Transcriptomic Reporting Framework (TRF)

Draft May 2021

This version of the TRF is currently under development by the Expert Group on Transcriptomics and is intended to be presented in draft form for the OECD Extended Advisory Group on Molecular Screening and Toxicogenomics (EAGMST) and Working Party and Hazard Assessment (WPHA) joint 'omics session in June 2021. The TRF is currently tested through case studies and will be distributed to the EAGMST for more formal review in the near future once members of the EG have revised and agreed to the text after the case studies. Given this is a draft that is subject to revision, members of EAGMST and WPHA are invited to acknowledge progress and ask questions for clarification or provide general comments, rather than specific technical comments on text in the draft. Please note that this is a draft document, and members are requested to not circulate the document. This document is accompanied with a Reporting Template in xl format.

JT03476751

Table of Contents

1. Introduction

1.1. Background

1.   Transcriptomic methodologies are increasingly applied in research and regulatory toxicology to provide a greater understanding of mechanisms of toxicity, inform read-across and identify point-of-departure(s). To facilitate use in regulatory decision-making in the absence of a test guideline, a comprehensive reporting framework is necessary to thoroughly document the components of a transcriptomic study (Sauer et al., 2017). Such a framework will increase the transparency for data processing methods used to convert raw transcriptomic 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). Thus, in 2018, the Organisation for Economic Co-operation and Development (OECD) Extended Advisory Group on Molecular Screening and Toxicogenomics (EAGMST) undertook the development of a Transcriptomic Reporting Framework (TRF) to foster and encourage further regulatory uptake of transcriptomic data. Reporting using the TRF will provide the essential information needed to evaluate transcriptomic study designs, data quality and applicability to regulatory decision-making processes and also maximise the likelihood that the analytical results of a transcriptomic experiment can be reproduced.

1.2. Scope

2.   The TRF was designed as a tool for documenting the details of laboratory-based toxicology studies that apply a transcriptomic technology: i.e. an assay that measures the abundance of many transcripts simultaneously and thus provides highly multiplexed outputs. The TRF 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 a transcriptomic 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.

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

4.   The TRF incorporates and refines the concepts described by an expert group convened by the European Centre for Ecotoxicology and Toxicology of Chemicals (ECETOC) published in Regulatory Toxicology and Pharmacology (Gant et al., 2017). Elements of the TRF are based in part on previously established frameworks for toxicology study annotation for data sharing and regulatory application (Fostel et al., 2007; McConnell et al., 2014; Schneider et al., 2019; Segal et al., 2015), as well as elements of previously

established frameworks for annotation of data from transcriptomic studies (Brazma et al., 2001; Conesa et al., 2016; Parkinson et al., 2005; Waters et al., 2003). Previous frameworks for annotation of transcriptomic data focus primarily upon annotation of data (raw and normalised), samples, sample to data relationships and technology-specific feature annotation (Brazma et al., 2001; Parkinson et al., 2007). The TRF includes all of these elements, but also provides a means to document the computational steps used to analyse the transcriptomic data and generate downstream results that may be of use in a regulatory decision-making context.

5.   Transcriptomic data sets can be generated using a variety of different technologies. Likewise, there are many different types of analyses and sequences of analyses that may be performed on transcriptomic data. As such, the TRF has been designed in a modular fashion. A toxicology experiment reporting module (TERM) was developed for reporting experimental design information. In addition, technology-specific data acquisition and processing reporting modules (DAPRM) that document data generation, handling, normalisation and filtering steps have been developed. Lastly, several downstream analysis reporting modules (DARMs) have been developed for a variety of analysis approaches. These DARMs detail the steps and resources necessary to reproduce computational analyses of transcriptomic data and are intended to be coupled with upstream TERM and DAPRM modules (e.g., identification of differentially expressed genes, benchmark dose modelling, etc.). The reporting modules prompt the data generator to list all components of an analysis necessary to experimentally and computationally reproduce their results, but do not prescribe best practices for any phase of the study; thus, the TRF is strictly a reporting tool.

