.

REPORT:

a. Labelling protocol

Specify whether the labelling protocol is manufacturer specific, modified or custom – describe the protocol in detail)

b. Labelling description

Specify whether the label has been added during the reverse transcription step or following amplification, whether the label is fluorescent (and the wavelength of the fluoroprobe) or radioactive, whether the labelling is direct or there is a pre-label and with a detection process after hybridisation. Specify if the hybridisation is two colour (two samples) or single (one sample).

c. Other relevant information

3.1.2.5. Hybridisation

Hybridisation is the process of complementary binding between a labelled target cRNA and an oligonucleotide probe on a microarray. While probe design and validation is a component of the microarray platform development process, other differences in microarray manufacture or variations in stringency of target cRNA binding to different microarrays will influence hybridisation outcome. In addition, systematic errors in microarray production can lead to variability in hybridisation (e.g. defective lot of microarrays). Conditions for microarray hybridisation should

be described as well as any experimental design features used to evaluate hybridisation performance.

Microarray platforms can be designed in a variety of ways. Multiple arrays may be present on a glass slide and samples can be labelled with one or two fluorescent dyes. Experimental designs in the latter case may include reference or dye-swap designs. Thus, it is critical to report which samples and dyes were labelled on specific arrays, including dates of hybridisation. This information is also essential for understanding whether the hybridisation design can introduce potential confounding variables (e.g. lack of randomised assignment of samples across batches, microarray slides, hybridisation dates, etc.). The association of samples and labels should be reported.

REPORT:

a. Hybridisation protocol: Provide the name of the protocol and state whether the protocol is manufacturer specific, modified or custom. Describe the protocol in detail.

b. Hybridisation information: In addition to the specific protocol used in the experiment, a full description of the design must be provided including the dates of hybridisation for slides/microarrays, association of physical unit identifiers with sample IDs, hybridisation oven temperature and hybridisation time (e.g. 12 hours).

3.1.2.6. Scanning of microarrays

The microarray slides are scanned using a scanner to obtain images of hybridised arrays. There are many companies (such as Agilent, Affymetrix, Illumina) that manufacture scanners to obtain microarray images from the hybridised microarray slides and the scanning protocol may differ from company to company. The following parameters associated with the scanner should be reported:

REPORT:

a. Scanning protocol

b. Name of instrument manufacturer

c. Type of instrument (Scanner type)

d. Scan rate

e. Scanning time

f. Lasers used (wavelength or type) and power

g. Other pertinent settings

h. Description of output type (e.g. .tiff image)

i. Other relevant information

3.1.2.7. Array Quality Control Metrics/Criteria

Quality control is among the most important of quality assurance measures. Controls are used to check assay performance, with special focus on the least robust components. Although traditional single-analyte assays require inclusion of a positive and a negative control in every run, it is clear that microarray runs cannot possibly include controls for each of the dozens to thousands of target analytes. Therefore, a new model of quality control has been developed to accommodate multi-analyte arrays. Multiple types of controls are used in RNA profiling.

Manufacturers of commercial microarray platforms (i.e. Affymetrix, Agilent, etc.) often recommend various types of quality control criteria and performance checks for their particular products. These may include visual inspection of scanned microarray images for bubbles, scratches and grid alignment, and computational evaluation of the homogeneity of hybridisation, uniformity of background hybridisation, dynamic range of gene expression and percentage of detectable genes (> 25%) and (if applicable) linearity of signal for spike-in RNAs.

Because of inherent biological variability in levels of any given gene product, several housekeeping genes that are consistently expressed at low to high levels in the relevant tissue or biofluid are often used to assess sample quality. For example, adequate expression of these housekeeping genes reflects suitable hybridisable RNA, thus allowing elimination of poor-quality samples. Manufacturers of commercial microarray platforms also often recommend various types of quality control criteria and performance checks for their products based on housekeeping gene measurements. These can include qRT-PCR based evaluation of housekeeping gene abundance prior to experimentation, and comparison of 3’ and 5’ expression ratios for housekeeping genes. In addition, the expression of housekeeping genes can be used to normalise the quantity of target RNAs.

Exogenous controls may be run alongside samples to evaluate assay performance in a general manner. Exogenous controls may be prepared by mixing a cell line or RNA derived from that cell line with an appropriate matrix, and serial dilutions can be used to challenge analytical sensitivity or linearity. Particular care needs to be exercised with this method as some lot-to-lot variation is expected even when precise criteria are defined for cell culture and harvest and there is the inherent danger of RNA degradation and variance in quantification. Some scientists prefer a mixture of several cell lines to fill in gaps that an individual cell line might have (e.g. non-expressed genes). When the same control material is used in multiple runs, selected numeric results can be tracked over time, e.g. using Levey-Jennings charts to visualise drift or shift. In addition, a “no template” control can be used to evaluate background signal and contamination by stray nucleic acids.

Spiked controls are another tool for assessing assay performance, and commercial RNA standards for this purpose are available. In this approach, exogenous RNAs are spiked into each sample at the earliest informative time point (e.g. with lysis buffer) to permit downstream evaluation of assay performance within the sample. This approach can detect technical failure or endogenous interfering substances (e.g. haemoglobin or background autofluorescence). Finally, combinations of spiked molecules have been proposed as a way of tracking sample identity through specimen preparation and analysis.

Generally, limits on acceptable performance of controls are empirically set by replicate analysis. Technical replicates can be run across different days, by different technicians, using different lot numbers (etc.) to assess the performance. When multiple controls are used, the expected failure rate increases accordingly. When combined with sample quality indicators, results of controls can help identify sources of technical and experimental errors.

REPORT:

a. Quality control approach/sample type(s)

   Examples are housekeeping genes, spiked controls, exogenous controls, no template controls, etc. State the type and source of each.

b. Quality control applicability

   Describe what aspects of sample processing they are designed to evaluate (e.g. efficiency of amplification or labelling, hybridisation, etc.). Describe what level of the experimental design they are intended to address (e.g. individual sample quality, sample batch quality).

c. Quality control performance metric(s)

d. Quality control accept/reject criteria

e. Technical and experimental replicates

   Describe the intended use of technical and experimental replicates for quality control. Describe summary/aggregation strategies across replicates.

f. Evaluation metrics for spike-in controls

g. Reproducibility for replicated probes (e.g. the median CV of replicate probes)

h. Summary measures of the negative control spots (e.g. mean and standard deviation)

i. Quality control results

   Provide a summary of quality control results as a data object. Example is a Pass/Fail score for each sample for each quality control metric. Indicate which samples were included or excluded based on quality control results.

j. Other relevant information

3.1.3. Specification of Raw Data

Images from a scanned microarray slide contain features (spots) of various signal intensities, comprising the raw data for a microarray experiment. The signal intensity of each feature denotes the magnitude of abundance of hybridised probes, which represents the degree of gene expression in each biological sample. The feature signal intensities are processed and translated into continuous data (numerical values) by a feature extraction software using a multi-step algorithm. As different platforms and software can produce different types of raw data, a detailed description of how and in what form the raw data is generated is crucial for downstream analysis such as identification of

differentially expressed genes. Reporting of the following parameters would help reproduce and verify the results of microarray studies.

3.1.3.1. Feature Extraction Software and Outputs

Several manufacturer-specific and some standalone software for the extraction of quantitative transcriptomic data from the scanned images are utilised by researchers. Please report the following parameters associated with the feature extraction software and data files generated as outputs:

REPORT:

a. Feature extraction software:
   - i. Name and version of feature extraction software
   - ii. Name of grid template/array design file
   - iii. Name of protocol used
   - iv. Other pertinent settings (e.g. if a manual fitting or adjustment has been performed)

b. Output files:
   - i. Type of files generated (e.g. raw intensity files, QC metrics, QC report)
   - ii. File extensions (.txt, .pdf)
   - iii. Naming convention
   - iv. Description of association of quantification matrices with raw data
   - v. Experimental metadata file
   - vi. Retention as part of experimental record? (YES/NO)
   - vii. Archiving location
   - viii. Other relevant information

3.1.3.2. Description of Raw Data

Different companies use their proprietary feature extraction algorithm to generate different types of raw data. The quantification of raw data starting from analysing the pixels of coloured spots (features) to final processed raw data includes multiple steps. These steps may include calculation of background noise, inter probe variability, correction based on background or a factor, flagging of outlier features, etc. Please report the following details associated with the generation of raw data and QC metrics:

REPORT:

a. Description of raw data table(s)

b. Type of raw data used (mean, median or processed signal intensities)

c. Background subtraction/correction (yes/no)

d. Multiplicative detrending or similar correction for probe variability (yes/no)

e. Removal of flagged features (yes/no) (reporting on how this was done is below)

f. Method (e.g. mean, median) of handling of replicate probes for calculation of gene level summaries

g. Applicability, performance metric and acceptability criteria of feature or probe level QC metrics

h. Other relevant information

3.1.3.3. Availability of Raw Data

Raw unprocessed gene expression data should be accessible to the public/researchers to facilitate reproduction of data processing steps and final results. Today most journals require researchers to submit their raw and/or processed transcriptomic data into public repositories such as the Gene Expression Omnibus (GEO) of the National Center for Biotechnology Information (NCBI) or ArrayExpress of the European Molecular Biology Laboratory - European Bioinformatics Institute (EMBL-EBI). Please report the following items:

REPORT:

a. Name of public repository or provision of link to private repository

b. Accession number or equivalent identification number of the submitted data to facilitate data retrieval

c. Format of submitted data (e.g. .txt, .xls, etc.)

3.1.4. Data Filtering (Pre-normalisation)

Filtering can be applied pre or post normalisation. This module specifies the information a regulatory scientist needs to understand the types of filtering that were applied to gene expression data. Filtering of gene expression data is dependent on what the researcher's downstream analysis goals are. Filtering impacts the power to detect differentially expressed genes, as well as pathway or gene set enrichment analyses. Filtering and the percentage of the data to be removed also impacts signature or biomarker development, cluster analysis and network construction as the inclusion of these features increases the risk of overfitting the training data. In toxicogenomics, the transcriptional benchmark dose is also influenced based on how genes are filtered. It has been shown that not applying appropriate gene filters impacts inter laboratory reproducibility studies and reproducibility across different microarray platforms. Filtering can be applied to single channel data and the ratios derived from an experiment to control comparison. These factors will impact the results and need to be considered and reported.

Reporting requirements for the most used pre-normalisation filtering have been included in the section below, as well as an option to report on other approaches and special scenarios.

3.1.4.1. Filtering by Signal Intensity

The objective of filtering by signal intensity is to remove genes that have signal intensities that are within the background noise. The rationale for filtering these features is that genes with background signal intensities are considered less

reproducible. It is assumed that genes with low signal intensities will likely not be differentially expressed; however, discarding low-intensity genes may potentially remove interesting differentially expressed genes.

REPORT:

a. Background Distribution

Describe how the background distribution was calculated. Report if and how the following (or other) elements were used in this calculation:

i. Background distribution calculation

ii. Local background from the quantification software

iii. Negative control spots

iv. Signal intensity distribution

v. Other relevant information

b. Background Threshold

Report how the background threshold was determined and indicate the specific threshold used. This may be based on:

i. Statistical test between the probe foreground and background intensity

ii. A statistic estimated from the negative control spots

iii. Quantile from the signal intensity distribution

iv. Other relevant information

3.1.4.2. Other Filtering Methods

If none of the above filtering methods was used, please report the method that was used following the aforementioned report elements.

REPORT:

a. Other relevant information

3.1.4.3. Special scenario

Procedures for values where no ratio (between test and control group) can be calculated because the expression valued is 0 in the control. This will only apply to upregulated genes in the test sample. Where there is an upregulated gene in the test sample that falls below a threshold for detection in the control, a ratio cannot be calculated. If ratios are used, please state how these values are handled in the statistical procedures and filtering methods.

REPORT:

a. Approach for handling expression ratios with 0 in the denominator.

3.1.5. Data Normalisation

Normalisation is the process by which data are adjusted to take account of technical variation in the study. For example, when using two fluorophores in microarray experiments, the two fluorescent probes can have different fluorescent properties, or the excitation lasers may have a different efficiency. In addition, there may be variances in the efficiency of extraction or quantification of the RNA. These and other factors create technical variation in the experiment that is accounted for in the process of normalisation. The process is similar to that used in blots for DNA, RNA and protein, when typically a second gene that is highly expressed, and considered to be stable in expression, is used to control for variances in the electrophoresis or gel loading. For omics methods, the process is applied across a large data set, which does give rise to some challenges. If a single gene is used for the normalisation, as used in blot analysis, then the whole omics data set (which can be substantial) will be subject to any variance in the gene used for the normalisation. Thus, microarrays are generally normalised through an accepted global-normalisation approach. There are many such approaches and frequently there are manufacturer-specific recommendations.

Reporting elements are provided below for a variety of approaches; the relevant methodology should be selected to report on the pertinent parameters applied in the analysis.

3.1.5.1. Data Normalisation by Manufacturer Process

Many manufacturers have proprietary methods for the normalisation of the data for their products. If these are used with no further processing of the data, then only minimal information needs to be provided.

REPORT:

a. Manufacturer of the product used for normalisation

b. Manufacturer's normalisation method used

c. Any deviance from the manufacturer method

d. Software package used and date/version

3.1.5.2. Data Normalisation by Mean or Median Centering Within Sample Data Sets (non-scaling)

In this global normalisation strategy, the mean or median of the data set for each sample is found and the rest of the data are centred by dividing by this value such that the mean or median of the sample data set is 1 (or 0 if the data are $\log_2$ transformed). The process does not change the distribution of the data. When plotted on a box plot if median centred, then the centres of each box on the plot of the medians will be aligned. In some cases the mean can be trimmed to only use a proportion of the data. If this is done it should be reported.

This process can be summarised:

For each column (j) of data (X_{ij}) where the columns are the data from each individual sample; for j = 1 to n, where n is the total number of data columns, and i = 1 to g, where g is the total number of rows, compute the mean or median (M), where $M_j = \text{median}_{i=1 \text{ to } g} \{X_{ij}\}$. Then, the normalised data for the sample $X_j^* = X_j / M_j$.

REPORT:

a. If the data are one channel (one fluorescent label) or two channels

b. If a background data subtraction was applied and at what point in the process relative to the normalisation step

c. The method of background calculation

d. Describe any weighting procedures that were applied

e. If the data are two channel, report if data were normalised before (on single channels) or after the calculation of the ratio

f. If the data were transformed (e.g. $\log_2$), and if so before or after the normalisation

g. If the data were trimmed before the mean or median (a trimmed mean) was calculated, and if yes, how

h. If control and/or negative sample data were removed from the data set prior to the normalisation

i. Other relevant information

3.1.5.3. Data Normalisation by Mean or Median Centering Across Sample Data Sets (scaling)

This process differs from that used in 3.1.5.2 in that one sample data set is chosen as the comparator, and the mean or median of this sample data set is calculated. The data elements in each sample data set are then adjusted such that the sample data set has the same mean or median as the reference sample data set that was chosen.

This process can be summarised:

For each column (j) of data (X) where the columns are the data from each individual sample; for X_1 to n where n is the total number of data columns, compute the mean or median for one sample data set X_j. For each of the other data sets calculate their mean or median and divide by the mean or median of the sample data set X_j to get a *scaling factor* for each sample. Divide each of the data elements by this scaling factor for each sample.

REPORT:

a. If the data are one channel (one fluorescent label) or two channels

b. If a background data subtraction was applied and at what point in the process relative to the normalisation step

c. The method of background calculation

d. Describe any weighting procedures that were applied

e. If the data are two channel, indicate if data were normalised before (on single channels) or after the calculation of the ratio

f. If the data were transformed (e.g. $\log_2$), and if so before or after the normalisation

g. If the data were trimmed before the mean (a trimmed mean) was calculated, and if yes, how

h. If control and/or negative sample data were removed from the data set prior to the normalisation

i. Which sample data set was used for the calculation of the mean or median to be used as reference mean or median (a sample data set is that data derived from one sample and consist of a number of data elements, each element corresponding to one gene or feature)

j. Other relevant information

3.1.5.4. LOWESS normalisation

This process of normalisation takes into account signal intensity in the normalisation, which is different from 3.1.4.2 or 3.1.4.3 above. In this method the ratio of the data between the experimental sample and the control is used. These data may be from two sets of single channel data or from two channels in a dual channel microarray hybridisation.

In this method the log ratio (M) and the log intensity (log(sqrt(experiment * control))) (A) of the experiment/control channel is calculated. An M/A plot is then produced where the log ratio is plotted against the intensity. A smoothed lowess fit is then produced through the data and the individual ratios across the set individually adjusted by reference to the smoothed fit such that the new smoothed fit lies on M=0 through the data. The new ratios are then used for the statistical analysis to statistically determine gene expression changes.

REPORT:

a. If the data are one channel (one fluorescent labelling) or two channels

b. If a background data subtraction was applied and at what point in the process relative to the normalisation step

c. The method of background calculation

d. If control and/or negative sample data were removed from the data set prior to the normalisation

e. Describe any weighting procedure that was applied

f. The process used for derivation of the M/A plot

g. The formula used for the calculation of the polynomial fit to the data

h. If the polynomial fit was calculated on the whole sample data set or on a print tip basis

i. Other relevant information

3.1.5.5. Quantile normalisation

This normalisation method, unlike the others above, standardises not only the mean or median of the data but also the distribution.

In this method each sample set of data (which could be single channel or two channel ratio data) is ranked from lowest to highest expressed gene. The mean of the ranks for all sample data across the experiment for each gene is derived. These means are then substituted for the ranks and used as the expression values. It is imperative that each sample set of data is the same length for this method.

REPORT:

a. If the data were ranked on single channel data or calculated ratio data

b. If a background data subtraction was applied and at what point in the process relative to the normalisation step

c. The method of background calculation

d. Any weighting procedures that were applied

e. If the data were transformed (e.g. $\log_2$), and if so before or after the normalisation

f. If the data were trimmed before the mean (a trimmed mean) was calculated, and if yes how

g. If control and/or negative sample data were removed from the data set prior to the normalisation

h. Other relevant information

3.1.5.6. Robust Multiarray Averaging (RMA)

The RMA method is a normalisation strategy designed for the Affymetrix GeneChip® system. RMA is a summary measure that is a robust multi-array average (RMA) of background-adjusted, normalised, and $\log_2$-transformed of the perfect match values. RMA normalises the arrays using the quantile normalisation approach, but also usually includes a calculation to remove background.