6.   To complete the TRF, researchers select the reporting modules (and reporting fields) that are relevant to their study (Figure 1) and report the information that would be required by an end-user to fully comprehend and replicate the analyses. Note that certain reporting specifications are required while others are considered optional. All reporting specifications that are required are noted as either "required" or are signified by the word "must". Specifications that are not required will be noted as "optional" or signified by the word "should". Although only select reporting elements are noted as ''required", it should be noted that a more comprehensive submission will enhance confidence and potential use in regulatory applications.

7.   A subset of mandatory reporting elements are included in a Study Summary Reporting Module (SSRM) to provide a high-level overview of the experimental and analytical information.

8.   This TRF guidance document provides narrative descriptions and basic background information for each reporting field within each module. The reporting itself should be done in the complementary multi-workbook excel file provided that contains each reporting field and within a workbook for each module.
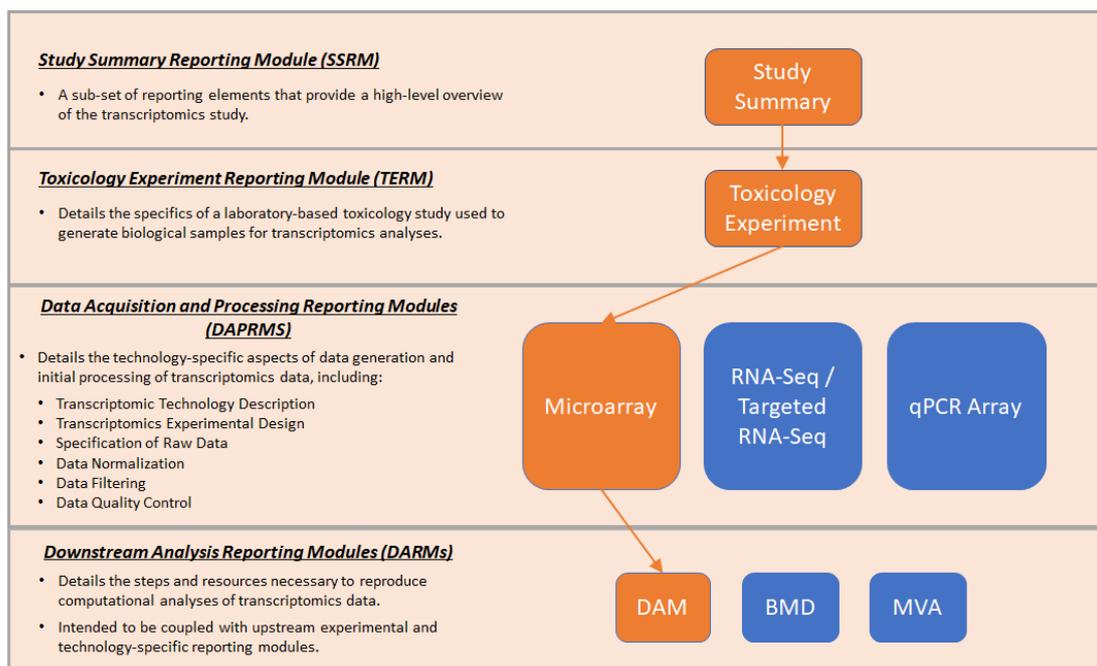
Study Summary Reporting Module (SSRM)

- A sub-set of reporting elements that provide a high-level overview of the transcriptomics study.

Study Summary

Toxicology Experiment Reporting Module (TERM)

- Details the specifics of a laboratory-based toxicology study used to generate biological samples for transcriptomics analyses.

Toxicology Experiment

Data Acquisition and Processing Reporting Modules (DAPRMS)

- Details the technology-specific aspects of data generation and initial processing of transcriptomics data, including:
  - Transcriptomic Technology Description
  - Transcriptomics Experimental Design
  - Specification of Raw Data
  - Data Normalization
  - Data Filtering
  - Data Quality Control

Microarray

RNA-Seq / Targeted RNA-Seq

qPCR Array

Downstream Analysis Reporting Modules (DARMs)

- Details the steps and resources necessary to reproduce computational analyses of transcriptomics data.
- Intended to be coupled with upstream experimental and technology-specific reporting modules.

DAM   **BMD**   **MVA**

Figure 1. Modular structure of the Transcriptomics Reporting Framework (TRF). To complete the TRF, researchers 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. An example is illustrated (orange highlights) where biological samples from a toxicology experiment were analyzed using microarrays and processed gene expression data is used to identify differentially expressed genes. DAM = Differentially Abundant Molecules. BMD = Benchmark Dose. MVA = Multivariate Analysis. Additional DAPRMs and DARMs can be developed as needed to address new transcriptomic technologies and additional analytical methods of interest.

1.3. Related 'Omics Standards Projects

9.   The TRF document was developed in parallel with a reporting framework for metabolomics under the ECETOC supported MERIT (Metabolomics standaRds Initiative in Toxicology) project and will also stylistically conform (where applicable) to previously published OECD guidance documents relevant to the field of toxicological research. Where possible, terminology and reporting fields in the DARMs have been developed to accommodate both transcriptomic and metabolomic data analyses.

1.4. 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

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

2. Toxicology Experiment Reporting Module

The Toxicology Experiment Reporting Module (TERM) serves to capture and report the key descriptors of the *in vivo* or *in vitro* toxicology study from which samples are derived for transcriptomics analysis.

2.1. Study Rationale

10.   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.

11.   Over the past decade a variety of applications of transcriptomic data to support decision-making in a regulatory context have been demonstrated, but to date have not been widely implemented in the regulatory community. Notable transcriptomic applications include: 1) acquiring information about the molecular changes that result from an exposure to inform mode of action and to facilitate assessment of probable human relevance (e.g. Kienhuis, van de Poll et al. 2009; Dong, Gill et al. 2016; Rodrigues, Heymans et al. 2016; Rooney, Ryan et al. 2017; Venkatratnam, House et al. 2018); 2) establishing chemical groups based on similar gene expression profiles (i.e. read-across) (e.g. De Abrew, Kainkaryam et al. 2016; Grimm, Iwata et al. 2016); 3) supporting weight of evidence approaches to identify hazard and risk (e.g. within Bourdon-Lacombe, Moffat et al. 2015; Cote, Andersen et al. 2016); 4) tiered assessment screens (e.g. Rowlands, Budinsky et al. 2013; Thomas, Philbert et al. 2013); 5) cross-species analysis (e.g. Black, Budinsky et al. 2012); and 6) deriving a transcriptomic point of departure (e.g. Farmahin, Williams et al. 2017; NTP 2018). This foundational work has shown that the optimal experimental design and analytical approach used in transcriptomic studies is highly dependent on the regulatory study objective.

12.   Experimental designs for each of the applications described above will differ. For example, if the identification of differentially expressed genes (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 gene expression 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 expression profiles against which comparisons can be made. The use of benchmark-dose analysis to identify a transcriptomic 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 platform used in deriving the transcriptomic profiles (e.g. whole genome approaches versus a targeted panel of genes) and clearly directs the downstream analysis applied.

13.   Overall, a clear rationale describing why a transcriptomic 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. 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).

2.2. Objectives

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

2.3. Test Guideline compliance

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

2.4. 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).

2.5. Model selection

For animal 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 the human biology. Provide a rationale for tissue or organ selection for the study.

Provide a rationale for the species, strain and reference genome used.

For *in vitro* studies, briefly describe the biological relevance of the test system used in relation to the tissue/organ/species of interest.

2.6. Dose level and dose interval selection

Provide a brief rationale for the selection of the employed doses levels and dose-intervals. For example, selection for *in vivo* experimentation may be based on known toxicological effects or molecular changes documented for the test article identified in prior studies, allowing for "read-across" between transcriptomic 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 exposures *in vivo*, test material concentrations may be chosen using a quantitative *in vitro/in vivo* extrapolation (IVIVE) rationale. Similarly, a rationale for *in vitro* and *in vivo* dose intervals should be provided.

2.7. 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.