REPORT:

a. Link to the protocol used

b. How the background was calculated and to which targets if this is applicable

c. Any weighting procedures that were applied.

d. If positive controls (spiked in probes) were used and how these were incorporated in the calculation

e. If the data were ranked on the single channel data or calculated ratio data

f. If the data were transformed (e.g. $\log_2$), and if so before or after the normalisation

g. If the data were trimmed before the mean (a trimmed mean) was calculated, and if yes how

h. How the ranking normalisation was achieved

i. Other relevant information

3.1.5.7. Other normalisation methods

If none of the above normalisation methods was used, please report the method you used following the aforementioned report elements.

REPORT:

a. All relevant information

3.1.5.8. Availability of normalised data

Normalised gene expression data should be accessible to the public/researchers to facilitate reproduction of data analysis steps and final results. Normalised data is often housed in public repositories such as GEO or ArrayExpress alongside raw data. Please report the following items:

REPORT:

a. Name of public repository or provision of link to private repository

b. Accession number or equivalent identification number associated with the normalised data to facilitate data retrieval.

c. Format of submitted data (e.g. .txt, .xls, etc.)

d. Description of raw data table(s).

3.1.6. Data Filtering (Post-Normalisation)

The rationale and importance of applying filtering methods have been discussed in aforementioned section "Data Filtering (Pre-Normalisation)".

This module specifies reporting requirements for the most used post-normalisation filtering, as well as an option to report on other approaches.

3.1.6.1. Filtering by Probe Level Variability

The objective of filtering by probe level variability is to remove genes that lack a degree of consistency between replicate probes. Replicate probes with large variabilities or coefficients of variation would be considered unreliable.

REPORT:

a. How the probe level variability was measured. Examples include variance or coefficient of variation of the technical replicates.

b. How the probe level variability cut-off was determined and applied. An example is use of a quantile estimate from the distribution of all probe level estimates

c. Other relevant information

3.1.6.2. Other Filtering Methods

If none of the above filtering methods was used, please report the method you used following the aforementioned report elements.

REPORT:

a. All relevant information

3.1.7. Identification and Removal of Low Quality or Outlying Data Sets

The primary purpose of removing an outlier sample is to decrease the leverage of the sample on any downstream analyses and the within group variance so that the statistical differences between comparison groups and identification of effects due to test article treatment can be maximised. However, identification and removal of outlier samples should be performed in a scientifically justified manner. Outlier samples (data sets) can be defined as samples containing extreme values, which are very different compared to the rest of the samples within a group. Outliers can result from variability in experimental steps, poor hybridisations, data acquisition or scanning errors such as misaligned grids or unique biological response. Identification of an outlier sample can be performed using different methods such as principal component analysis, cluster analysis, box plots, etc. There can be several additional justifiable reasons for removal of an outlier such as failed microarray QC metrics, low dye incorporation, low signal to noise ratio, failure of spiked in controls, etc. Please report the following parameters used to identify and remove outlying dataset:

3.1.7.1. Outlier and Low Quality Data Removal

REPORT:

a. Describe how outlier samples were identified and reason for removal

b. Threshold, if any

c. Processing step where exclusion occurs

d. List of samples excluded and per sample justification for exclusion

e. Removal of outliers before normalisation? (if so: provide justification and describe applied algorithm)

f. Removal of additional outliers after normalisation? (if so: provide justification and describe applied algorithm)

g. Other relevant information

3.1.8. References

Bryant S., Manning D.L. (1998) “Isolation of messenger RNA”. Methods Mol Biol. 86:61-4.

Chomcynski, P., Sacchi, N. (1987) “Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction.” Anal Biochem 162(1):156-159.

Chomcynski, P., Sacchi, N. (2006) “The single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction: twenty-something years on.” Nat Protoc 1(2): 581-585.

Chung, J.Y., Cho, H., et al. (2016) “The paraffin-embedded RNA metric (PERM) for RNA isolated from formalin-fixed paraffin-embedded tissue.” Biotechniques 60(5): 239-244.

Grillo M., Margolis F.L. (1990) ”Use of reverse transcriptase polymerase chain reaction to monitor expression of intronless genes.” Biotechniques. 9(3):262, 264, 266-8.

Vomelová I, Vanícková Z, et al. (2009). “Methods of RNA purification. All ways (should) lead to Rome.” Folia Biol (Praha). 55(6):243-51.

Wilfinger, W.W., Mackey, K., (1997) “Effect of pH and ionic strength on the spectrophotometric assessment of nucleic acid purity.” Biotechniques 22(3): 478 – 481.

3.2. RNA Sequencing and Targeted RNA Sequencing

RNA Sequencing (RNA-Seq) and targeted RNA-Seq technologies allow both the identification and quantification of RNA molecules expressed in a given biological sample. The most commonly used methods for RNA-Seq applications in toxicological research generally involve whole transcriptome sequencing and alignment against a reference genome or transcriptome. Methods for targeted RNA-Seq, such as Templated Oligo with Sequencing Readout (TempO-Seq; Yeakley et al. 2007) and RNA-mediated oligonucleotide Annealing, Selection and Ligation with Next-Gen sequencing (RASL-seq; Li et al. 2012), use targeting probes to quantify gene expression. These targeting probes anneal to specific RNA sequences, become ligated and are used as input for sequencing to measure levels of transcript abundance through sequencing and counting of ligated probes. Targeted RNA-Seq uses similar bioinformatic pipelines to RNA-Seq, including alignment to a reference genome or transcriptome, or alternatively, using a targeting probe manifest. For both RNA-Seq and targeted RNA-Seq, following alignment, the number of reads assigned to each gene or transcript are counted to determine the level of gene expression. However, RNA-Seq and targeted RNA-Seq applications are extremely diverse, and the output will be determined by many parameters (library preparation methods, sequencing platform, coverage, data processing pipeline, normalisation methods). To be used in regulatory decision making, the complete experimental protocol (for molecular techniques, sequencing and informatics) needs to be fully documented and reported.

This module provides a reporting framework for describing the RNA-Seq or targeted RNA-Seq technology and methodology used in a toxicology experiment. It aims to guide the users on how to document all of the study design parameters required to understand and analyse the experiment, including the platform-specific sample processing steps, library preparation protocol details, raw data acquisition and how quality control, alignment, gene quantification, normalisation, data filtering and outlier identification were conducted.

3.2.1. Technology

This section describes the information a regulatory scientist requires to determine analytical sensitivity, limits of detection, interference, and precision (reproducibility and repeatability) of the RNA-Seq or targeted RNA-Seq technology used in a transcriptomic experiment.

Since RNA-Seq and targeted RNA-Seq can be performed using a variety of technologies and configurations within each technology, the main output may vary substantially across different protocols. The following parameters will be the main source of variability that will impact the outcome:

- RNA Extraction method [if relevant]
- RNA Enrichment method [if relevant] (e.g. poly[A] enrichment, rRNA depletion, etc.)
- Targeting probe mixture (for targeted RNA-Seq only)
- Library preparation method (including sample indexing and batch processing)
- Sequencing Platform
- Sequencing Coverage

- Sequencing pre-processing (QC and Trimming)
- Alignment tool
- Genome Reference
- Gene quantification methods
- Normalisation of the raw data
- Data Filtering
- Outlier detection

All these steps will be described individually in this reporting framework for RNA-Seq and targeted RNA-Seq transcriptomic analysis. The general feature and goal of the analysis should be reported first. If possible, provision of a link to the pipeline or code repository used to process the data should be provided.

REPORT:

3.2.1.1. Type and version of the sequencing platform (e.g. Illumina HiSeq2500)

3.2.1.2. Size and type of sequencing (e.g. 100 bp paired-end)

3.2.1.3. Flow cell used (type and catalogue number)

3.2.1.4. Targeting probe annotation, including the list of attenuated genes (if any) (for targeted RNA-Seq only)

3.2.1.5. Library type (e.g. mRNA libraries)

3.2.1.6. Purpose (e.g. target gene expression, quality control, etc.)

3.2.1.7. Other relevant information

3.2.2. Transcriptomics Experimental Design

The experimental design starts with defining the RNA extraction method (if used), the sample pooling strategy, batch processing (if applicable), the possible use of spike-in (or other internal control), the library preparation method and the analytical processes applied to assess sample and library preparation quality.

3.2.2.1. RNA Processing

The methodologies used for RNA sample collection, processing and storage will impact the final research results and should be reported in detail. The RNA extraction method will have a direct consequence on the sequencing results (for example, RNA extraction methods that utilise binding to a silica matrix usually will not recover RNA molecules under 200 nucleotides efficiently, which makes library preparations for small RNA (e.g. micro RNA) impossible). The RNA quality is usually assessed by measuring the integrity of the ribosomal RNA, either through manual gel electrophoresis, or various automated methods that provide a numerical quality score of integrity (such as the RNA integrity number (RIN) (Schroeder et al. 2006), Bioanalyzer/tape station or RNA Quality Score (RQS, LabChip).

The procedures used to quantify and qualify RNA should be reported to enable an appropriate evaluation of the RNA extraction steps and suitability for use in downstream transcriptomic analyses. Quality thresholds used to define samples of sufficient RNA integrity for transcriptional analysis should be defined.

RNA-Seq technology is not typically applied directly on total RNA, since sequencing ribosomal RNA would rarely be interesting. Generally, either an enrichment method to specifically select the RNA to be sequenced (such as poly[A] enrichment to isolate the mRNA or targeted gene amplification) or a depletion method to remove an RNA target (most commonly the ribosomal RNA but other types of RNA may be targeted) is applied. Targeted RNA-Seq measures specific transcripts and thus does not require elimination of ribosomal RNA. However, some library preparation protocols start from total RNA and include a specific enrichment method. For instance, Combo-Seq (from Perkin Elmer) can be used to sequence both mRNA (based on poly[A] selection) and miRNA in one run.

Together, these parameters should be considered to evaluate the quality of the generated transcriptome. For example, RINs showing degraded RNA (RIN <7) make poly(A) enrichment of total RNA inadvisable, since many mRNAs will have lost their poly(A) tail integrity during degradation.

Targeted RNA-Seq technologies are compatible with purified RNA prepared as described above. However, some targeted RNA-Seq technologies are also compatible with cell or tissue lysates and thus may not require RNA extraction. If this is the case, the reagents and methods associated with creating the cell or tissue lysates should be described in detail. Steps taken to evaluate RNA integrity or quantity cell of tissue lysates should also be described in detail.

REPORT:

Describe method used for preparation of RNA samples if relevant.

a. RNA extraction
    i. Type of extracted RNA (e.g. total RNA, mRNA, miRNA, etc.)
    ii. Extraction and purification techniques
    iii. Procedures for specific RNA enrichment or depletion procedures (e.g. poly[A] enrichment for mRNA, Ribosomal RNA depletion...)
    iv. Storage conditions
b. Quantification and Qualification of RNA
    i. Tool for RNA assessment (e.g, Nanodrop, QuBit, Bioanalyzer…)
    ii. RNA quality (e.g. A_{260}, A_{260}/A_{280} ratio, RIN for eukaryotic RNA, PERM number (Chung et al. 2016) for eukaryotic RNA extracted from formalin-fixed paraffin embedded tissues
    iii. RNA quantity (e.g. RNA/g of tissue, RNA/10^6 cells, Other).

3.2.2.2. Library Preparation

Many library preparation kits are offered by a variety of companies. While most RNA library preps share some essential steps (fragmentation of the RNA, adapters

and/or index ligation, reverse transcription and amplification of cDNAs), the differences in the details of these methods can be important for interpretation and must be reported.

For targeted RNA-Seq studies, the collection of targeting probes used to quantify gene expression can vary from study to study and have a large impact on the identity and abundance of genes that are measured. The targeting probe mixture should be described in detail or a link to a targeted probe manifest available on a public facing repository should be provided. In addition, the protocol used to perform annealing, probe ligation, PCR amplification and sample barcoding should be described in detail. Alternatively, a link to a protocol available on a public facing website or repository should be provided. Deviations from said protocols should be documented within the OORF reporting structure.

Library preparation methods for RNA-Seq and targeted RNA-Seq must be reported fully, since not all library preparation methods will be comparable. Having the complete methodological details can also reveal potential methodological problems that can impact the reproducibility of results (e.g. if a poly[A] selection strategy has been applied to low RIN samples).

REPORT:

a. Library preparation applied (full name, manufacturer and catalogue number of the kit used)
   i. Manual library preparation or automated system (if yes, which automation system)
   ii. Fragmentation strategy (if applicable)
   iii. Probe manifest (if applicable)
   iv. Number of PCR amplification cycles
   v. Targeting probe annealing, ligation and PCR amplification steps (if applicable)
   vi. Other relevant information

3.2.2.3. Sample Pooling

With the exception of low throughput Illumina sequencing platforms (i.e. MiSeq, iSeq or MiniSeq), the sequencing output of the majority of sequencing platforms is too high to cost-effectively sequence a single transcriptome sample at a time. Instead, samples are pooled and sequenced together. To achieve this and still be able to associate sequence reads to the original samples, sample specific indices (barcodes) are added during the library preparation methods and sequenced either as a separate index read, or from the beginning of the sequence read. There are a variety of strategies for indexing, with indices ranging in size (most frequently 8 bp but ranging between 6 and 12 bp), and number (used either as single barcodes or with two barcodes in combination for the forward and reverse directions). Indexing barcodes are used for library preparation in both RNA-Seq and targeted RNA-Seq approaches.

When considering an indexing strategy, the decision will depend on the sequencing platform, number of samples to be pooled together and library type being sequenced. The sequence of the pooled indices must be considered to give a balance of the different nucleotides in index reads, and the similarity between index sequences needs to be sufficiently distinct so as to allow for correcting basecall errors. While traditionally a single-indexing strategy was sufficient for many applications, the increased throughput of modern sequencers requires greater levels of multiplexing. The increase in index mis-assignment due to index swapping has changed modern best practices over to the use of Unique Dual Indices (UDIs), where all samples have a barcode on their 5'and 3'adaptors with neither index shared with any other samples in the pool.

REPORT:

a. Whether any sample pooling has occurred (samples being combined into one sequencing run) – yes/no

b. Indexing strategy (e.g. single/dual barcodes, sizes)

c. Number of samples pooled per sequencing unit (lane or flow cell)

d. Table of barcodes assigned to each sample

e. Other relevant information

3.2.2.4. Batch Processing of Samples (if applicable)

While the correlation between two independent sequencing runs of a given library pool is usually very high (above 0.9), an important part of the variability between samples comes from the library preparation itself. The number of samples to be processed does not usually allow all library preparations to be prepared in a single batch. It is therefore important to assign samples from different treatment groups (and/or different treatment duration, cell types, doses etc.) to different batches and to try to include equal number of samples from each group in order to minimise batch effects and maximise biological differences (or to carry out batches in replicate sets). Please report the following parameters:

REPORT:

a. Number of batches

b. Number of samples used in each batch of the sequencing experiment

c. Method of sample assignment to different batches (for example, simple randomisation, randomised block design etc.)

3.2.2.5. Wet Lab Quality Control

Several quality control steps should be used to assess the quality and concentration of the generated libraries and ensure the robustness of the produced sequence data. An analysis of the produced libraries with a chip-based automated electrophoresis system to identify the size distribution of the generated libraries is often conducted prior to any sequencing. This applies to libraries produced as part of RNA-Seq or targeted RNA-Seq workflows.

The use of spiked controls is another tool for assessing the library preparation performance, and commercial RNA standards are available for this purpose. In this approach, exogenous RNAs are spiked into each sample at the earliest informative time point (e.g. with lysis buffer) to permit downstream evaluation of assay performance within the sample. This approach can detect technical failure or endogenous interfering substances.

Finally, since most library preparation protocols include a PCR amplification step of the cDNA to reach a sufficient amount of molecules to sequence and add the sequencing adapters, the sequencing of PCR clones can be an important source of bias in the analysis. To limit the amount of impact that PCR bias has on the final library, the number of PCR cycles should be minimised. Recent library preparation methods have also introduced the use of Unique Molecular Identifiers (UMIs), which adds one of a pool (typically of several million) random barcodes to PCR amplified products, allowing for identification of PCR products that came from the same original fragment.

All these quality control steps will help in judging the quality of the produced sequences.

REPORT:

a. Use of chip-based automated electrophoresis system on the libraries (join profile if available)

b. Use of spiked-in control
   i. Type of spike-in used (sequences)
   ii. Amount of sequence used
   iii. Source of spike-in (manufacturer, catalogue number, etc.)
   iv. Step of introduction of the spiked control in the protocol

c. Use of Unique Molecular Identifiers (UMIs)
   i. Type of UMIs
   ii. Source of UMIs (manufacturer, catalogue number, etc.)

d. Other relevant information

3.2.2.6. Sequencing Quality Control Metrics/Criteria

Manufacturers of commercial sequencing platforms often recommend that a certain percentage of reads on a flow cell include an internal control in the library to assess the sequencing quality run and obtain a reference GC balance. Illumina recommends using the genome of the phage PhiX. However, technically, any other source of internal control could be used to assess the sequencing quality in the different lanes and flow cells. Other sources of sequencing quality control are the number of clusters passing the filtering step (% PF). The difference between the expected number of clusters for a given flow cell and the % PF, together with a high duplication rate, could indicate a clustering issue (under or over clustering).

REPORT:

a. Quality control standard (e.g. PhiX genome)
   i. Type of standard
   ii. Source of standard (manufacturer, catalogue number, etc.)
   iii. Quantity used
b. Sequencing quality metrics
   i. Number of clusters passing filtering
   ii. Average quality score
   iii. Other relevant information

3.2.3. Analysis of Raw Data

The true raw data of Illumina sequencing platforms are high resolution pictures of each sequenced cycle. These pictures are automatically converted into compressed files (in BCL format), which are usually not stored but instead are directly converted into FASTQ files. The conversion of BCL to FASTQ files is usually done by the sequencing platform software, and following the recommended manufacturer's protocol (applying various pre-processing steps, such as demultiplexing with error correction, automated adapter removal, splitting into quality scores, etc).

The ultimate goal of a transcriptomic analysis is the identification and quantification of all genes expressed in a biological sample. The data processing pipeline used to obtain a final read count per gene will directly depend on the software and tools included in the workflow. Therefore, the complete framework needs to be documented (including software versions, genome version, etc.) and reported to facilitate assessment of data provenance and quality.