2.8. Time point selection

Provide a brief rationale for the exposure durations and sampling time points. Gene expression profiles are highly dependent on both the duration of exposure and the time post-treatment of sample collection. Expression changes may adapt over time, with early time points reflecting molecular initiating and early key events, and later times points reflecting pathological changes or adaptation. For *in vivo* studies, the interval between the final dosing and sample procurement should be specified. In addition, sampling-time post-treatment is critical to define. For example, a gene expression profile 4 hours following a treatment *in vitro* can be significantly different from the profile 24 hours following the treatment.

2.9. 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.

2.10. Limitations

To facilitate regulatory evaluation, when appropriate, indicate the study limitations that could affect the outcome or the interpretation of the results. These can include technical or mechanistic limitations in relation to known modes of action. For example, if the culture conditions include poorly described systems that would be a source of uncontrollable variability (such as serum or matrices like matrigel); such information should be available to the evaluator. Likewise, some of the test articles used might have physicochemical properties (lipophilicity, volatility, etc.) that might lead to a cell exposure that is different from the expected exposure. For example, the compound might interact with plastic or, proteins in the serum medium before it enters the cell, or even (partially) evaporate. 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

14.   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 2018.a). Studies submitted for analysis to regulatory agencies should be 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. Information regarding the test item characterisation is needed to inform potential route of exposure, as well as physicochemical properties that might influence the study design and transcriptomic experimentation (i.e., solubility, volatility, etc.).

15.   Regulatory scientists must have all test items, vehicle, and control identification and characterisation information in order to accurately interpret transcriptomic study results. Detailed reporting of all test items used in transcriptomic experimentation will aid regulators' ability to confidently interpret 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 2018.b), 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 2016. a).

REPORT:

2.11.   Test item name

2.12.   Formulation composition if test item is a mixture

   a.   Identify substances that make up the mixture

   b.   Percentage of substances (if known)

2.13.   Preparation of test item (composition)

   a.   Concentration of test items

   b.   Concentration of diluent(s)

   c.   Identification of impurities

2.14.   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.   Vapor pressure

   i.   Partition coefficient

   j.   Water solubility

k. Solubility in organic solvents/fat solubility

l. Particle size distribution/fiber length and diameter distribution

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

2.15. Chemical stability

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

b. Storage stability and reactivity towards container material

c. Stability: thermal, sunlight, metals

d. Stability: dissociation constant

2.16. Commercial source

a. Vendor

b. Manufacture ID

c. Lot (Batch) number

d. Purity

e. Salt form

f. Expiration date

g. Storage conditions

2.17. Test Item specific identifiers

a. CAS

b. SMILES

c. IUPAC name

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

The data submitter should provide the source of the substance-specific identifier for each of the identifiers reported.

2.3. Test System Characteristics

16.   Transcriptomic studies suitable for use in a regulatory context may be performed in a wide range of model species and/or *in vitro* test systems. During the introduction of transcriptomic technologies into the field of toxicology, researchers were limited in terms of species selection by the small variety of species-specific platforms (mostly microarrays) available from commercial vendors. The traditional use of rodent species (i.e. mouse and rat) in toxicity testing and the availability of annotated genomes for these species partially drove commercial vendors to develop transcriptomic platforms specific for rodents (i.e.

mouse and rat). The proliferation of annotated genomes and the introduction of RNA-Seq, targeted RNA-Seq and quantitative real-time PCR arrays into the field of toxicology greatly expanded the number of species from which transcriptomic data could be generated for use in evaluating chemical toxicity and also roughly coincided with increased emphasis on the use of alternative species (i.e. fish, invertebrates, etc.) and *in vitro* models (many of which are human-derived) for toxicity testing.

17.   Similar to traditional toxicity testing, it is critical that regulatory scientists applying transcriptomic 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 transcriptomic data in order to critically evaluate the results and accurately compare the results across studies and data types.

18.   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 individuals 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 individuals were examined, then this should be explicitly described by the researcher. Researchers should also include detailed information on the housing conditions for all individuals included in a study including number of individuals 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 (Animal Research: Reporting of In Vivo Experiments) guidelines should be included (Kilkenny, Browne et al. 2010).

19.   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. 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 characterizing the model system and describe any other relevant characteristic that might not be listed here.