3.2.3.1. Base Calling

The base calling step, consisting of converting BCL files to FASTQ, is usually made directly on the sequencing instrument. Illumina recommends using bcl2fastq software, which has evolved over the years and different versions of the algorithm exist. The number of FASTQ files generated depends on the experimental design and the sequencing protocol applied. Paired-end sequencing typically produces two files per sample (R1 and R2, or forward and reverse). It is also common to run a single sample on different sequencing lanes of a flow cell, which will then produce a file (or two, if paired-end reads are used) per sample per lane. These files are then usually concatenated across the lane and named with a convention that should be explained (for instance, Cell_Compound_dose_time_replicate.fastq). If a number (or string of characters) is used for sample identification, a metadata file with the sample description must be included.

Please report the following parameters associated and data files generated as outputs:

REPORT:

a. Base calling software:

i. Name and version of base calling software

ii. Quality score version (e.g. Phred33)

iii. Other relevant information.

b. Output files:

i. Naming convention (or sample ID metadata)

ii. Experimental metadata file

iii. Other relevant information

3.2.3.2. Availability of Raw Data

Raw unprocessed gene expression data should be accessible via public databases or through communication with researchers themselves to facilitate reproduction of data processing steps and final results. Most journals require researchers to submit their raw and/or processed transcriptomic data into public repositories such as the Gene Expression Omnibus (GEO) of the National Center for Biotechnology Information (NCBI) or (especially for RNA-Seq data) the European Nucleotide Archive (ENA) of the European Molecular Biology Laboratory - European Bioinformatics Institute (EMBL-EBI). Please report the following items:

REPORT:

a. Output files:

i. Name of public repository or provision of link to private repository.

ii. Accession number or equivalent identification number of the submitted data to facilitate data retrieval.

iii. Format of submitted data (e.g. .fastq, .fastq.gz, .bam, etc.)

3.2.3.3. Raw Data Filtering

RNA-Seq and targeted RNA-Seq data analysis usually starts by removing the samples with insufficient sequencing depth. Indeed, while the number of reads per sample is expected to be close to the total number of reads in the lane divided by the number of indexed samples in the lane, this calculation is more accurately an estimate of average reads per sample with individual samples both above and below this estimate. In instances where there is high variability in reads per sample, it is sometimes better to exclude samples with too few reads (since they would have an important impact on the normalisation).

Following this exclusion based on read depth, FASTQ reads can be trimmed to either shorten the effective read length, or to remove the reads that would have too low a quality score. Filtering and trimming will have important consequences on the gene quantification and must be reported.

Please report the following items:

REPORT:

a. Data filtering and trimming
    i. Minimum read count
    ii. List of samples excluded and rationale (thresholds) for exclusion
    iii. Trimming algorithm/software (e.g. FastQC v0.11.8, Trimmomatic v0.39)
    iv. Trimming parameters (e.g. leading 3, trailing 3, sliding window 4:15)
    v. Other relevant information

3.2.3.4. Sequence Alignment

Once trimmed, the reads need to be aligned (or mapped) to a reference genome, transcriptome, or probe manifest (in the case of targeted RNA-Seq). *De novo* transcriptome assembly, while possible from high coverage data, will not be considered herein for regulatory application. Mapping will generally produce an alignment file (of .bam or .sam extension), which will be used for quantifying the level of expression of genes (or transcripts). Please report the following items:

REPORT:

a. Alignment
    i. Source and version of the genome reference and/or the transcriptome (e.g. Ensembl Homo sapiens GRCh38 release 99, Top level, primary assembly, masked or not, etc.).
    ii. Mapping software (e.g. STAR v2.7.0a)
    iii. Mapping parameters (gene or transcript level, number of mismatch, gap penalty, etc.)
    iv. Other relevant information

3.2.3.5. Gene Quantification

Once an alignment has been generated, a quantification tool is usually used to attribute reads to each annotated transcript. Depending on the application, the gene expression value can be obtained at the level of each individual transcript, the individual probe in targeted RNA-Seq applications, or at the gene level (which is then usually the sum of all reads mapping to all transcripts of a gene). Quantification is usually made using a genome annotation file (such as a .gtf or .gff3 files). Some software can perform both mapping and quantification in a single command (e.g. “Salmon” for transcript level), and should then be reported in both sections (3.2.3.4.a. Alignment and 3.2.3.5.a. Quantification).

Please report the following items:

REPORT:

a. Quantification:

i. Source and version of the software (e.g. RSEM v1.1.17)

ii. Accession number or equivalent identification number of the submitted data to facilitate data retrieval

iii. Format of submitted data (e.g. .fastq, .bam, etc.)

iv. Description of method for summarising transcript or probe counts to gene counts (if applicable)

v. Other relevant information

3.2.4. Data Normalisation

Normalisation is critical for RNA-Seq and targeted RNA-Seq data analysis, since the generated number of reads per sample (and thus the number of detected biological entities) can be highly variable between samples. While historically, reads counts were simply scaled to the same order of magnitude between each samples (in counts per million, or CPM for instance), more sophisticated normalisation methods have now been developed, such as the trimmed mean of M values (TMM) and fitting the count to an expected negative binomial distribution.

3.2.4.1. Normalisation of the raw count

Normalisation is usually applied using an R package and some statistical methodologies which incorporate the normalisation and experimental design as part of the method (e.g. DESeq2).

Finally, the use of spike-in or UMIs at the library preparation step can also be used to normalise the raw reads counts.

REPORT:

a. Normalisation method applied

b. Package and version used (provide link if possible)

c. Factor used as design (if applicable)

d. Other relevant information

3.2.5. Post-normalisation Data Filtering

3.2.5.1. Identification and removal of low quality or outlying data sets

The primary purpose of removing an outlier sample is to decrease the influence of the sample on any downstream analyses and the within-group variance so that identification of effects due to the test treatment can be maximised. However, identification and removal of outlier samples should be performed in a scientifically justified manner. Outlier samples (data sets) can be defined as samples containing extreme values that are very different when compared to the rest of the samples within a group. Outliers can result from variability in experimental steps (well

contamination, pipetting error, etc.), poor library preparation or unique biological response.

Identification of an outlier sample can be performed using different methods such as principal component analysis, cluster analysis, box plots, etc. Please report the following parameters used to identify and remove outlier samples:

REPORT:

a. Describe how outlier samples were identified and the reason for removal

b. Threshold, if any

c. Processing step where exclusion occurs

d. List of samples excluded and per-sample justification for exclusion

e. Other relevant information

3.2.5.2. Specific gene filtering

Once the data are normalised (and thus the normalised counts are available), post-normalisation filtering is often required to exclude genes with certain behaviour (removing any gene before normalisation should not be done, to avoid creating bias due to different coverage between samples). For instance, DESeq2 includes a default parameter to evaluate the Cook's distance for excluding genes where the expression would be too influenced by a single data point. Independent filtering, usually performed to remove genes below a certain expression threshold (and thus decrease the impact of multiple testing correction), is also commonly applied (and is applied by default in DESeq2). All these steps will have consequences on the determination of differentially expressed genes and on future biological interpretation of the results, and should be reported.

REPORT:

a. Specific gene filtering

   i. Low read count filtering step (if any)

   ii. High variation among replicate filtering step

   iii. Other post-normalisation gene filtering steps (e.g. gene flagging)

   iv. Other relevant information

3.2.6. References

Chung, J.Y., Cho, H., et al. (2016) "The paraffin-embedded RNA metric (PERM) for RNA isolated from formalin-fixed paraffin-embedded tissue." Biotechniques 60(5): 239-244.

Schroeder, A., Mueller, O., et al. (2006) "The RIN: an RNA integrity number for assigning integrity values to RNA measurements." BMC Mol Biol 31(7): 3.

3.3. Quantitative Reverse Transcription Polymerase Chain Reaction (qRT-PCR)

Quantitative reverse transcription PCR (qPCR) determines transcript abundance by measuring a fluorescent signal emitted throughout the PCR amplification of a (set of) targeted gene(s). This technology can be used to measure a single gene target, or can be adapted to measure multiple gene targets via multiplexing (i.e. amplifying more than one gene per PCR reaction) and/or by employing a multi-well format (i.e. PCR array).

This module provides a reporting framework for describing a qPCR-based toxicogenomic experiment, including sample preparation, assay design, and data collection and processing. This module is largely based on the existing minimum information for publication of quantitative real-time PCR experiments (MIQE) guidelines (Bustin et al. 2009) but has been adapted for the OORF.

NOTE: To be consistent with the MIQE guidelines and Real-Time PCR Data Markup Language (RDML) data standard (Lefever et al. 2009), the *quantification cycle* (Cq) will be the term used to describe the fractional PCR cycle at which quantification is determined. Other common terms for this in the literature or by manufacturers include *threshold cycle* (Ct), *crossing point* (Cp), and *take-off point* (TOP).

3.3.1. Sample Processing

This section describes all of the sample processing that is required after tissues are collected from an exposure experiment up to before the qRT-PCR amplification reaction. This includes the RNA extraction protocol, assessment of RNA quantity and quality, and the reverse transcription of RNA into cDNA.

3.3.1.1. RNA Extraction

The extraction of RNA for qRT-PCR experiments requires many of the same considerations as other transcriptomic technologies. To review these considerations please refer to the RNA extraction section of microarray OORF module (Section 3.1.2).

REPORT:

a. Description of instrument(s) used

b. Protocol or kit and any modifications

c. Source of additional reagents

d. Details of DNase or RNase treatment

e. Contamination assessment (DNA or RNA)

3.3.1.2. Quantification and Qualification of RNA

REPORT:

a. Description of instrument(s) used

b. Protocol or kit and any modifications

c. Purity (A_{260}/A_{280}, and optionally A_{260}/A_{230})

d. Yield

e. RNA integrity: method/instrument

f. RIN/RQI or ratio of 3' to 5' transcript Cqs

g. Electrophoresis traces

h. Inhibition testing (Cq dilutions, spike, or other)

3.3.1.3. Reverse Transcription of RNA to cDNA

REPORT:

a. Description of instrument(s) used

b. Protocol or kit and any modifications

c. Amount of RNA and reaction volume

d. Priming oligonucleotide (specify if custom gene-specific primer, random hexamers, or oligo[dT], and percentage of each if mixture) and concentration:

e. Reverse transcriptase and concentration

f. Temperature and time

g. Manufacturer of reagents and catalogue numbers

h. Cqs with and without reverse transcription

i. Storage conditions of cDNA

3.3.2. qRT-PCR Assay Design

This section describes which qRT-PCR assay-specific details should be reported. These include a description of the assay format (single- or multi-gene), technology (double-stranded DNA dye, probe-based, etc.) and protocol employed, information about the target genes and the oligonucleotides used to detect them, and any assay validation data.

3.3.2.1. qRT-PCR Protocol

REPORT:

a. Target detection technology (dye-based, probe-based, other)

b. Protocol or kit and any modifications

c. Number of reactions per sample

d. Number of gene targets per reaction if multiplexing

e. Description of quality control reactions (No template control, PCR positive control, genomic contamination controls, etc.)

f. Reaction volume and amount of cDNA/DNA

g. Primer, (probe), Mg_2^+, and dNTP concentrations

h. Polymerase identity and concentration

i. Buffer/kit identity and manufacturer

j. Additives (SYBR Green I, DMSO, and so forth)

k. Exact chemical composition of any custom buffers or reagents, if applicable

l. Manufacturer of plates/tubes and catalogue number

m. Description of instrument(s) used

n. Complete thermocycling parameters

o. Fluorescence excitation and emission wavelengths used for data collection

p. Other relevant instrument/software settings

q. Reaction setup (manual/robotic)

3.3.2.2. Gene Target Information

REPORT:

a. Target identifier: Entrez ID preferred, or other identifier (Gene symbol or other database ID)

b. Sequence accession number

c. What splice variants are targeted?

3.3.2.3. Oligonucleotides and Amplicon

REPORT:

a. Primer sequences

b. Probe sequences (if applicable)

c. Location and identity of any modifications on oligos

d. Primer/probe database and identification number

e. In silico specificity screen (BLAST, or other)

f. Location of amplicon in target gene

g. Location of each primer (and probe, if applicable) by exon or intron, and indicate if intron-spanning

h. Amplicon length

i. Secondary structure analysis of amplicon

j. Manufacturer of oligonucleotides

k. Purification method

3.3.2.4. Assay Validation

REPORT:

a. Evidence of optimization (from gradients)
b. Specificity (gel, sequence, melt, or digest)
c. Description of target (primer and probes, if applicable) used for no-template controls (NTC)
d. Cq of NTC reactions
e. Calibration curves with slope and y intercept
f. PCR efficiency calculated from slope
g. Confidence intervals for PCR efficiency or standard error (SE)
h. R2 of calibration curve
i. Linear dynamic range
j. Cq variation at limit of detection (LOD)
k. CIs throughout range
l. Evidence for LOD

3.3.3. Data Analysis

This section describes the details that should be reported for the analysis of qRT-PCR data. This includes information about how the Cq values are determined, quality control results, pre-processing of low/high-signal samples, and the data normalisation method employed. The most common method for data normalisation for qRT-PCR experiments is to normalise to a set of so-called *reference genes* (often also referred to as *housekeeping genes*). As such, this reporting framework provides more details on the reference gene normalisation method. However, as multigene qRT-PCR platforms are becoming more common (e.g. 96- and 384-gene PCR arrays), methods originally designed for microarrays are increasingly being applied. In cases where such methods are used, sufficient detail should be provided so that normalisation results can be replicated (see microarray normalisation section 3.1.4).

3.3.3.1. Data Acquisition and Quality Control

REPORT:

a. qRT-PCR analysis program (source, version)
b. Method of Cq determination
c. Results for quality control reactions (No template controls, PCR positive controls, Genomic contamination controls, 3’ to 5’ Cq ratios, etc)

d. Number of technical replicates and how they were produced (e.g. a single RNA sample that was used for qRT-PCR or a single RT product that was split for qPCR)

e. Treatment of technical replicates (average, median, etc.)

f. Number and concordance of biological replicates

g. Repeatability (intraassay variation)

h. Reproducibility (interassay variation, CV)

i. Submission of Cq or raw data in public repository

3.3.3.2. Data Pre-Processing

REPORT:

a. Identification of low/high expression targets/samples (Cq thresholds)

b. Treatment of low/high expression targets/samples (removed, replaced, etc.)

c. Outlier identification and treatment

d. Any transformations applied to data prior to normalisation (if any)

3.3.3.3. Data Normalisation

REPORT:

a. Description of normalisation method (reference gene (RG), or other)

b. RG method: identification of reference genes

c. RG method: justification for selection of reference genes (evidence of stability)

d. RG method: was variable amplification efficiency considered in normalisation calculations, or assumed to be equal across targets

e. RG method: formula used for normalisation

f. For all other methods: Provide sufficient information to reproduce normalisation results

3.3.4. References

Bustin, S.A., Benes, V., et al. (2008) “The MIQE guidelines: Minimum information for publication of quantitative real-time PCR experiments.” Clin Chem. 55(4):611–622.

Lefever, S., Hellemans, J., et al. (2009). “RDML: Structured language and reporting guidelines for real-time quantitative PCR data.” Nucleic Acids Res. 37(7):2065-9.

3.4. Mass Spectrometry Metabolomics

Applications of metabolomics in research utilise a broad array of analytical technologies, typically involving a separation technology (e.g. chromatography) and a detection device (e.g. mass spectrometry). As this reporting framework focuses on the application of metabolomics in regulatory toxicology, only the most mature, stable and proven technologies are considered. Based on two international surveys (Weber et al., 2015, Weber et al., 2017), the most widely applied analytical methods for metabolite analysis include liquid chromatography mass spectrometry (LC-MS; Dunn et al., 2011), gas chromatography mass spectrometry (GC-MS; Beale et al., 2018) and direct infusion mass spectrometry (DIMS; Southam et al., 2017).

These technologies can be broadly categorised as either untargeted, targeted or a hybrid of one or more assay types. In an **untargeted assay**, no analytes are pre-selected for measurement; any metabolites that are detected can be (i) unknown (MSI level 4), annotated (MSI levels 2-3) or identified (MSI level 1) (Sumner et al., 2007), and (ii) are all relatively quantified (i.e. relative between different samples such as treated vs. controls). In practice the majority of features measured in an untargeted assay will typically be either unknown (MSI level 4) or annotated (MSI levels 2-3). In contrast, all analytes are pre-selected for measurement with **targeted assays**. Traditionally this would mean all metabolites that are detected are (1) identified to MSI level 1, and are (2) quantified using a metabolite-specific standard. Traditional assays require metabolite-specific standards acquired in the same laboratory using the same analytical methods where the absolute concentration of metabolites in the biological sample are measured. See FDA guidelines for traditional targeted assays to measure metabolic biomarkers (FDA 2015, FDA 2018). However, depending on the availability and use of reference standards in the assay, targeted assays can also "semi-quantify" metabolites rather than provide absolute quantification. Moreover, targeted assays can also be used to reliably detect and relatively quantify 'analytically known' metabolites, the annotation of which is unclear at the time of measurement. In addition, **hybrid assays** combine components of untargeted and targeted analyte selection into a single assay. Hybrid assays typically have a higher information content than untargeted assays by attempting to measure pre-selected (and often toxicologically important) metabolic biomarkers. Untargeted, targeted and hybrid LC-MS, GC-MS and DIMS assays can all be reported using this OECD framework.

3.4.1. Overall description and rationale for mass spectrometry metabolomics approach

In this **Data Acquisition and Processing Reporting Module** we describe the reporting required for a **mass spectrometry-based metabolomics study**, including sample processing, metabolite extraction and preparation of standards; analytical QA/QC samples; acquisition and processing of mass spectrometry data; demonstrating the quality of the data; and the data matrices produced by this technology, including metabolite annotation/identification and intensities.

REPORT:

1. Overall description and rationale for mass spectrometry metabolomics approach, for example including sample type, extraction method, technology type, processing methods and QA/QC.
2. Technology type(s) used, e.g. LC-MS, LC-MS/MS, GC-MS, GC-MS/MS, DI-MS, DI-MS/MS, or any other mass spectrometry approach used.
3. Link(s) to relevant SOP(s) or publication(s) of approach used.

3.4.2. Sample processing: metabolite extraction and addition of chemical reference standards

While the specific protocols for metabolite extraction will depend on the sample type (i.e. biofluid, cells, tissue, etc.) being measured, a reporting framework that can capture the most important information about a diverse range of methods is presented here.

3.4.2.1. Metabolite extraction from biofluids, cells and tissues

Extraction methods differ for liquid, cellular and tissue sample types, though typically require addition of an organic solvent to denature any proteins present and therefore stop any enzymatic activity that would otherwise change the metabolome.

REPORT:

1. Extraction method general description, including per-sample amounts, solvent(s) used and their ratios, means of agitation/maceration, temperatures and times, and post extraction handling.