REPORT:

2.18.   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 carried out before or during the experiment

    i. Quality criteria before use

    j. Health status and acclimation prior to study start

    k. Randomization of animals to groups

    l. Other …

In vitro

    a. Cell type (cell line or primary cells, tumor cells, etc…)

    b. Origin (animal/organ/tissue)

    c. Cell passage number

    d. Differentiation stage

    e. Absence of mycoplasma

    f. Metabolic competence

    g. Supplier

    h. Quality criteria before use

    i. Other…

2.19. Housing, husbandry and culture conditions

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 etc. (for fish)

    j. Type of food

k. Access to food and water

l. Environmental enrichment

m. Methods for fertilisation/collection of eggs, if applicable

n. Other appropriate characteristics

In vitro

a. Incubation characteristics:

- Temperature

- CO_2/O_2 conditions

- Humidity

- Other

b. State of the cells before use:

- Viability (including test used)

- Quality control

- Morphology

- Recommended confluency of use

- Other

c. Culture media (in case of multiple)

d. Use of serum (with details such as species of origin, age, sex, etc.)

e. Use of antibiotics

f. Use of feeder cells

g. Use of matrixes or scaffolds

2.4. Study Design

20.   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 transcriptomic 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 transcriptomic data is based on previously published OECD guidance documents for the respective areas (OECD 2014; OECD 2015; OECD 2017). Much of the reporting guidance is based on the application of good laboratory practice (GLP) principles and good *in vitro* method practice (GIVMP) according to an internationally accepted definition, ensuring mutual acceptance of data (MAD) across OECD countries (OECD 1998; OECD 2018.c). 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 transcriptomic 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. In addition to the relationship of mechanistic arguments employed (see Section 2.1 on study rationale), 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 on study rationale and Section 2.2 on test system characteristics) (EC 2010).

REPORT:

2.20. 3Rs considerations

Briefly describe how the study addresses the 3R principles.

2.21. Dosing

a. Dose levels

Indicate each of the dose levels/concentrations used in the study and identify the matched vehicle/solvent controls to be used.

b. Dose intervals

Indicate:

- Dose interval (acute single or chronic dosing).
- Time of day of dosing.
- Frequency of dosing.

2.22. 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 transcriptomic analysis, with details concerning:

a. Commercial source, detailing 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.

2.23. 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 article. All control types should be reported (i.e., positive control, negative control, vehicle control, blank, etc.) and following the criteria described in Section 2.2. 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 animal 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).

2.24. Numbers of individuals/samples per treatment

a. Clearly describe the number of biological replicates in each treatment condition.

b. Describe the number of technical and analytical replicates.

2.25. 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.

b. Sampling schemes
Describe sampling schemes used to prevent sample collection bias, and to ensure proper sample labelling post-collection, using methods such as:

- Sequential
- Stratified
- Systematic
- Randomized
- Ranked set

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 2.8 Sample Identification Codes) that does not represent the sample or treatment type.

2.26. Observations/examinations during treatment

Where appropriate, include details of other experimental observations used in the experiments generating the toxicogenomic 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 chemistry and 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

h. Metabolomics, proteomics, biochemical and molecular biological analyses, etc. in in vivo and in vitro experiments

2.27. Time-table

Detail the time-tables used to perform the study protocol with respect to:

a. Treatments

b. Sample collections

c. Time since last dose administered

d. Time to sample preservation

2.5. Treatment Conditions

21.   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 the modulation of genes and pathways representative of key events possibly related to the final adverse outcome. The objective of well-designed transcriptomic studies informs the selection of the treatment conditions, which, in turn, impact the final outcome. Transcriptomic 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 transcriptomic results.

22.   For both *in vivo* and *in vitro* transcriptomic studies (as in any study), 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.

23.   The present guidance for reporting treatment conditions in transcriptomic studies is based mainly on previously published OECD documents (OECD 2014; OECD 2017; OECD 2018.c; OECD 2018.a) and standard harmonised templates for reporting of information derived from *in vivo* or *in vitro* studies for the risk assessment of chemicals (OECD 2016; OECD 2018). The ARRIVE guidelines for reporting animal research were

also 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 gene expression studies.