3.4.2.2. Derivatisation of samples

Only if GC-MS is used:

REPORT:

1. Derivatisation method general description, including reagents and reaction (including temperature and time) and clean-up/partitioning (if used).

3.4.2.3. Extract concentration and reconstitution in solvent for mass spectrometry analysis

For many sample types, before analysis, the metabolite extract is evaporated to dryness and reconstituted in solvents that (1) facilitate their ionisation, and (2) are compatible with the LC-MS or GC-MS mobile and stationary phases.

REPORT:

1. Drying and reconstitution method general description, including final reconstitution solvent(s) and final volume (if used), storage temperature (if relevant) and duration of reconstituted extracts (if relevant).

3.4.2.4. Chemical reference standard(s) and their preparation

Reference standards can be prepared for a variety of purposes including standards for quality assessment and to correct for variability, standards used for aiding identification/annotation, and standards used for both identification and quantification. Key considerations include if the reference standard was used to annotate, identify and/or quantify an actual metabolite or used to annotate and/or quantify a (set of) chemically related metabolite(s) (e.g. a class of metabolites); if the standard used is a surrogate of the metabolite of interest; also if the reference standard was spiked directly into the biological sample or not.

Internal reference standards for assessing (and correcting) the variability of sample extraction

and/or mass spectrometry detection

Internal standards can correct the experimental variability (e.g. extraction and/or the instrument performance) for each sample. Specifically, if the internal standard is added at the start of the extraction procedure (pre-spiked) it can provide information on the extraction (optionally matched against a post-spiked representative sample extract) - and serve as an 'extraction standard'. If the internal standard is added after the last step of the extraction procedure (typically added to the reconstitution solvent) it can correct instrument performance - and serve as an 'injection standard'. Both an extraction standard and injection standard can be used in the same assay.

Reference standards to aid metabolite annotation/identification **only**

This refers to reference standards prepared to either accurately identify one or more individual metabolites or annotate metabolite classes in a biological sample. These standards are only used for annotation and identification purposes and not used for quantification. All the standards should ideally be acquired on the same instrument type and method as applied to measure the biological samples. The standards can be spiked internally into a biological sample (typically through the spiking of an isotopically-labelled reference material) or spiked into an external alternative matrix (e.g. surrogate or working solvent only)

Existing in-house as well as external (both public and commercial) mass spectral libraries of standards can also be used for annotation purposes. In these cases the full details regarding the reference standard preparation might not be known or available but the source of the mass spectral library and version should be reported in Section 3.4.5.11.

Reference standards to aid metabolite annotation/identification **and** quantification

This refers to reference standards prepared to either identify and quantify one or more metabolites, or annotate and (semi-) quantify one or more metabolite classes. All the standards need to be acquired on the same instrument type and method as the biological samples. To quantify, single or multi-point calibration data is required.

Traditionally, these reference standards have mostly been made 'in-house' by individual laboratories and are largely defined by the specific metabolites or metabolic pathways of interest. More recently, commercial kits containing a panel of reference standards have become available, often in convenient multi-well plate formats. Identical reference standards should be available for all targeted mass spectrometry assays within a study (both intra- and interlaboratory), with mixtures ideally being prepared by/sourced from one laboratory. Furthermore, the ability to report full reference standard content and associated methods for metabolite identification and quantification should be ensured before commencing a project.

Often considered the most reliable approach in metabolomics for calculating absolute abundances using LC-MS or GC-MS is the inclusion of isotopically-labelled standards and the use of SRM (single reaction monitoring) or MRMs (multiple reaction monitoring, i.e. the selected monitoring of multiple product ions from one or more precursor ions) for each metabolite that is measured. *Here, known amounts of an array of isotope-labelled relevant metabolites are spiked into each sample. Absolute quantification can then be achieved by comparison of the peak areas of the labelled and unlabelled versions of each metabolite.* This method also gives an additional level of confidence to metabolite identification. Guidelines for traditional targeted assays for quantification using metabolite standards are available from the FDA (FDA 2015, FDA 2018).

Throughout this Guidance Document the term "absolute quantification" refers only to the process when a reference standard has been used to measure an actual metabolite (or a metabolite surrogate that is very similar to the actual metabolite, e.g. a stable isotope labelled metabolite) and any matrix effects have been accounted for. All other quantification methods using reference standards, described in this Guidance Document, are considered "semi-quantification" and if **no**

reference standards are used only “relative-quantification” can be achieved. See Section 3.4 - Table 1 for a description of the quantification methods using reference standards.

For LC-MS and DIMS, to account for the ‘matrix effect’ when performing quantification, the reference standard typically needs to be internally spiked into the biological sample matrix. For GC-MS, matrix effects are generally less prominent, so the reference standard does not always need to be “internally spiked” to account for such effects but this will depend on the compound being used for the reference standard. Full details on the assessment of the matrix effect should be detailed in Section 3.3.

Section 3.4 - Table 1: Summary of quantification methods based on the type of reference standards.

		How similar is the reference standard to the metabolite of interest?	
		Actual metabolite (including stable isotopically labelled)	A surrogate that is similar to the actual metabolite (e.g. the surrogate is the same metabolite class)
Have matrix effects been accounted for?	Yes	**Absolute quantification**	**Semi quantification**
	No	**Semi quantification**	**Semi quantification**

REPORT:

1. For each new standard (or each named panel of standards) prepared for this study:
    a. Purpose of standard (e.g. extraction (and injection) standard; injection standard only; annotation/identification **only**; annotation/identification **and** quantification)
    b. Quantification type (if relevant; e.g. absolute quantification, semi-quantification or relative quantification - see Table 1)
    c. Calibration methodology (if relevant; e.g. external calibration (without matrix), external matrix-matched calibration, calibration by internal normalisation, standard addition calibration, internal calibration)
    d. Identity (including common name and, for example, PubChem CID, HMDB id, KEGG id, CAS, InChI, InChIKey, SMILES)
    e. Purity (if known)
    f. Supplier
    g. Preparation method, including whether standard spiked internally into the biological sample or alternative external solution
    h. Concentration(s)

3.4.3. Analytical quality assurance and preparation of quality control samples

When conducting a metabolomics study using LC-MS, GC-MS or DIMS it is essential to have a quality assurance (QA) framework and use quality control (QC) samples, as described in the MERIT best practice guidelines (Viant et al., 2019). Here, the reporting of the analytical QA/QC is described for LC-MS, GC-MS and DIMS instrument set up and calibration, and for analysis of a set of biological and QC samples. The QC results are reported in the Section 3.4.6 - *Demonstration of quality of mass spectrometry metabolomics analysis*, below.

3.4.3.1. QC samples

System suitability QC sample (if used)

Used to ensure that LC or GC retention times (if chromatography is used), *m/z* measurements and feature intensities are within specification of the instrument. Should be consistent over long periods and potentially usable across multiple laboratories. A synthetic sample comprising a mixture of metabolite standards or reference material can be used for this purpose.

Intrastudy QC sample

Used to pre-condition the chromatography (if used) and mass spectrometry, and analysed repeatedly throughout an analytical batch to assess, and potentially correct, any drifts in measurement performance. It is essential that the intrastudy QC is highly representative of the biological samples in the study. Typically, this type of QC is derived from a small aliquot of the biological samples within the study.

Intralaboratory QC sample (if used)

Used to assess (and potentially correct for) any differences between separate studies within one laboratory. Should be representative of the biological sample type in the study and hence derived from a one-time pool of multiple extracted samples by a specific laboratory using a defined protocol, or a synthetic sample covering the relevant metabolite space, or a reference material of sufficiently similar metabolic composition to the biological sample type.

Interlaboratory QC sample (if used)

Used to assess (and potentially correct for) any differences between individual laboratories. Should be accessible to multiple laboratories, has known providence, is stable, characterised and available in controlled batch numbers. Ideally this type of QC has a similar metabolic composition or matrix to the biological samples in the study, although this is not always possible, in which case use as close to the same composition as possible.

REPORT:

1. Sample details for system suitability QC sample, intrastudy QC sample, intralaboratory QC sample (if used) and interlaboratory QC sample (if used). Should include details of source, preparation details, batch number (if relevant), certificate of analysis (COA) (if relevant), storage conditions and days since preparation.

3.4.3.2. Process blank sample

Used to provide a measure of background contamination arising from the extraction and LC-MS, GC-MS or DIMS analysis. It is study specific, prepared in the same manner as the biological samples except that no biological material is present. It is important to define the start and end points of the ‘process’ used to prepare this type of QC sample.

REPORT:

1. Type of process blank, start and end points of the 'process' used to prepare this sample, and storage conditions.

3.4.3.3. Assessment of sample matrix effects (optional)

Assessing sample matrix effects is particularly important for quantitative targeted metabolomics, and are mostly related to ESI sources. Matrix effects can be estimated using various methods, for example by post-column infusion for qualitative estimation (Bonfiglio et al., 1999) or by the method described by Matuszewski et al. for quantitative estimation (Matuszewski et al., 2003), after selective sample preparation. The latter approach estimates analyte loss during the extraction step and the signal alteration (ion enhancement or suppression) due to the interfering compounds from the matrix. Additional care must be taken with this phenomenon, especially in quantitative analysis or relative quantification, as mentioned by the FDA, which recommends identifying any matrix effects (FDA 2018).

REPORT:

1. Method to assess matrix effects.

3.4.4. Acquisition of mass spectrometry metabolomics and metabolite data

3.4.4.1. Mass spectrometry assay type(s)

Mass spectrometry metabolomics assays can be untargeted, targeted, or a hybrid. Each assay can have one or more levels of annotation/identification and quantification. Given there are multiple subtypes of these assays, Section 3.4 - Table 2 describes the list of possible metabolomics/metabolite assays. A single toxicology study can comprise a combination of several assays.

REPORT:

1. Mass spectrometry assay type(s) used (see Section 3.4 - Table 2)
2. Mass spectrometry assay type description

Section 3.4 - Table 2: Summary of all relevant mass spectrometry assay types for metabolomics/metabolite analysis.

	Name	**Notes**
Untargeted assays (no features are pre-selected for data acquisition)		
1	Untargeted with relative quantification	No standards measured (for identification or quantification) at time of assay.
Targeted assays (all features are pre-selected for data acquisition)		
2	Targeted with relative quantification only	This includes targeting of 'known unknown' metabolites for which the metabolite identity is unknown.
3	Targeted with semi-quantification only	This includes measuring intensities when using a metabolite class standard, and/or measuring intensities when matrix effects have not been accounted for.
4	Targeted with absolute quantification only	Traditional targeted assay.
5	Targeted with combination of quantification methods	Any combination of the targeted approaches 2-4.
Hybrid assays (some features are pre-selected and some features are not pre-selected for data acquisition)		
6	Hybrid with relative quantification	Combination of untargeted approach (1) and targeted approaches (2).
7	Hybrid with combination of quantification methods	Combination of untargeted approach (1) and targeted approaches (2-5). Need ≥ 1 reference standard used for either semi or absolute quantification

3.4.4.2. Mass spectrometry configuration(s) and method(s)

REPORT:

1. File(s) summarising instrument configuration and method details of analysis performed. See Section 3.4 - Table 3 for suggested reporting elements depending on the instrumentation and method applied.

Section 3.4 - Table 3: Suggested reporting elements for mass spectrometry configuration and methods.

	Suggested reporting elements
LC Configuration	LC configuration name; LC manufacturer; LC model number/name; Software package(s) and version number(s); Column and pre/guard column manufacturer; Column model number/name; Stationary phase composition and particle size; Column internal diameter and length; Injection vials or plate manufacturer and model number
LC Method	LC method name; Mobile phase composition; Mobile phase flow rate; Composition of the wash solvent; Column temperature and pressure; Gradient profile; Amount of sample injected.
GC Configuration	GC configuration name; GC manufacturer; GC model number/name; Software package and version number; Column manufacturer; Column model number/name; Stationary phase composition; Column internal diameter and length; Injection vials manufacturer and model number; Plates manufacturer and model number
GC Method	GC method name; Inlet system (e.g. split/splitless); Inlet temperature; Transfer line temperature; Gas flows and pressure; Temperature gradient; Amount of sample injected
MS Configuration	MS configuration name; MS manufacturer; MS model number/name; Software package(s) and version number(s); Ionisation source; Type of mass analyser (triple quadrupole MS, Orbitrap, time-of-flight, FT-ICR, ion-trap, etc.)
MS Method	MS method name; MS acquisition mode(s) (full scan, MRM, SRM, DDA, DIA, MS/MS, MSn, etc.); Polarity (positive or negative ion analysis); *m/z* scan range; Mass resolution; Collision energies; Isolation width; Lock spray parameters; Source parameters; Gas flows

3.4.4.3. Acquisition order for QC samples, biological samples and reference standard samples

REPORT:

1. Describe how sample acquisition order was defined, e.g. frequency of intrastudy QCs, method to randomise biological sample order.
2. Acquisition order of all types of QC samples, biological samples and reference standard samples (if relevant), thereby indicating the number of pre-conditioning QC samples, the number of process blank samples, and the frequency of analysis of intrastudy QC samples. Columns of reported table should include:
    a. Run order
    b. File name
    c. Sample type (must be able to distinguish between QC samples, biological samples and process blanks)

3.4.5. Processing of mass spectrometry metabolomics and metabolite data

This section covers the reporting of the data processing steps that are required to transform raw data into a form amenable to statistical analysis. This usually requires production of a 2-dimensional data table with each sample represented in one dimension (typically a row) and each metabolic feature (e.g. peak) in the other dimension (typically a column).

The entries in the data table will either be used for relative quantification, absolute quantification or “semi-quantification” of the specific feature in each sample. Each metabolic feature in the data table may also have different levels of metabolite annotation/identification (Sumner et al., 2007).

This section incorporates the reporting of both ‘data processing’ and ‘data post-processing’ as described by the MERIT guidelines (Viant et al., 2019).

3.4.5.1. Data processing workflow(s)

Data processing can be performed across various software and platforms and a single workflow could potentially include both open source (e.g. XCMS (Smith et al., 2006), MS-DIAL (Tsugawa et al., 2015)), proprietary software (e.g. Compound Discoverer, Symphony), dedicated data analysis/processing workflow platforms (e.g. Galaxy (Afgan et al., 2018), KNIME (Berthold et al., 2009), Nextflow (Tommaso et al., 2017)) and/or custom programming script(s).

Sharing sufficient details to be able to re-run the full data processing workflow is encouraged and ultimately improves the reproducibility of the data processing.

REPORT:

1. Data processing workflow description
2. Name(s) and version(s) of the software used
3. Script(s)/workflow file(s) or a reference to a URL or DOI of the script(s)/workflow file(s) (optional)
4. Data processing history(ies) or log(s) (optional)
5. Reference(s) to relevant protocol(s) or publication(s).

3.4.5.2. Centroiding, baseline correction and noise reduction

Mass spectrometry data is typically measured in 'profile mode'. To reduce file sizes considerably the raw data files are often reduced to a form where each feature is represented as an individual *m/z* with zero line width, a process called centroiding. Baseline correction and noise reduction can also be performed.

REPORT:

1. Centroiding, baseline correction and/or noise reduction method(s) and parameters if used.

3.4.5.3. Data reduction

Data reduction involves four steps: feature detection/picking; retention time alignment to take into account shifts in retention time of the same analytes in different samples; grouping/matching of features from the same analyte across different samples (if relevant); and feature integration when estimating the abundance of a metabolite.

REPORT:

1. Data reduction method(s) and parameters used - including (where relevant) details of feature detection/picking, retention time alignment, feature grouping/matching, and feature integration.

3.4.5.4. Feature intensity drift and/or batch correction

While every effort should be made experimentally to minimise variations in mass spectrometry signal intensity within and between analytical batches, these can be (partially) corrected in the data processing step, typically using signals recorded in the intrastudy QC samples.

Assessing within- and/or between-batch signal intensity drift (optional for absolute quantification/semi-quantification)
Evaluation of the presence of such signal intensity drift typically includes a PCA analysis, showing both biological and intrastudy QC samples and relative standard deviation measurements (RSD; also known as coefficient of variation (CV)) for some metabolites present in QC samples (covering a wide range of physicochemical properties and concentrations). If the intrastudy QC samples show significant differences in scores or RSD values within any batch, or a drift in scores values or RSD values between batches, then drift and/or batch effects are present.

Signal intensity batch correction
If used, the effects of batch correction should be demonstrated, for example with a PCA analysis and by inspecting the intensities of representative analytes, both before and after the batch correction.

Signal intensity drift correction within a batch
If used, the effects of drift correction should be demonstrated, for example with a PCA analysis and by inspecting the intensities of representative analytes, both before and after the correction is made.

REPORT:

1. Feature intensity drift and/or batch correction methods and parameters used - including (where relevant) details of signal intensity drift assessment, batch correction, and drift correction within a batch.

3.4.5.5. Identification and removal ('filtering') of features

Particularly untargeted mass spectrometry analysis may detect a significant number of low intensity noisy features and/or systematically biased features, which should be considered for removal to improve the overall reliability of the data set. Several optional filtering processes may be conducted and reported as described here.

Removal of features that are sparsely detected across biological and/or intrastudy QC samples
Features can be removed if they are only present below a defined percentage of biological and/or intrastudy QC samples, or below a defined percentage of a biological sample group.

Filtering features for repeatability
A widely used procedure in untargeted metabolomics is to remove features with high analytical variability in intensity. The RSD of the intensity of each feature can be estimated from the intrastudy QC samples and features with an RSD greater than a threshold are removed.

Filtering features for a proportional response function (i.e. linearity)
Some analytical designs may include a dilution series, where an intrastudy QC is diluted by known factors. Within the instrument's linear range, reliable features should exhibit intensities which correlate very strongly with the known dilution factors. A feasible strategy is therefore to remove features whose intensities do not correlate well to the dilution factors.

Removal of features present in process blanks
Features detected in process blanks are thought to result from the solvents or plasticware rather than the biological sample and therefore can be considered for removal from the final data set.

Removal of features from dosed substance(s)
In toxicology studies, in which the biological system is deliberately exposed to an exogenous chemical, it is common to observe the dosed parent substance and/or its biotransformation products in the resulting data. For the purposes of analysing the endogenous metabolic effect of the exposure, it is important that these signals are removed from the data set. Identification of the relevant features to remove will typically involve comparison to control spectra from non-dosed animals, spectra from a chemical standard, and/or literature. Typically features identified for removal will simply be deleted from the data set.

REPORT:

1. Methods and parameters used to identify and remove features - including (where relevant) details of removal of features that are sparsely detected across biological and/or intrastudy QC samples, filtering features for repeatability, filtering features for a proportional response function (i.e. linearity), removal of features present in process blanks, and removal of features from dosed substance(s).

3.4.5.6. Identification and removal ('filtering') of outlying samples

It is important to identify potential outlying samples and to remove them, if necessary, before applying statistical analysis. Reasons for removing each outlier must be clearly explained in terms of biological, analytical or data analytical aspects; e.g. sample removed due to limited volume, contamination was detected, results of an independent assay indicating abnormality, etc.

Removal of samples with a sparsely detected number of features

If multiple features are missing for a particular sample, above a defined threshold (e.g. 50%), then that sample should be considered for removal. A further option, before applying this method to filter samples, is to apply a univariate outlier analysis to flag outlying values of each feature as missing data. Then the filter is applied to remove any samples with a number of missing values above the defined threshold.

Removal of outlying samples

Multivariate methods are recommended for outlier detection, e.g. PCA using Hotelling's T^2 distribution on the scores and/or F-tests on the residuals.

REPORT:

1. Methods and parameters used to identify and remove samples (if used; see section 3.4.7.1 for actual samples removed).

3.4.5.7. Normalisation

Normalisation is the process of removing technical or otherwise irrelevant variation from the data on a sample by sample basis. Typically, for relative quantification data, the intensity for each sample is multiplied by a scalar factor, which is different for each sample. Normalisation is usually applied to take account of uncontrolled factors such as dilution or overall instrument response.

REPORT:

1. Normalisation method(s) and parameters (if used).

3.4.5.8. Missing value imputation

Missing values occur in mass spectrometry metabolomics datasets for a variety of reasons, such as loss of samples, failure to detect a feature in a given sample, or data processing effects. Their presence can significantly affect the performance of the statistical analysis and thus influence the results of the study.

REPORT:

1. Missing value imputation method(s) and parameters (if used).

3.4.5.9. Normality testing, scaling and/or transformations

These processes are applied to each metabolite feature and are particularly important for multivariate analysis. They are typically performed to allow all features to contribute more evenly to a model, or to bring distributions closer to normality. Normality testing is of particular importance for selecting the appropriate statistical approach to use. Types of scaling and transformations include unit variance, Pareto, log, generalised log, range, level and no scaling/transformation. The appropriate type will depend on the nature of the data.

REPORT:

1. Normality testing, scaling and/or transformation methods and parameters (if used).

3.4.5.10. Processing methods for metabolite quantification

Due to the importance of the type of metabolite quantification used in a regulatory toxicology study, this section reports the methods used even though they may already have been listed in Section 3.4.5.3. Data reduction.

REPORT:

1. Description of method(s) and parameters for metabolite quantification - including (where relevant) for relative- and/or semi- and/or absolute quantification.

3.4.5.11. Processing methods for metabolite annotation and/or identification

Metabolite annotation/identification for mass spectrometry datasets may involve the use of multiple approaches and/or standards, depending upon the nature of the data.

Metabolite identification (i.e. MSI level 1 identification) can only be achieved by matching the retention time and *m/z* value(s) of a reference standard (representing a single metabolite) with the retention time and *m/z* value(s) in a biological sample (representing a single metabolite) - where both the biological sample data and reference standard data were acquired in the same laboratory with the same analytical methods.

Metabolite annotation (i.e. MSI levels 2-3) can be achieved by a plethora of approaches, some of which use libraries of either in house or public reference standards, and other approaches that do not require libraries of standards at all. In cases where no reference standard is available, annotation approaches include searching either the experimental *m/z*, calculated monoisotopic molecular mass or calculated molecular formula of the unknown metabolite feature against public and/or commercial libraries of compounds. If fragmentation spectra have been collected for the feature of interest, features can also be annotated to *in silico* fragments and/or predict a metabolite structure using machine learning approaches. In some cases the annotation will only be to a metabolite class level rather than a single metabolite structure.

REPORT:

1. Description of method(s) and parameters for metabolite annotation/identification - including (if relevant) whether reference standards are from in-house or external spectral library(ies) and the spectral library(ies) used.

3.4.6. Demonstration of quality of mass spectrometry metabolomics analysis

This section covers the reporting of data that will allow the quality of the mass spectrometry metabolomics dataset to be assessed, by the regulator, after the processing described in the section above. Measures of quality are derived from the appropriate uses of different types of QC samples.

3.4.6.1. Mass spectrometer performance report

REPORT:

1. Performance achieved using system suitability QC, relative to a laboratory's acceptance criteria, to confirm instrumentation is fit for purpose. Depending on the assay selected, relevant criteria can include: quantitative reporting of *m/z* shift; retention time shift; shape and/or intensity of selected peaks.

3.4.6.2. Intrastudy QC precision report

REPORT:

1. Measure of a study's analytical precision achieved using intrastudy QC samples, relative to a laboratory's acceptance criteria, to confirm analyses are of sufficient quality for regulatory purposes. To include quantitative reporting of the RSDs of feature intensity (mandatory), *m/z* (optional) and retention time (if relevant; optional), for example reported as the distribution and median of RSD values of all feature intensities across all intrastudy QC samples; and a qualitative assessment of the similarity of intrastudy QCs using PCA, reported as a PCA scores plot from a global analysis of all the biological and intrastudy QC samples.

3.4.6.3. Intralaboratory QC reproducibility report (optional)

REPORT:

1. Measure of intralaboratory (and interstudy) reproducibility, using intralaboratory QC samples, to assess any long term differences between separate studies within the laboratory.

3.4.6.4. Interlaboratory QC reproducibility report (optional)

REPORT:

1. Features (*m/z*, intensities) detected in interlaboratory QC (e.g. defined features in a (standard) reference material). This reporting is likely to evolve as the community improves its use of interlaboratory QC samples.
2. Measure of interlaboratory reproducibility, using interlaboratory QC samples, to assess any differences between separate laboratories; *i.e.* performance standard achieved relative to specified amounts of metabolites.

3.4.7. Outputs: Data matrices from metabolomics assays

The outputs include a data matrix that conveniently summarises the biological samples that were analysed, and whether any of these samples were removed from the study, with a justification.

Reporting the *methods* used for metabolite annotation/identification and for determining metabolite intensities was addressed in Section 3.4.4 (analytical) and Section 3.4.5 (computational) of this reporting module. Here we describe how to report the *results* of those procedures on a feature-by-feature and/or metabolite-by-metabolite basis. In addition, we describe how to report the level of confidence in annotation/identification, utilising the international criteria established by the Metabolomics Standards Initiative (MSI) in 2007 (Sumner et al., 2007) that are currently being reviewed by the Metabolite Identification task group of the International Metabolomics Society. We also describe how to report the level of confidence in the quantification of features/metabolites.

The reporting below should describe data derived from one or more mass spectrometry assays used in the toxicological study.

3.4.7.1. Sample list

REPORT

1. For each biological sample:
    a. Unique biological sample identifier (as an example, a concatenation of selected parameters such as: m(001)p(7d)DG0MOAXY where m = male, (001) = animal number, p = matrix (p=plasma), (7d) = day of sampling after study start, DG0 = dose group (0 = control group), MoAXY = study identifier of study XY);
    b. Order of extraction of biological samples;
    c. Order of mass spectrometric data acquisition;
    d. Whether any biological samples were removed from the study and justification for doing so (e.g. outlier detected using PCA due to limited sample volume)**.**

3.4.7.2. Metabolite annotation/identification

REPORT:

1. For each feature and/or metabolite:
    a. Analytical identifiers (*m/z*, retention time (if relevant), fragmentation data (optional));
    b. Ion form (*i.e.* adduct, isotope) (if known);
    c. Molecular formula(e) of metabolite (if known);
    d. Monoisotopic molecular mass (if known);
    e. Common metabolite name (if known);
    f. Structural code (e.g. standard InChI string or SMILES) (if known);
    g. Metabolite identifier(s) from relevant database(s) (e.g. PubChem, HMDB);
    h. Common metabolite class name (if relevant).
    i. MSI level of identification (levels 1-4; (Sumner et al., 2007))
        i. Level 1 - Identified compound (i.e. confirmed by a reference standard);
        ii. Level 2 - Putatively annotated compound (i.e. without a chemical reference standard, based upon physicochemical properties and/or spectral similarity with public/commercial spectral libraries);
        iii. Level 3 - Putatively characterised compound class (e.g. based upon characteristic physicochemical properties of a chemical class of compounds, or by spectral similarity to known compounds of a chemical class);
        iv. Level 4 - Unknown compound—although unidentified or unclassified these metabolites can still be differentiated and quantified based upon spectral data.
    j. For level 4 only, indicate whether the feature is a ‘known unknown’ or not. Here we define a ‘known unknown’ as meaning a consistently observed feature (i.e. consistent *m/z* and retention time) in the sample matrix under investigation that has been detected repeatedly in one or more laboratories.
    k. *m/z* experimental error from MS1 data (or threshold).
    l. Retention time error (or threshold).
    m. Score for metabolite identification derived from fragmentation data (or threshold).

3.4.7.3. Metabolite quantification

For a number of reasons, including the potential commercial sensitivity associated with untargeted mass spectrometry measurements of the parent substance and its biotransformation products, the mandatory reporting of raw data (i.e. obtained directly from the data acquisition files) is not currently possible.

Furthermore, different statistical analysis methods have differing requirements for the type of processing applied to the metabolomics data. For example, univariate statistical analysis is often applied to the normalised (but not missing value imputed or transformed) data matrix. Benchmark dosing typically requires the imputation of missing values. Multivariate statistical analysis typically requires the data to be normalised, missing value imputed and transformed.

Therefore, multiple data matrices should be reported here.

REPORT:

1. Data matrices of all feature/metabolite intensities, across all remaining samples (biological and intrastudy QCs)
2. For each feature/metabolite
    - Quantification unit (if relevant)
    - Type of quantification (relative quantification, semi-quantification, absolute quantification)
    - RSD of the technical variability of the feature/metabolite intensity derived from repeated measurements of a representative sample, e.g. intrastudy QC sample (optional).

3.4.8. References

Afgan, E., Dannon Baker, Bérénice Batut, Marius van den Beek, Dave Bouvier, Martin Čech, John Chilton, Dave Clements, Nate Coraor, Björn Grüning, Aysam Guerler, Jennifer Hillman-Jackson, Vahid Jalili, Helena Rasche, Nicola Soranzo, Jeremy Goecks, James Taylor, Anton Nekrutenko, and Daniel Blankenberg. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update, Nucleic Acids Research, Volume 46, Issue W1, 2 July 2018, Pages W537–W544, doi:10.1093/nar/gky379

Beale, D.J., Pinu, F.R., Kouremenos, K.A. et al. Review of recent developments in GC–MS approaches to metabolomics-based research. Metabolomics 14, 152 (2018). https://doi.org/10.1007/s11306-018-1449-2

Berthold, M.R., Cebron, N., Dill, F., Gabriel, T.R., Kötter, T., Meinl, T., Ohl, P., Thiel, K. and Wiswedel, B., 2009. KNIME-the Konstanz information miner: version 2.0 and beyond. *AcM SIGKDD explorations Newsletter*, *11*(1), pp.26-31.

Bonfiglio R, King RC, Olah TV, Merkle K. The effects of sample preparation methods on the variability of the electrospray ionization response for model drug compounds. Rapid Commun Mass Spectrom. 1999 Jun;13(12):1175-1185. doi: 10.1002/(SICI)1097-0231(19990630)13:12<1175::AID-RCM639>3.0.CO;2-0. PMID: 10407294.

Di Tommaso, P., Chatzou, M., Floden, E.W., Barja, P.P., Palumbo, E. and Notredame, C., 2017. Nextflow enables reproducible computational workflows. *Nature biotechnology*, *35*(4), pp.316-319.

Dunn, W., Broadhurst, D., Begley, P. et al. Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. Nat Protocol 1060–1083 (2011). https://doi.org/10.1038/nprot.2011.335

FDA (2015) - Analytical Procedures and Methods Validation for Drugs and Biologics - Guidance for Industry. July 2015. U.S. Department of Health and Human Services. Food and Drug Administration. Center for Drug Evaluation and Research (CDER). Center for Biologics Evaluation and Research (CBER). Pharmaceutical Quality/CMC.

FDA (2018) - Bioanalytical Method Validation - Guidance for Industry. May 2018. U.S. Department of Health and Human Services. Food and Drug Administration. Center for Drug Evaluation and Research (CDER). Center for Veterinary Medicine (CVM). Biopharmaceutics.

Matuszewski, B.K., Constanzer, M.L. and Chavez-Eng, C.M., 2003. Strategies for the assessment of matrix effect in quantitative bioanalytical methods based on HPLC− MS/MS. *Analytical chemistry*, *75*(13), pp.3019-3030.

Smith, Colin A., et al. "XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification." *Analytical chemistry* 78.3 (2006): 779-787.

Southam, A., Weber, R., Engel, J. et al. A complete workflow for high-resolution spectral-stitching nanoelectrospray direct-infusion mass-spectrometry-based metabolomics and lipidomics. Nat Protocol 12, 310–328 (2017). https://doi.org/10.1038/nprot.2016.156

Sumner, L.W., Amberg, A., Barrett, D. et al. Proposed minimum reporting standards for chemical analysis. Metabolomics 3, 211–221 (2007). https://doi.org/10.1007/s11306-007-0082-2

Tsugawa, H., Cajka, T., Kind, T., Ma, Y., Higgins, B., Ikeda, K., Kanazawa, M., VanderGheynst, J., Fiehn, O. and Arita, M., 2015. MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis. *Nature methods*, *12*(6), pp.523-526.

Viant, M.R., Ebbels, T.M.D., Beger, R.D. et al. Use cases, best practice and reporting standards for metabolomics in regulatory toxicology. Nat Commun 10, 3041 (2019). https://doi.org/10.1038/s41467-019-10900-y

Weber, R.J., Winder, C.L., Larcombe, L.D., Dunn, W.B. and Viant, M.R., 2015. Training needs in metabolomics. *Metabolomics*, *11*(4), pp.784-786. https://doi.org/10.1007/s11306-015-0815-6

Weber, R.J., Lawson, T.N., Salek, R.M., Ebbels, T., Glen, R.C., Goodacre, R., Griffin, J.L., Haug, K., Koulman, A., Moreno, P. and Ralser, M., 2017. Computational tools and workflows in metabolomics: An international survey highlights the opportunity for harmonisation through Galaxy. Metabolomics 13(2), pp.1-5. https://doi.org/10.1007/s11306-016-1147-x

3.5. NMR Metabolomics

As this OECD reporting framework focuses on the application of metabolomics in regulatory toxicology, only the most mature, stable and proven technologies are considered. Based on two international surveys (Weber et al., 2015, Weber et al., 2017), nuclear magnetic resonance spectroscopy (NMR; Beckonert et al., 2007, Smolinska et al., 2012) is a widely applied, non-destructive and highly precise analytical tool used in metabolomics. This technology is typically applied in an untargeted manner, in which no analytes are pre-selected for measurement; any metabolites that are detected can be unknown (MSI level 4), annotated (MSI levels 2-3) or identified (MSI level 1) (Sumner et al., 2007). NMR spectroscopy can also be used as a targeted assay. Depending on how the NMR data is acquired, metabolites can be either relatively quantified (i.e. relative between different samples such as treated vs. controls) or their absolute concentrations can be measured.

3.5.1. Overall description and rationale for NMR spectroscopy metabolomics approach

In this **Data Acquisition and Processing Reporting Module** we describe the reporting required for an **NMR-based metabolomics study**, including sample processing, metabolite extraction and addition of standards; analytical QA/QC; acquisition and processing of NMR data; demonstrating the quality of the data; and the data matrices produced by this technology, including metabolite annotation/identification and intensities.

REPORT:

1. Overall description and rationale for NMR spectroscopy metabolomics approach, for example including sample type, extraction method, technology type, processing methods and QA/QC.
2. Technology type(s) used (e.g. “1D ^{1}H NMR”, “2D ^{1}H JRES”, “2D ^{1}H–^{1}H TOCSY”, “2D ^{1}H–^{13}C HSQC NMR”, or any other NMR approach used).
3. Link(s) to relevant SOP(s) or publication(s) of approach used (optional).

3.5.2. Sample processing: metabolite extraction and addition of chemical reference standards

While the specific protocols for metabolite extraction will depend on the sample type (i.e. biofluid, cells, tissue, etc.) being measured, a reporting framework that can capture the most important information about a diverse range of methods is presented here.

3.5.2.1. Metabolite extraction from biofluids, cells and tissues

Extraction methods differ for liquid (e.g. urine), cellular and tissue sample types, though typically require addition of an organic solvent to denature any proteins present and therefore stop any enzymatic activity that would otherwise change the metabolome.

REPORT:

1. Extraction method general description, including per-sample amounts, solvent(s) used and their ratios, means of agitation/maceration, temperatures and times, and post extraction handling.

3.5.2.2. Extract concentration and reconstitution in solvent for NMR analysis

For many sample types, before analysis, the metabolite extract is evaporated to dryness and reconstituted in (1) deuterated versions of the relevant solvent (e.g. deuterated water), and (2) if necessary, provided with a pH buffer to minimise inter-sample and interstudy inconsistencies in chemical shift values.

REPORT:

1. Drying and reconstitution method general description, including final reconstitution solvent(s) and final volume (if used), pH buffer type and concentration (if used), storage temperature (if relevant) and duration of reconstituted extracts (if relevant).

3.5.2.3. Chemical reference standard(s) and their preparation

Reference standards can be added for a variety of purposes including standards for quality assessment, metabolite standards used for annotation/identification purposes only, metabolite standards used for annotation/identification and quantification purposes, and a chemical shift reference standard to facilitate metabolite annotation/identification and optionally for quantification.

Internal reference standards for assessing the quality of sample extraction (quality assessment reference standards)
Internal standards can provide information on the extraction efficiency for each sample. The internal standard is added at the start of the extraction procedure (pre-spiked) - and serves as an 'extraction standard'.

Reference standards to aid metabolite annotation/identification **only**
This refers to metabolite reference standards prepared to accurately identify one or more individual metabolites in a biological sample. These standards are only used for annotation/identification purposes and not used for quantification. All the standards should ideally be acquired on the same instrument type and method as applied to measure the biological samples. The standards can be internal to the biological sample (internal reference standards). Otherwise, the reference standard can be spiked into alternative solutions (e.g. buffer only) - these standards are referred to here as external reference standards.

Existing in-house as well as external (both public and commercial) NMR spectral libraries of standards can also be used for annotation/identification purposes. In these cases the full details regarding the reference standard preparation might not be known or available, but the source of the NMR spectral library and version should be reported in Section 3.5.5.10.