REPORT:

2.28. Route of administration

Indicate the selected route of administration.

In vivo

    a. Oral

- Gavage
- Diet

    b. Dermal

    c. Inhalation (specify the type of exposure)

    d. Implantation

    e. Parenteral

    f. Exposure regime, if applicable (static, semi-static, flow through)

    g. Other …

In vitro

    a. Direct addition of the test item or a preparation of the test item to cultures

    b. Substitution of culture medium

    c. Exposure at the air-liquid interface

    d. Other…

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 'Test System Characteristics'), in addition to:

In vivo

    a. Fasting period

    b. Anesthesia and/or analgesia

    c. Other …

In vitro

    a. Serum free or serum depleted medium

    b. Use of items for limiting media evaporation

    c. Other…

2.29. Test item preparation

Provide the dose/concentration levels and the dose/concentration level spacing. 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. For dose levels/concentration to be used, see paragraph 2.21:

    a.   Dilution in a vehicle

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

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

    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

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

For *in vivo* studies, provide details about:

    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 material 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.

2.30. Vehicle description and delivery volume

If the test chemical is dissolved or suspended in a suitable vehicle, provide all the relevant information on the vehicle, the delivery volume and the final concentration of the vehicle in the test item preparation. Indicate 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 'Test Item and Control Items').

2.31. Exposure duration / schedule

Indicate exposure duration (hours, days, weeks or months) and the frequency of the administration.

Indicate the recovery period (in days, weeks, months, if any) after the exposure to the test substance.

2.32. Housing / culture conditions during treatment

If different from the normal maintenance (refer to section 'Test System Characteristics'), please report on the environmental conditions recorded during treatment. Describe also the 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

24.   Transcriptomics studies can be conducted using samples from animal studies and from *in vitro* studies. When properly performed, transcriptomic profiling of such a sample produces a "snapshot" of its gene expression levels and allows assessment of alterations in gene expression due to perturbation of biological processes. It is widely understood that proper sample preparation is a key step in transcriptomic studies and adequate care must be taken to ensure sample fidelity. Several factors relating to preparation for sampling need to be taken into account because of their potential for causing alterations in the transcriptome, which may confound biological interpretation of the data. This applies to both animal and *in vitro* studies, although the type of biological sample drives the selection of the subsequent handling steps.

25.   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, anesthetics 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 studies (HK 2004; Overmyer, Thonusin

et al. 2015; Nakatsu, Igarashi et al. 2017). Also, for *in vitro* studies, the methods used for harvesting the samples may influence the transcriptome.

Special notes for transcriptomic analyses

These methods may vary largely, ranging from the use of a detergent to lyse the cells to scraping of adherent cells. A detailed report on the methodologies used for collecting specimens, stabilizing RNA before/during the extraction steps and during storage will allow reviewers to appropriately evaluate the quality of the transcriptomic studies.

Additionally, biotic and abiotic information at the time of harvesting should be collected to allow for assessment of sample fidelity. This also includes the conditions under which

---

REPORT:

2.33.   **Type of biological sample** (*in vitro*: single cell, cell culture, 2D or 3D culture, single or multi type cell culture; *in vivo*: cells, tissue, organ, organism in toto, fresh, frozen or paraffin-embedded specimens)

In vivo

    a.   Anesthetic 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 (e.g. methods for dissection, isolation of tissues or organs, etc.)

    f.   If relevant, procedures for RNA stabilisation (RNA-later, flash-freeze in liquid nitrogen, disruption and homogenisation of samples in the presence of RNase-inhibiting or denaturing reagents...)

In vitro

    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)

    d.   Number of culture passage

    e.   Morphology

    f.   Mechanical methods used for collection of biological samples (e.g. removal of media, washes, scraping into sampling vial, etc.)

g. Sampling vial (e.g. type of vial or tube, chemicals within the sample tube added to preserve sample (RNA-later, etc.))

2.34. Pooling (or aliquoting) of samples

Provide details of sample pooling or aliquoting scheme used prior to omics analysis.