Reference standards to aid metabolite annotation/identification **and** quantification
This refers to reference standards prepared to identify and quantify one or more metabolites. All of the standards should be acquired using the same instrument type and method as the biological samples.

Chemical shift reference standard
An internal chemical shift reference standard is required for spectral alignment and identification of NMR resonances. It is also useful for assessing spectral quality, and may be used for the quantification of metabolites.

REPORT:

1. For each new standard (or each named panel of standards) prepared for this study:
    a. Purpose of standard (e.g. extraction standard; annotation/identification **only**; annotation/identification **and** quantification; chemical shift reference standard)
    b. Quantification type (if relevant; e.g. absolute quantification or relative quantification)
    c. Calibration methodology (if relevant; e.g. external calibration (without matrix), external matrix-matched calibration, calibration by internal normalisation, standard addition calibration, internal calibration)
    d. Identity (including common name and, for example, PubChem CID, HMDB id, KEGG id, CAS, InChI, InChIKey, SMILES)
    e. Purity (if known)
    f. Supplier
    g. Preparation method (including whether standard spiked internally into the biological sample or alternative external solution)
    h. Concentration(s)

3.5.3. Analytical quality assurance and preparation of quality control samples

When conducting an untargeted metabolomics study using NMR, it is essential to have a quality assurance (QA) framework and use quality control (QC) samples, as described in the MERIT best practice guidelines (Viant et al., 2019). Here, the reporting of the analytical QA/QC is described for NMR instrument set up and calibration, and for NMR analysis of a set of biological and QC samples. The QC results are reported in Section 3.5.6 - *Demonstration of quality of NMR spectroscopy metabolomics analysis*.

3.5.3.1. QC samples

System suitability QC sample to assess NMR instrument calibration
For a regulatory toxicology study, a system suitability QC must be used to ensure that sample temperature is properly calibrated, and that adequate water suppression has been achieved (for aqueous samples). Temperature calibration (using deuterated methanol (MeOD)) and water suppression (using a sucrose solution) should be conducted using, for example, the methods described by (Dona et al., 2014).

Intrastudy QC sample

Used to provide measures of intrastudy precision and monitor, assess and potentially correct for systematic errors in measurements (e.g. drift in chemical shift, baseline fluctuations, shimming problems). It is essential that the intrastudy QC is highly representative of the biological samples in the study. Typically this type of QC is made by pooling small aliquots of all biological samples within the study.

Intralaboratory QC sample (optional)

Used to assess (and potentially correct for) any differences between separate studies within one laboratory. Should be representative of the biological sample type in the study and hence derived from a one-time pool of multiple extracted samples by a specific laboratory using a defined

protocol, or a synthetic sample covering the relevant metabolite space, or a reference material of sufficiently similar metabolic composition to the biological sample type.

Interlaboratory QC sample (optional)

Used to assess (and potentially correct for) any differences between individual laboratories. Should be accessible to multiple laboratories, has known providence, is stable, characterised and available in controlled batch numbers. Ideally this type of QC has a similar metabolic composition or matrix to the biological samples in the study.

REPORT:

1. Sample details for system suitability QC sample, intrastudy QC sample, intralaboratory QC sample (if used) and interlaboratory QC sample (if used). Should include details of source, preparation details, batch number (if relevant), certificate of analysis (COA) (if relevant), storage conditions and days since preparation.

3.5.3.2. Process blank sample(s)

There are two types of process blank that are commonly used in NMR metabolomics. The first is used to provide a measure of any background contamination arising from the reconstitution solvent/buffer itself, hence the 'process' includes making the solvent/buffer, adding it to an NMR tube, and the NMR analysis. This is often termed a 'buffer blank'. The second process blank is used to provide a measure of background contamination arising from the extraction of the biological sample as well as from the NMR solvent/buffer, hence the process includes extraction, making the solvent/buffer, adding that solution to the extracted sample, and the NMR analysis. This is often termed an 'extraction blank'.

REPORT:

1. Type of process blank(s), start and end points of the 'process' used to prepare this sample, and storage conditions.

3.5.4. Acquisition of NMR metabolomics and metabolite data

Despite shortcomings in sensitivity, NMR metabolomics can reliably detect a variety of metabolites with outstanding precision and robustness. Furthermore, NMR is capable of detecting a considerable range of nuclei (^{1}H, ^{13}C, ^{15}N, etc.) and collecting valuable chemical structural information using two dimensional (2D) experiments (both homonuclear and heteronuclear). However, NMR spectra collected from one dimensional (1D) ^{1}H-NMR experiments are the primary source of metabolomics and metabolite data. Thus, the reporting requirements detailed in this document will be restricted to instrument configuration and calibration, and the subsequent data acquisition and processing associated with 1D ^{1}H-NMR spectra (Viant et al., 2019).

3.5.4.1. NMR spectroscopy assay type(s)

NMR spectroscopy metabolomics assays can be untargeted (i.e. the traditional approach in which no particular metabolites are pre-selected for study) or targeted (e.g. Bruker B.I.Methods™ such as B.I.QUANT-UR that targets and quantifies up to 150 metabolites). Each assay can have one or more levels of metabolite annotation/identification and quantification. A single toxicology study can comprise a combination of several assays.

REPORT:

1. NMR spectroscopy assay type(s) used
2. NMR spectroscopy assay description

3.5.4.2. NMR instrument configuration(s) and method(s)

REPORT:

1. File(s) summarising instrument configuration and method details of analysis performed. See Section 3.5 - Table 1 for suggested reporting elements depending on the instrumentation and method applied.

Section 3.5 - Table 1: Suggested reporting elements for NMR spectroscopy configuration and methods.

	Suggested reporting elements
NMR Configuration	NMR configuration name; NMR magnet manufacturer; NMR magnet model number/name; Magnetic field strength as proton NMR frequency; NMR console manufacturer; NMR console model number/name; Probe type (e.g. 5 mm RT probe); Probe manufacturer; Probe model number/name; Software packages and version number(s)
NMR Method	NMR method name; Sample temperature; Pulse sequence type (e.g. 1D ^{1}H NOESY presat); Sweep width (ppm); Pre-delay (i.e. relaxation delay) (seconds); Carrier frequency (MHz); 90-degree pulse width (microseconds); Number of accumulated scans; Acquisition time (seconds); Confirm sample spinning not used

3.5.4.3. Acquisition order for QC samples, biological samples and reference standard samples

REPORT:

1. Describe how sample acquisition order was defined, e.g. frequency of intrastudy QCs, method to randomise biological sample order.
2. Acquisition order of all types of QC samples, biological samples and reference standard samples (if relevant), thereby indicating the number of process blank samples, and the frequency of analysis of intrastudy QC samples, etc. Columns of reported table should include:

a. Run order
b. File name
c. Sample type (must be able to distinguish between QC samples, biological samples and process blanks)

3.5.5. Processing of NMR metabolomics and metabolite data

This section covers the reporting of established processes for converting raw NMR data acquired on an instrument to processed NMR data that is in a form amenable for statistical analysis. This processing typically produces a 2-dimensional (2D) data table with each sample represented in one dimension (typically a row) and its resonance intensity values in the other dimension (typically a column). For a traditional untargeted NMR metabolomics study (for which no metabolites were pre-selected for analysis and the NMR pulse sequence uses a short relaxation delay), the entries in the data table typically indicate the relative quantities of the metabolites in each sample. This can be in the form of 'binned' (or 'bucketed') data, for which there can be several bins per metabolite, or in the form of one relative quantity per metabolite, depending on how the data is processed. For a targeted NMR metabolomics study (targeting a defined list of metabolites), the entries in the data table typically describe the absolute concentrations of the metabolites. In general, only 1D ^{1}H-NMR data are acquired on each sample, with 2D data acquired primarily for metabolite identification purposes (with the potential exception of 2D *J*-resolved NMR spectroscopy which is gaining in popularity as a high throughput NMR metabolomics method (Ludwig and Viant 2010). Hence, here we focus on reporting for 1D ^{1}H-NMR data processing, based largely on the MERIT guidelines (Viant et al., 2019).

3.5.5.1 Data processing workflow(s)

Data processing can be performed across various software and platforms and a single workflow could potentially include both open source, proprietary software, dedicated data analysis/processing workflow platforms (e.g. Galaxy, KNIME, Nextflow) and/or custom programming script(s).

Sharing sufficient details to be able to re-run the full data processing workflow is encouraged and ultimately improves the reproducibility of the data processing.

REPORT:

1. Data processing workflow description
2. Name(s) and version(s) of the software used
3. Script(s)/workflow file(s) or a reference to a URL or DOI of the script(s)/workflow file(s) (optional)
4. Data processing history(ies) or log(s) (optional)
5. Reference(s) to relevant protocol(s) or publication(s).

3.5.5.2. Spectral pre-processing

In comparison to mass spectrometry metabolomics datasets, the pre-processing steps for NMR are relatively straightforward. At a minimum, these include application of an apodisation function prior to Fourier transformation, phase and baseline correction, and chemical shift calibration. These steps are essential for the pre-processing of raw NMR spectra prior to subsequent metabolomic data analyses and thus the parameters used should be reported as part of any

NMR-based metabolomics study. Other aspects of pre-processing such as zero-filling and linear prediction are occasionally employed and should be reported when used.

REPORT:

1. Spectral pre-processing method(s) and parameters used - including (where relevant) forward and/or backward linear prediction and the number of points for each, window function type and magnitude used for apodisation, zero-filling original and final points count, phasing method (manual or automatic) and parameters, baseline correction method and parameters, and chemical shift calibration method (manual or automatic).

3.5.5.3. Data reduction

This section covers the process of converting pre-processed spectral data into a tabular form for statistical analysis. Steps to report in this process include: (1) peak alignment and matching, followed by either (2) use of full resolution NMR spectra, (3) binning, and/or (4) peak fitting.

REPORT:

1. Data reduction method(s) and parameters used - including (where relevant) peak alignment and matching; whether data reduction was applied (i.e. applying binning, and reporting the bin start, bin end and bin width) or full resolution NMR spectra were used; and peak fitting.

3.5.5.4. Resonance intensity drift and/or batch correction (optional)

While every effort should be made experimentally to minimise variations in signal intensity within and between batches, these can be (partially) corrected in the data processing step, typically using signals recorded in the intrastudy QC samples.

Assessing within- and/or between-batch signal intensity drift

Evaluation of the presence of such signal intensity drift typically includes a PCA analysis, showing both biological and intrastudy QC samples and relative standard deviation measurements (RSD) for some metabolites present in QC samples (covering a wide range of physicochemical properties and concentrations). If the intrastudy QC samples show significant differences in scores or RSD values within any batch, or a drift in scores values or RSD values between batches, then drift and/or batch effects are present.

Signal intensity batch correction

If used, the effects of batch correction should be demonstrated, for example with a PCA analysis and by inspecting the intensities of representative analytes, both before and after the batch correction.

Signal intensity drift correction within a batch

If used, the effects of drift correction should be demonstrated, for example with a PCA analysis and by inspecting the intensities of representative analytes, both before and after the correction is made.

REPORT:

1. Resonance intensity drift and/or batch correction methods and parameters used - including (where relevant) details of signal intensity drift assessment, batch correction, and drift correction within a batch.

3.5.5.5. Identification and removal ('filtering') of NMR resonances

Untargeted NMR analysis may detect certain resonances that should be considered for removal to improve the overall reliability of the data set.

Removal of resonances present in process blanks

Resonances detected in process blanks are thought to result from the solvents or plasticware rather than the biological sample and therefore can be considered for removal from the final data set.

Removal of resonances from dosed substances

In toxicology studies, in which the biological system is deliberately exposed to an exogenous chemical, the dosed parent substance and/or its biotransformation products may be observed in the resulting data. For the purposes of determining the endogenous metabolic effect of the exposure, it is important that these resonances are removed from the data set. The appropriate resonances for removal are typically determined by comparison to control spectra, spectra from a chemical standard, and/or the literature.

REPORT:

1. Methods and parameters used to identify and remove unwanted resonances - including (where relevant) details of removal of resonances present in process blanks, and removal of resonances from dosed substance(s).

3.5.5.6. Identification and removal ('filtering') of outlying samples

It is important to identify potential outlying samples and to remove them, if necessary, before applying statistical analysis. Reasons for removing each outlier must be clearly explained in terms of biological, analytical or data analytical aspects; e.g. sample removed due to limited volume, contamination was detected, results of an independent assay indicating abnormality, etc.

Removal of samples with a sparsely detected number of resonances

If multiple resonances are missing for a particular sample, above a defined threshold (e.g. 50%), then that sample may be considered for removal. A further option, before applying this method to filter samples, is to apply a univariate outlier analysis to flag outlying values of each resonance as missing data. Then the filter is applied to remove any samples with a number of missing values above the defined threshold.

Removal of outlying samples

Multivariate methods are recommended for outlier detection, e.g. PCA using Hotelling's T^2 distribution on the scores and/or F-tests on the residuals.

REPORT:

1. Methods and parameters used to identify and remove samples (if used; see section 3.5.7.1 for actual samples removed).

3.5.5.7. Normalisation

Normalisation is the process of removing technical or otherwise irrelevant variation from the data on a sample by sample basis. Typically, intensity data for each sample is multiplied by a scalar factor, which is different for each sample. Normalisation is usually applied to take account of uncontrolled factors such as dilution or overall instrument response.

REPORT:

1. Normalisation method and parameters (if used).

3.5.5.8. Normality testing, scaling and/or transformations

These processes are applied to each metabolite resonance and are particularly important for multivariate analysis. They are typically performed to allow all resonances to contribute more evenly to a model, or to bring distributions closer to normality. Normality testing is of particular importance for selecting the appropriate statistical approach to use. Types of scaling and transformations include unit variance, Pareto, log, generalised log, range, level and no scaling/transformation. The appropriate type will depend on the nature of the data.

REPORT:

1. Normality testing, scaling and/or transformation methods and parameters (if used).

3.5.5.9. Processing methods for metabolite quantification

Due to the importance of the type of metabolite quantification used in a regulatory toxicology study, this section reports the methods used even though they may already have been listed in Section 3.5.5.3. *Data reduction*.

REPORT:

1. Description of method(s) and parameters for metabolite quantification - including (where relevant) for relative and/or absolute quantification.

3.5.5.10. Processing methods for metabolite annotation and/or identification

Metabolite annotation/identification for NMR datasets may involve the use of multiple data types (e.g. 1D NMR spectra of reference compounds, 2D homonuclear and/or heteronuclear NMR spectra, etc.). As a result, the processing method(s) used to facilitate metabolite annotation/identification will depend upon the nature of the NMR data.

Method of annotating/identifying metabolite resonances when a metabolite reference standard is used

Metabolite identification based on the match of chemical shift value(s) and relative peak intensities of a reference standard with those of the biological sample.

Method of annotating metabolite resonances when no metabolite reference standard is available
NMR spectroscopy is a powerful analytical tool for the *de novo* determination of the structure of small molecule metabolites. While this requires specialist expertise, it is a viable strategy for the reporting of an NMR metabolomics study.

REPORT:

1. General description of method(s) and parameters for metabolite annotation/identification - including the type of NMR data (e.g. 1D, 2D, etc.) collected and the processing parameters, the use of commercial and/or in-house software, and whether reference standards are from external or in-house spectral library(ies), and the spectral library(ies) used.

3.5.6. Demonstration of quality of NMR spectroscopy metabolomics analysis

This section covers the reporting of data that will allow the quality of the NMR metabolomics dataset to be assessed by the regulator. Measures of quality are derived from the appropriate uses of different types of QC samples.

3.5.6.1. NMR spectrometer performance report

REPORT:

1. Performance achieved using system suitability QC, relative to a laboratory's acceptance criteria, to confirm instrumentation is fit-for-purpose. To include quantitative reporting of chemical shift reference peak full width at half maximum height (FWHM) without window function applied, 90 degree pulse width, and probe temperature variation (plus or minus value in ^{o}C). Reporting an image(s) of aligned and stacked spectra for visual inspection is recommended to facilitate an assessment of the general quality of all spectra.

3.5.6.2. Intrastudy QC precision report

REPORT:

1. Measure of a study's analytical precision achieved using intrastudy QC samples, relative to a laboratory's acceptance criteria, to confirm analyses are of sufficient quality for regulatory purposes. To include quantitative reporting of the RSDs of feature intensity (mandatory), for example reported as the distribution and median of RSD values of all feature intensities across all intrastudy QC samples; and a qualitative assessment of the similarity of intrastudy QCs using PCA, reported as a PCA scores plot from a global analysis of all the biological and intrastudy QC samples.

3.5.6.3. Intralaboratory QC reproducibility report (optional)

REPORT:

1. Measure of intralaboratory (and interstudy) reproducibility, using intralaboratory QC samples, to assess any long-term differences between separate studies within the laboratory.

3.5.6.4. Interlaboratory QC reproducibility report (optional)

REPORT:

1. Resonances detected in interlaboratory QC (e.g. defined resonance in a (standard) reference material). This reporting is likely to evolve as the community improves its use of interlaboratory QC samples.
2. Measure of interlaboratory reproducibility (using interlaboratory QC samples) to assess any differences between separate laboratories; *i.e.* performance standard achieved relative to specified amounts of metabolites.

3.5.7. Outputs: Data matrices from metabolomics assays

The outputs include a data matrix that conveniently summarises the biological samples that were analysed, and whether any of these samples were removed from the study, with a justification.

Reporting the *methods* used for metabolite annotation/identification and for determining metabolite intensities was addressed in Section 3.5.4 (analytical) and Section 3.5.5 (computational) of this reporting module. Here we describe how to report the *results* of those procedures on a resonance-by-resonance and/or metabolite-by-metabolite basis. In addition, we describe how to report the level of confidence in annotation/identification, utilising the international criteria established by the Metabolomics Standards Initiative (MSI) in 2007 (Sumner et al 2007) that are currently being reviewed by the Metabolite Identification task group of the International Metabolomics Society. We also describe how to report the level of confidence in the quantification of resonances/metabolites.

The reporting below should describe data derived from one or more NMR spectroscopy assays used in the toxicological study.

3.5.7.1. Sample list

REPORT:

1. For each biological sample:
    a. Unique biological sample identifier (as an example, a concatenation of selected parameters such as: m(001)p(7d)DG0MOAXY where m = male, (001) = animal number, p = matrix (p=plasma), (7d) = day of sampling after study start, DG0 = dose group (0 = control group), MoAXY = study identifier of study XY);
    b. Order of extraction of biological samples;
    c. Order of NMR data acquisition;
    d. Whether any biological samples were removed from the study and justification for doing so (e.g. outlier detected using PCA due to limited sample volume)**.**

3.5.7.2. Metabolite annotation/identification

REPORT:

1. For each resonance and/or metabolite:
    a. Chemical shift value(s) (ppm) for all relevant nuclei;
    b. Multiplicity/splitting pattern (e.g. doublet) for each relevant ^{1}H peak;
    c. Molecular formula (if known);
    d. Common metabolite name (if known);
    e. Structural code (e.g. standard InChI string or SMILES) (if known);
    f. Metabolite identifier(s) from relevant database(s) (e.g. PubChem, HMDB);
    g. Common metabolite class name (if relevant).
    h. MSI level of identification (levels 1-4; (Sumner et al 2007))
        i. Level 1 - Identified compound (i.e. confirmed by a reference standard);
        ii. Level 2 - Putatively annotated compound (i.e. without a chemical reference standard, based upon physicochemical properties and/or spectral similarity with public/commercial spectral libraries);

iii. Level 3 - Putatively characterised compound class (i.e. based upon characteristic physicochemical properties of a chemical class of compounds, or by spectral similarity to known compounds of a chemical class);
iv. Level 4 - Unknown compound—although unidentified or unclassified these metabolites can still be differentiated and quantified based upon spectral data.

i. For level 4 only, indicate whether the resonance is a 'known unknown' or not. Here we define a 'known unknown' as a resonance in the sample matrix under investigation that has been detected repeatedly in one or more laboratories.

3.5.7.3. Metabolite quantification

For a number of reasons, including the potential commercial sensitivity associated with untargeted NMR measurements of the parent substance and its biotransformation products, the mandatory reporting of raw data (i.e. obtained directly from the data acquisition files) is not currently possible.

Furthermore, different statistical analysis methods have differing requirements for the type of processing applied to the metabolomics data. For example, univariate statistical analysis and benchmark dosing are often applied to the normalised (but not transformed) data matrix. Multivariate statistical analysis typically requires the data to be normalised and transformed.

Therefore, multiple data matrices should be reported here.

REPORT:

1. Data matrices of all resonance/metabolite intensities, across all remaining (i.e. not removed as outliers) biological and intrastudy QCs samples.
2. For each resonance/metabolite
   - Quantification unit (if relevant)
   - Type of quantification (relative quantification, absolute quantification)
   - RSD of the technical variability of the resonance/metabolite intensity derived from repeated measurements of a representative sample, e.g. intrastudy QC sample (optional).

3.5.8. References

Beckonert, O., Keun, H., Ebbels, T. et al. Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. Nat Protoc 2, 2692–2703 (2007). https://doi.org/10.1038/nprot.2007.376

Dona, A.C., Jiménez, B., Schäfer, H., Humpfer, E., Spraul, M., Lewis, M.R., Pearce, J.T., Holmes, E., Lindon, J.C. and Nicholson, J.K., 2014. Precision high-throughput proton NMR spectroscopy of human urine, serum, and plasma for large-scale metabolic phenotyping. Analytical chemistry, 86(19), pp.9887-9894.

Ludwig, C. and Viant, M.R., 2010. Two-dimensional J-resolved NMR spectroscopy: review of a key methodology in the metabolomics toolbox. Phytochemical Analysis: An International Journal of Plant Chemical and Biochemical Techniques, 21(1), pp.22-32.

Smolinska, A., Blanchet, L., Buydens, L.M. and Wijmenga, S.S., 2012. NMR and pattern recognition methods in metabolomics: from data acquisition to biomarker discovery: a review.

Analytica chimica acta, 750, pp.82-97.

Sumner, L.W., Amberg, A., Barrett, D. et al. Proposed minimum reporting standards for chemical analysis. Metabolomics 3, 211–221 (2007). https://doi.org/10.1007/s11306-007-0082-2

Viant, M.R., Ebbels, T.M.D., Beger, R.D. et al. Use cases, best practice and reporting standards for metabolomics in regulatory toxicology. Nat Commun 10, 3041 (2019). https://doi.org/10.1038/s41467-019-10900-y

Weber, R.J., Winder, C.L., Larcombe, L.D., Dunn, W.B. and Viant, M.R., 2015. Training needs in metabolomics. *Metabolomics*, *11*(4), pp.784-786. https://doi.org/10.1007/s11306-015-0815-6

Weber, R.J., Lawson, T.N., Salek, R.M., Ebbels, T., Glen, R.C., Goodacre, R., Griffin, J.L., Haug, K., Koulman, A., Moreno, P. and Ralser, M., 2017. Computational tools and workflows in metabolomics: An international survey highlights the opportunity for harmonisation through Galaxy. Metabolomics, 13(2), pp.1-5. https://doi.org/10.1007/s11306-016-1147-x

4. Data Analysis Reporting Modules

Statistical analysis will depend heavily on the objective and design of the study. For example, a single treatment versus control may only require a pairwise analysis, whereas an experiment to derive molecular points-of-departure requires dose-response modelling. While there is no universally acceptable best practice, some statistical analyses are widely applicable to many study designs, for example univariate and multivariate analyses. An optimal data analysis framework has been published for transcriptomics methods that is not prescriptive for the way that data should be processed but does provide a reference against with other analyses can be compared (Verheijen et al., 2022). This section of the OORF describes the following reporting modules:

- Section 4.1 - Differentially Abundant Molecules (using univariate methods)
- Section 4.2 - Multivariate Statistical Analysis
- Section 4.3 - Benchmark Dose Analysis

4.1. Differentially Abundant Molecules (using univariate methods) Module

This multi-omics compliant module specifies the information needed by a regulatory scientist in order to assess univariate statistical analysis as used to discover differentially abundant molecules (DAMs); i.e. for transcriptomics, to identify differentially expressed genes (DEG). Henceforth DEG are referred to as DAM.

The goals of this module are to specify what needs to be reported while not being prescriptive about the analysis performed. We leave judgement about the appropriateness of the analysis plan to the users.

4.1.1. Inputs

Input for identifying DAMs, in the form of processed omics data (e.g. normalised gene expression data, normalised metabolome data) is generated using processes described in a DAPRM. The user should indicate what file(s) / data are used as the input for the DAM analysis. Metadata files containing information associating omics samples with experimental parameters (e.g. treatments, dose levels, time points, sex, etc.) are often also necessary for conducting an DAM analysis. The user should specify these files as well.

REPORT:

4.1.1.1. Data Input

Provide the name of the input data object(s)

4.1.1.2. Metadata Input

Provide the name of the input metadata object(s)

4.1.2. Software Documentation

The requirements in this section are intended to facilitate reproducibility of analyses and make quality assessment possible. Freely available open source, commercially available, or proprietary software may be used. The name and version of all software that is used to identify DAMs must be specified. In addition, all add-ons, plug-ins, packages, or libraries (hereafter “libraries”) that are called, used, required, loaded, or otherwise brought into the software for use in identifying DAMs must be specified. This specification must include the name and version of such libraries, and a hyperlink to download the libraries. In those cases where the libraries are locally developed and controlled, such that a hyperlink to download does not exist, specify the following: “this library is not available for download” and should then make the code available as specified in the reporting template.

REPORT:

4.1.2.1. Software

The name of the software/analysis package along with the version.

4.1.2.2. Operating System

In some cases, modelling of data can be impacted by the operating system that the software is run on, therefore it is important to document this information.

4.1.2.3. Additional Libraries Used

A list of libraries used for analysis along with the version for each.

4.1.2.4. Software Availability

If the software is open source, a hyperlink to the software and source code (if available) is desirable.

4.1.3. Contrasts for Which Differentially Abundant Molecules were Identified

To ensure clarity, the contrasts (or group comparisons) that were performed to identify the DAMs must be detailed. Generally, for simple control vs treated study designs, the contrast of interest is treated vs control. This can become more complicated when the number of treatment levels increase, or if a time-course design is being used. In cases where the term "control" can be confusing, a more specific term must be used. For instance, if the study design uses a laboratory control and a field control then which one is being used should be clearly specified for each contrast, or a description of some transformation based on these controls provided. For instance, "treated time 12 h vs field control 12 h" is more clear than stating "treated time 12 h vs control" in the case where control could refer to laboratory control or field control. For the purpose of clarity, the *factor* (alternatively called the independent variable) is an explanatory variable manipulated by the experimenter. Each factor has two or more *levels* (i.e. different values of the factor). Combinations of factor levels are often called treatments.

REPORT:

4.1.3.1. Contrasts

A table or listing that must include each factor and level within each factor that is being considered in the DAM analysis.

4.1.4. Assay Experimental Design

In order to ensure that the omics data were analysed properly, the end-user must have a clear description of which samples were used in the DAM analysis, how many samples there are per experimental group, and how they may be assayed together/measured simultaneously or not. In addition, the end-user can use this information to perform post hoc power analyses, as well as estimates for Type M (magnitude) and Type S (sign) error, which are equally, if not more, informative in assessing data quality and the likelihood of drawing erroneous conclusions. For more on Type M and Type S errors, see Gelman and Carlin (2014).

Report the relevant information in the reporting template as detailed below:

REPORT:

4.1.4.1. Group Sizes

A table or listing of the number of samples used for identifying DAMs in each group or factor at a minimum. Should include the sample IDs that can be used to match across the report.

4.1.4.2. Covariance

Identification of which samples may exhibit covariance (examples include: assayed on the same day, assayed in parallel, assayed on the same physical platform (e.g. chip/glass slide, cartridge, etc. for transcriptomics; LC-MS analytical batch, etc. for metabolomics), processed using the same wash solutions or reagents, processed using the same master mixes (transcriptomics only), littermates in the exposure study, animals which are housed together, etc.). Provide an explicit statement if there is no reason to believe covariance exists between samples, and a justification to support this assertion.

4.1.4.2. Technical Replicates

Identify any samples that are technical replicates of each other, and how those samples are technical replicates (i.e. define what makes the samples technical replicates).

4.1.5. Statistical Analysis to Identify Differentially Abundant Molecules

Today there are a myriad of ways for identifying DAMs, all with their own strengths and weaknesses. It is also recognized that each experimental design may require its own set of assumptions and caveats in the analysis that makes it difficult to prescribe a universal standard. Thus, it is critical to clearly communicate how the statistical analysis was performed so that the end-user can understand what was done and establish if the approach taken was sound. To accomplish this, the statistical analysis plan must be supplied. The following specifies the minimum requirements for a sound statistical analysis plan.

REPORT:

4.1.5.1. Statistical Approach

The name and description of statistical approach

4.1.5.2. Data Transformation

Clearly state all data transformations (e.g. $\log_2$ transformation) that are performed in the course of analysis. Alternatively, indicate if no data transformations were performed to ensure clarity.

4.1.5.3. Effects Model

If using a general linear model, general linear mixed model, ANOVA, or something similar, specify the effects being modelled including all

fixed and random effects, as well as their interactions, if any, and any nesting.

4.1.5.4. Modelling Input

If using a pairwise comparison approach, such as a Wald's test, Student's or Welch's t-test, or non-parametric analogues, specify what values are being used (model-based values or the transformed/non-transformed values without adjustment from a model).

4.1.5.5. Bayesian Approaches

If using a Bayesian approach, then standard reporting requirements are required, including specification of any priors, explicit specification of the posterior, specification of what is being modelled to identify the DAMs.

4.1.5.6. Decision Criteria

If using a *p*-value criteria to identify DAMs then specify the nominal alpha value and the *p*-value threshold. If a multiple-testing correction is being performed, specify the nominal alpha value, the multiple-testing correction method (exact type, e.g. Bonferroni for family wise error rate correction or Benjamini & Hochberg for false discovery rate control), and any adjusted threshold value. If a fold-change or log fold-change criterion is used alone or in combination with nominal or adjusted *p*-values, then specify the level (e.g. 2x change). Also specify the exact order of operations and how the decision criterion is applied – this should be written clearly such that anyone could replicate this work should it be necessary.

4.1.5.7. Other

If using another approach for ranking and prioritisation of DAMs, specify the procedure clearly such that anyone could replicate the work should it be necessary.

4.1.6. Outputs

Because there are numerous approaches for the analysis and modelling of DAM data, there are also numerous formats for outputting this information. It would be impossible to enumerate all of the potential output types, styles, and other information here. Instead, we provide general guidance regarding outputs that may be submitted in support of a regulatory application.

REPORT:

4.1.6.1. Outputs and Supporting Files

Complete the file manifest included as part of the reporting template. The manifest must include the listing of all files included in the regulatory application package specific to the DAM analysis. Each file in the manifest must be accompanied by a description of the file. If the file being described is a tabular file, then the rows and columns must be described so that anyone can understand the contents of the file. Supporting files may include analysis scripts, software configurations or tables of metadata. The phrase `data object` refers to any machine readable input, output or metadata file.

4.1.7. References

Gelman, A. and Carlin, J. (2014) Beyond Power Calculations: Assessing Type S (Sign) and Type M (Magnitude) Errors. *Perspect. Psychol. Sci.*, **9**, 641–51.

Verheijen MC, Meier MJ, Asensio JO, Gant TW, Tong W, Yauk CL, Caiment F. (2022) R-ODAF: Omics data analysis framework for regulatory application. *Regul. Toxicol. Pharmacol*. **131**, 105143.

4.2. Multivariate Statistical Analysis Module

This multi-omics compliant module specifies the information needed by a regulatory scientist in order to assess multivariate statistical analysis.

The goals of this module are to specify what needs to be reported while not being prescriptive about the analysis performed. We leave judgement about the appropriateness of the analysis plan to the users.

4.2.1. Inputs

Input for multivariate statistical analysis, in the form of processed omics data (e.g. normalised gene expression data, normalised metabolome data) is generated using processes described in a DAPRM. The user should indicate what file(s) / data are used as the input for the analysis. Metadata files containing information associating omics samples with experimental parameters (e.g. treatments, dose levels, time points, sex, etc.) may also be necessary for the analysis. The user should specify these files as well.

REPORT:

1. **Data Input:** Provide the name of the input data object(s)
2. **Metadata Input:** Provide the name of the input metadata object(s)

4.2.2. Software, method and parameters

The requirements in this section are intended to facilitate reproducibility of analyses and make quality assessment possible. Operators are free to choose freely available open source, commercially available, or proprietary software. Operators must specify the name and version of all software that is used. Operators must also specify all add-ons, plug-ins, packages, or libraries (hereafter “libraries”) that are called, used, required, loaded, or otherwise brought into the software. This specification shall include the name and version of such libraries, and a hyperlink to download the libraries (or if download not available - hyperlink for further information on the software and/or library). In those cases where the libraries are locally developed and controlled by the operator, such that a hyperlink to download does not exist, the operator must specify the following: “this library is not available for download” and should then make the code available as specified below.

REPORT:

1. **Software:** The name of the software/analysis package along with the version.
2. **Operating system:** In some cases, modelling of data can be impacted by the operating system that the software is run on, therefore it is important to document this information.
3. **Additional libraries used:** A list of libraries used for analysis along with the version for each.
4. **Software availability:** A hyperlink to the software and source code (if available) is desirable.

4.2.3. Experimental conditions for multivariate analysis

To ensure clarity, operators must specify the conditions which were included or compared in the multivariate analysis. For unsupervised analysis, specific contrasts may not be defined (e.g. one may just model the control group). For supervised analysis, the operator will define specific conditions which are compared or specific factors which are examined (e.g. dose). For simple control vs. treated study designs, the contrast of interest is treated vs. control. This can become more complicated when the number of treatment levels increase, or if a time-course design is being used. In cases where the term "control" can be confusing, the operator must use a more specific term. For instance, if the study design uses a laboratory control and a field control then the operator must specify whether the laboratory control or the field control is being used, or if it is some transformation based on these controls. For instance, "treated time 12-hr vs field control 12-hr" is clearer than stating "treated time 12-hr vs control" in the case where control could refer to laboratory control or field control. For the purpose of clarity, the *factor* (alternatively called an independent variable) is an explanatory variable usually manipulated by the experimenter. Discrete factors have two or more *levels* (i.e. different values of the factor). Continuous factors (e.g. time) may take a range of values. Combinations of factor levels are often called treatments.

REPORT:

1. A table or listing that must include each factor, and level within each factor, that is being considered in the analysis.

4.2.4. Assay experimental design

In order to ensure that the omics data were analysed properly, the regulatory scientist must have a clear description of which samples were used in the multivariate analysis, how many samples there are per experimental group, and how they may be assayed together/measured simultaneously or not. In addition, the regulatory scientist can use this information to perform post hoc power analyses, as well as estimates for Type M (magnitude) and Type S (sign) error, which are equally, if not more, informative in assessing data quality and the likelihood of drawing erroneous conclusions. For more on Type M and Type S errors, see Gelman and Carlin (2014).

REPORT:

1. Report the relevant information in the reporting template as detailed below:
    a. **Group Sizes:** A table or listing of the number of samples used in each group or factor at a minimum. Should include the sample IDs that can be used to match across the report.
    b. **Covariance:** Identification of which samples may exhibit covariance (examples include: assayed on the same day, assayed in parallel, assayed on the same physical platform (e.g. chip/glass slide, cartridge, etc. for transcriptomics; LC-MS analytical batch, etc. for metabolomics), processed using the same wash solutions or reagents, processed using the same master mixes (transcriptomics only), littermates in the exposure

study, animals which are housed together, etc.). Provide an explicit statement if there is no reason to believe covariance exists between samples, and a justification to support this assertion.

c. **Technical Replicates:** Identify any samples that are technical replicates of each other, and how those samples are technical replicates (i.e. define what makes the samples technical replicates).

4.2.5. Multivariate statistical analysis – unsupervised

Many different types of unsupervised multivariate analysis can be applied. The goal of the unsupervised analysis is to provide an overview of the data to explore structures such as the major sources of variance, clustering or trends. These structures may result from, but are not limited to, the following: the biological effects being studied (e.g. dose effect), uncontrolled natural biological variation, or residual variance in the analytical procedure.

REPORT:

1. Name and description of the multivariate method
2. The groups or conditions included in the analysis
3. If variable selection used
    a. The name and description of the variable selection method
    b. Selected variables
4. Parameter settings (e.g. scaling method such as Pareto or unit variance)
5. Citation(s) to relevant literature describing methods (if available)

4.2.6. Multivariate statistical analysis – supervised

Supervised methods (usually classification or regression methods) are commonly used to focus the analysis on specific questions, e.g. whether the metabolic data can classify a sample into a chemical MoA, and to find which metabolic variables are most responsible for this classification. These methods are able to model data in cases where the treatment effect is small compared to other sources of variation. Many methods exist, but the chosen method should exhibit an ability to handle data with a) many variables, b) high degree of correlation between variables, c) high levels of noise, and d) missing data (if any). Common methods include partial least squares (PLS) or Orthogonal PLS (OPLS) regression when the outcome is continuous, or the equivalent discriminant analysis (PLS-DA and OPLS-DA) when the outcome is discrete (e.g. classification).

REPORT:

1. Name and description of the multivariate method used
2. The groups, conditions or factors contrasted or examined in the analysis
3. If variable selection used
    a. The name and description of the variable selection method used
    b. Selected variables
4. Parameter settings (e.g. scaling method)
5. Citation(s) to relevant literature describing methods (if available)

4.2.7. Multivariate statistical analysis - validation

All models, both supervised and unsupervised, must be statistically validated to show that they are robust and predictive. This should be done by either a) separating the metabolomics data into independent training (typically a maximum of 70% of dataset) and test sets (remaining 30% minimum of dataset), or b) internal cross-validation. In both cases, summary statistics such as Q^2 or misclassification rate should be calculated. Model complexity (e.g. number of principal components) should be chosen based on predictivity of the model.

REPORT:

1. Name and description of the model validation method used
2. Parameter settings
3. Citation(s) to relevant literature describing methods (if available)

4.2.8. Multivariate statistical analysis - variable importance for feature selection

One of the main objectives of multivariate analysis in metabolomics studies is usually to determine which variables (e.g. metabolites) are driving the observed effect, e.g. those which show differential regulation between a chemical treatment and control. Many approaches are based on assessing the weight of each variable in the developed model; additionally some methods can assess the statistical significance of these weights. For example, bootstrap procedures may be used to estimate confidence intervals on PCA loadings, allowing selection of variables where confidence intervals do not contain zero. Methods such as Variable Importance in Projection (VIP) or S-plot are able to rank variables by importance in PLS models. They can be used as long as a statistically sound approach to determining the significance threshold is used (e.g. bootstrap resampling).

REPORT:

1. Name and description of method used for variable importance/selection
2. Details of how significance threshold determined
3. Citation(s) to relevant literature describing methods (if available)

4.2.9. Outputs

Because there are numerous approaches for the analysis and modelling of multivariate data, there are also numerous formats for outputting this information. It would be impossible to enumerate all of the potential output types, styles, and other information here. Instead, we list general guidance regarding outputs that may be submitted in support of a regulatory application.

REPORT:

General requirements for outputs and supporting files:

Complete the file manifest included as part of the reporting template. The manifest must include the listing of all files included in the regulatory application package

specific to the analysis. Each file in the manifest must be accompanied by a description of the file. If the file being described is a tabular file, then the rows and columns must be described so that anyone can understand the contents of the file. Supporting files may include analysis scripts, software configurations or tables of metadata. The phrase `data object` refers to any machine readable input, output or metadata file.

Method specific requirements:

1. If unsupervised methods used, output of the unsupervised methods including plots (e.g. PCA scores, loadings, proportion of variance explained)
2. If supervised methods used, output of the supervised method including plots (e.g. PLS scores, loadings/weights, regression coefficients, proportion of variance in outcome explained by the model)
3. If validation performed, validation statistics (Q2, error rate, etc.)
4. If feature selection was performed, list of features selected by the method including
    a. Feature importance (e.g. VIP value)
    b. Feature significance (e.g. *p*-value, if available)

4.2.10. References

Gelman, A. and Carlin, J., 2014. Beyond power calculations: Assessing type S (sign) and type M (magnitude) errors. *Perspectives on Psychological Science*, *9*(6), pp.641-651.

4.3. Benchmark Dose Analysis and Quantification of Biological Potency

This multi-omics compliant module specifies the information needed in order to assess benchmark dose (BMD) analysis of omics data and the subsequent derivation of an estimated BMD for a molecule (e.g. a transcript, protein or metabolite) or a set of molecules (e.g. a gene set, a metabolite panel, etc.).

The goals of this module are to specify what needs to be reported to enable another scientist to reproduce the data analysis, while not being prescriptive about the analysis performed. Judgement relating to the suitability of the analysis plan is left to the end-user.

4.3.1. Inputs

Input for BMD analysis, in the form of processed omics data (e.g. normalized gene expression data, normalized metabolome data) is generated using processes described in a DAPRM. The user should indicate what file(s) / data are used as the input for the DAM analysis. Metadata files containing information associating omics samples with experimental parameters (e.g. treatments, dose levels, time points, sex, etc.) may also be necessary for BMD analysis. The user should specify these files as well.

REPORT:

4.3.1.1. Data Input

Provide the name of the input data object(s)

4.3.1.2. Metadata Input

Provide the name of the input metadata object(s)

4.3.2. Software Documentation

The requirements in this section are intended to facilitate reproducibility of analyses and make quality assessment possible. Freely available open source, commercially available, or proprietary software may be used. The name and version of all software that is used to perform the modelling and gene set analysis must be specified. All add-ons, plug-ins, packages, or libraries (hereafter all referred to as “libraries”) that are called, used, required, loaded, or otherwise brought into the software for use in identifying BMD values from omics data must also be specified. This specification shall include the name and version of such libraries, and a hyperlink to download the libraries or other source documentation. In those cases where the libraries are locally developed and controlled, such that a hyperlink to download does not exist, specify the following: “this library is not available for download” and should then make the code available as specified in the reporting template.

REPORT:

4.3.2.1. Software

The name of the software/analysis package along with the version.

4.3.2.2. Operating System

In certain cases dose response model fitting can be impacted by the operating system that the software is run on therefore it is important to document this variable.

4.3.2.3. Additional Libraries Used

A list of libraries used for analysis along with the version for each.

4.3.2.4. Software Availability

If the software is open source, a hyperlink to the software and source code (if available) is desirable.

4.3.3. Description of the Data to be Modelled

In order to perform dose-response modelling a test article will need to be assessed for its effects on omics level changes at multiple dose levels. This module must be paired with relevant experimental and data processing modules that describe the relevant meta-data (including dose levels/units) and how the annotated/normalised final data set used as input for the BMD analysis was derived.

Omics metrics vary by technology and platform. These differences can impact the coverage of information captured (e.g. whole genome vs partial genome), accuracy of annotation (e.g. probe to gene mapping for transcriptomics, mass accuracy and chromatographic retention time for metabolomics) and the ability to quantify molecules (e.g. differences in dynamic range in the case of RNA-seq vs hybridization technologies, the use of reference standards in metabolomics). In order to maximise reproducibility of the analysis it is important to capture this information; this information should be reported in the complementary Data Acquisition and Processing Reporting Modules (DAPRMs) provided with the omics reporting frameworks.

Omics data are often transformed to a variety of different levels. In the case of individual features (i.e. the individual measurements for transcripts/probe sets in the case of transcriptomics, or metabolites/mass spectral features in the case of metabolomics that are derived from an omic technology) algorithmic scaling (normalisation) and log transformation is often performed. In addition, individual features can be merged into meta-features, such as genes, and further collapsed into biologically meaningful sets such as pathways or gene ontology biological processes (e.g. fatty acid metabolism). Henceforth in this document the actual data that are fit to dose response models will be called model substrate data (MSD). Much of this is captured in other sections of this Guidance Document; however, it is important to emphasise the level of data abstraction that was performed on the data prior to BMD modelling in order to ensure reproducibility and in some cases to ensure the replicability of the reported output from the analysis.

REPORT:

4.3.3.1. Dose Levels

Identification of dose levels included in the BMD analysis and rationale for exclusion of any dose levels. Please specify the N for each dose.

4.3.3.2. Annotation for Model Substrate Data (MSD)

With certain platforms (e.g. most transcriptomic platforms) there are existing annotations that will be associated with the data as a part of the data acquisition and processing procedures (as reported in the technology-specific DAPRMs). If MSD for BMD modelling corresponds to the measured features and annotations, without additional data abstraction (e.g. aggregation, summarization or merging of features), then this annotation and citation to the associated DAPRM is sufficient and must be provided. If features are merged or computed into meta-features (e.g. genes, gene sets, metabolite collections), then precise annotation of the meta-features must be reported here. In some cases, measurements from certain technologies do not have a clear mapping to specific annotations. In the latter case the user should indicate that annotations are not available and provide any annotations that were generated to support the analysis.

4.3.3.3. Version of Platform Annotation that Corresponds to Technology

Annotations for the measured MSD in a data set are in certain cases updated frequently. The updated mapping causes changes in the association of the features measured by the platform/technology therefore impacting the biological mapping and interpretation of the data. The annotations are often associated with a date and/or a version. This information must be captured to ensure reproducibility of the findings. Identification of the platform and exact version allows for appropriate linkage of the platform biological annotations.

4.3.4. Pre-filtering of Data Prior to Dose-Response Modelling

To reduce the burden of model fitting, which can be resource intensive, a prefiltering step is sometimes used to remove MSD that do not demonstrate a significant dose-related change. Often a statistical test in combination with an absolute change filtering is performed to identify the MSD that are responding to chemical treatment. In addition, due to the high dimensionality of omics data, these statistical tests are often associated with false discovery rate or a multiple testing correction. The details of the prefiltering analysis must be fully described in order to provide transparency as to the molecules that were passed into the subsequent modelling process.

Report the relevant information in the reporting template as detailed below:

REPORT:

4.3.4.1. Statistical Test Performed to Identify Dose-Responsive MSD

There are a variety of methods for performing statistical filtering of dose response data. These range from pairwise methods such as ANOVA, to trend-based tests such as the Williams trend test. Record the

statistical method that was applied to the data. If additional filtering was not applied indicate NA (not applicable)

4.3.4.2. Statistical Threshold Applied

Indicate the statistical threshold that was applied when performing any statistical test to pre-filter the data.

4.3.4.3. Statistical Multiple Testing Correction Method

Because of the large number of replicate statistical tests that are performed when analysing an omics data set, an adjusted statistical threshold is often applied that corrects for a large number of tests, therefore reducing the false discovery rate. There are multiple methods for performing this correction. Report the method that was used to perform the multiple testing correction. If no correction was applied, indicate NA (not applicable).

4.3.4.4. Additional Statistical Test Parameters

This section is to describe any additional parameters that were applied and can impact statistical testing results. An example would be the number of iterations that were performed when employing a bootstrap method for *p*-value determination.

4.3.4.5. Additional Filtering

In addition to a statistical filter, additional filters are sometimes applied to ensure that only responses representing biologically meaningful changes are retained. One of the most common is a fold change filter. If more complicated filter methods are applied a detailed description of the method and all critical parameters should be provided. If no additional filtering was applied, indicate NA (not applicable).

4.3.5. Benchmark Dose Modelling

BMD modelling is the process of fitting mathematical models to dose-response data to identify a modelled dose level where a predefined level (i.e. the benchmark response or BMR) of activity is predicted to occur. As these are modelled data, there is uncertainty in the potency estimates and therefore uncertainty bounds are often reported in addition to central estimates. Two general approaches can be taken to dose response modelling: parametric (i.e. predefined mathematical models) and non-parametric (model free). These methods are fundamentally different, have different assumptions about the data and may be viewed in different lights by regulatory authorities; therefore, it is important to document which approach was employed. Once the overall approach is chosen a number of choices must be made including (in the case of parametric modelling): which of the models are run; the BMR; whether/how a best model is selected and by what method; or, if model averaging is performed, by what method. In addition to these parameters there are often other options that can be modified, all of which can influence the fitting process and

subsequent potency estimates of the MSD. For this reason, all of these choices should be documented.

Report the relevant information in the reporting template as detailed below:

REPORT:

4.3.5.1. BMD Modelling Approach

As noted above there are different approaches (e.g. parametric and non-parametric approaches) to dose response modelling. The specific method used should be specified.

4.3.5.2. Method for Final Estimation of MSD BMD

There are two general approaches for identifying a final potency estimate for a given feature - best model selection and model averaging. With the best model approach a single best model is used to derive the BMD; whereas, model averaging uses a weighted average to derive a final BMD estimate. In the case of non-parametric modelling, typically only a single model fitting process is performed, hence there isn't a best model or models to average. In this case the method for final estimation of BMD is simply the modelling approach.

4.3.5.3. List of Models Fit to the Data

In the case of parametric modelling (with or without model averaging) a set of models that represent different dose-response shapes are used to model the data. Report all models fit to the data in the modelling process.

4.3.5.4. Model Averaging Approach

When performing model averaging, a variety of mathematical approaches can be used that can impact the eventual BMD estimate of an MSD. Record what method was employed and any other associated critical parameters associated with the averaging process. If model averaging is not employed, indicate NA (not applicable) in this section.

4.3.5.5. Method for Determining Benchmark Response (BMR; a.k.a. BMR Type)

There are a number of ways of determining a critical response threshold, i.e. a BMR. Two examples are the multiple of standard deviation from control or the relative deviation (percent change) from control. Report the method used to determine the BMR.

4.3.5.6. Benchmark Response (BMR, a.k.a. BMR factor)

The BMR is the degree of change relative to control that is deemed to either be statistically or toxicologically relevant for the study. The threshold that was applied in the modelling process should be recorded.

4.3.5.7. Number of Filtering Iterations

In the case of individual parametric model fitting, a maximum number of iterations is typically allowed. This setting reflects the number of times a model is fit to an MSD to obtain a convergent estimate of the fit. If applicable, report the maximum number of iterations that were employed.

4.3.5.8. Confidence Interval of the BMD (a.k.a. Confidence Level)

This modelling parameter sets the confidence interval range that will be determined for the BMD values derived from each model fit to the data. Report how and what confidence intervals were derived.

4.3.5.9. Power Parameter Restriction

Often it is desirable to restrict the value of certain modelling parameters when performing parametric modelling. The most commonly restricted parameter is power and the restriction is most commonly applied when modelling the Hill and power models. Restricting the power parameter avoids biologically implausible fits to the data. Report if any of the model parameters were restricted in the modelling process and the specific restriction threshold.

4.3.5.10. Variance Assumption

Toxicology data are often heteroscedastic (i.e. the spread on the response distribution increases with dose). This is often referred to as non-constant variance. Different assumptions are made when data are assumed to be hetero vs homoscedastic, which can impact model fitting and BMD derivation. It is therefore necessary to document if constant or non-constant variance was assumed.

4.3.5.11. Model Selection

In the case of the best model approach to parametric modelling, criteria for selection of the best model must be selected. There are several approaches that can be used to select the best model, all of which can impact eventual potency estimates for the MSDs. It is therefore necessary to document the method used to select the best model. If a best model approach is not applied (e.g. non-parametric modelling or model averaging, indicate NA (not applicable)).

4.3.5.12. Criteria to Identify Models that Extrapolate Outside the Dose Range

When using the best model approach it is sometimes the case that due to dose selection the estimated BMD is well below the dose range. In this case the BMD can not be accurately estimated but it is desired to retain the MSD in the analysis. There are several options to identify and deal with the extrapolated (a.k.a. flagged) BMD value. An example of this is when the Hill model is identified as the best model, but its k-parameter is at, or less, than the lowest non-zero dose. If this criterion is met, a number of options to either accept the model and report the BMD value, or modify it in some way/ pick an alternative best model, can be employed. These choices will impact the BMD estimate/reporting of the set of flagged MSD and potentially impact downstream analysis. Document how flagged models were dealt with in the analysis.

4.3.6. Determination of Biological Entity of Biological Set Level BMD Values

Often MSD are fractional representations of a biological entity (e.g. a gene or metabolite) or a biological set (e.g. a collection of genes or metabolites that make up a pathway). Organising the MSD into recognized biological entities helps in contextualising a biological response and allows for more robust BMD estimates at a molecular or biological process level and therefore can provide greater confidence and context to biological potency estimates. Several options are available to perform this part of the analysis, all of which can impact the eventual biological entity or biological set activity estimate and potency characterization. Thus, the parameter selections that were made when performing this step must be clearly described.

REPORT:

4.3.6.1. Biological Entity or Biological Set Annotation Used for the Analysis

In this section, document which biological entities (e.g. genes, metabolites) or biological sets (e.g. GO terms, KEGG pathways) were used in the analysis. The platform annotation (described above) should capture the versioning of the annotations used; however, if this is not the case (i.e. independent mappings of MSD to biological entities or sets were used) the source of these entities and any versioning should be documented

4.3.6.2. MSD Acceptance Criteria for Use in Biological Entity or Set Analysis

A number of criteria can be applied to the MSD models (e.g. the best fit model) to remove the fits where the quality is poor or the potency estimates are highly uncertain either due to noisy data or model extrapolation. Some examples of the metrics that can be applied for removing such MSD include only including BMD < highest dose and setting thresholds on the BMD/BMDL ratio and global goodness of fit *p*-value. Document all metrics and associated thresholds used in the analysis.

4.3.6.3. Criteria for Identification of “Active” Biological Sets

In the case of parsing MSD into biological sets (e.g. pathways or gene ontologies), criteria must be set to identify the sets that are “active” (i.e. adequately populated by dose responsive MSD with acceptable model fits deemed to be responsive to treatment). There are a number of different types of criteria that can be used to gauge the active status of a biological set. Examples include percentage populated based on annotation size (MSD/Total annotated*100), minimum number of dose-responsive MSD in a biological entity set, and enrichment tests such as the Fisher’s exact test. Report the methods and thresholds set to define ‘active’ biological sets.

4.3.6.4. Method of Estimating the BMD, BMDL and BMDU of the Individual Biological Entity or Biological Sets

When MSD are sorted into sets this means a set contains several BMD values representative of each MSD in the set. When this happens, and the biological set is determined to be active (in the case of sets), a representative BMD, BMDL and BMDU value for the entire set can be determined. There are a number of simple mathematical methods that can be employed to obtain a representative value for each active biological entity set. These most commonly are the mean, median or 5th percentile (by total annotated biological entities) dose-responsive MSD BMD, BMDL and BMDU values contained in each active biological set; however, alternative methods are also acceptable. The choice of method will impact the estimate of potency for most biological entity sets so it is important that the method for deriving a representative BMD, BMDL and BMDU value be very clearly documented.

4.3.7. Outputs and Supporting Files

Complete the file manifest included as part of the reporting template. The manifest must include the listing of all files included in the regulatory application package specific to the BMD analysis. Each file in the manifest must be accompanied by a description of the file. If the file being described is a tabular file, then the rows and columns must be described so that anyone can understand the contents of the file. Supporting files may include analysis scripts, software configurations or tables of metadata. The phrase `data object` refers to any machine readable input, output or metadata file.

REPORT:

4.3.7.1. Output and Supporting Files

4.3.8. References

None.