2.35. Sample storage and transport

Provide logistical details pertaining to sample storage duration and conditions before omics analysis.

a. Sample storage (refrigeration requirements) prior to analysis.
b. Transportation between experimental facilities where appropriate

2.7. Sample Identification Codes

26. Sample management is a critical component of regulatory and non-regulatory experimentation that should be carefully planned, especially for high-throughput data generation systems like transcriptomic technologies. 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 for transcriptomic experimentation should be given a unique identification code and information stored in a secure LIMS where available.

27. 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 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:

2.36. Laboratory information management system (LIMS), if applicable
a. Software used for information management

2.37. Unique code for each sample
a. Method of code generation

2.38. Information stored for each sample code:

a. Sample name

b. Sample source

c. Sample storage

d. Characterisation

e. Stability

f. Vehicle and/or buffer

g. Test item information

h. Receiving date

i. Test date

j. Metadata

k. Experimental results

2.8. Supporting Data Streams

28. Transcriptomic studies can be used to address different types of regulatory questions. In order to be able to fully appreciate a transcriptomic 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.

29. However, there may be situations where even a higher level of detail is needed to allow for use of transcriptomic 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 Standard Operating Procedures (SOP) for RNA processing 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

Black, M. B., R. A. Budinsky, et al. (2012). "Cross-species comparisons of transcriptomic alterations in human and rat primary hepatocytes exposed to 2,3,7,8-tetrachlorodibenzo-p-dioxin." Toxicol Sci **127**(1): 199-215.

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

Bourdon-Lacombe, J. A., I. D. Moffat, et al. (2015). "Technical guide for applications of gene expression profiling in human health risk assessment of environmental chemicals." Regul Toxicol Pharmacol **72**(2): 292-309.

Bryant, S. and D. L. Manning (1998). "Isolation of messenger RNA." Methods Mol Biol **86**: 61-64.

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

Chomczynski, P. and N. Sacchi (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.

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.

De Abrew, K. N., R. M. Kainkaryam, et al. (2016). "Grouping 34 Chemicals Based on Mode of Action Using Connectivity Mapping." Toxicol Sci **151**(2): 447-461.

Dong, H., S. Gill, et al. (2016). "Toxicogenomic assessment of liver responses following subchronic exposure to furan in Fischer F344 rats." Arch Toxicol **90**(6): 1351-1367.

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.

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.

Farmahin, R., A. Williams, et al. (2017). "Recommended approaches in the application of toxicogenomics to derive points of departure for chemical risk assessment." Arch Toxicol **91**(5): 2045-2065.

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

Foster, W. R., S. J. Chen, et al. (2007). "A retrospective analysis of toxicogenomics in the safety assessment of drug candidates." Toxicol Pathol **35**(5): 621-635.

Grillo, M. and F. L. Margolis (1990). "Use of reverse transcriptase polymerase chain reaction to monitor expression of intronless genes." Biotechniques **9**(3): 262, 264, 266-268.

Grimm, F. A., Y. Iwata, et al. (2016). "A chemical-biological similarity-based grouping of complex substances as a prototype approach for evaluating chemical alternatives." Green Chem **18**(16): 4407-4419.

Gwinn, W.M., Auerbach, S.S. (2020) "Evaluation of 5-day in vivo rat liver and kidney with high-throughput transcriptomics for estimating benchmark doses of apical outcomes. Toxicol Sci **176**(2): 343-354.

HK, H. (2004). Toxicogenomics: Principles and Applications. New Jersey, ohn Wiley & Sons, Inc. Hoboken.

Kienhuis, A. S., M. C. van de Poll, et al. (2009). "Parallelogram approach using rat-human in vitro and rat in vivo toxicogenomics predicts acetaminophen-induced hepatotoxicity in humans." Toxicol Sci **107**(2): 544-552.

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

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.

NTP (2018). NTP Research Report on National Toxicology Program Approach to Genomic Dose-Response Modeling: Research Report 5 [Internet].

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. (2015). "Guidance Document on Revisions to OECD Genetic Toxicology: Test Guidelines ", from https://www.oecd.org/env/ehs/testing/Draft%20Guidance%20Document%20on%20OECD%20Genetic%20Toxicology%20Test%20Guidelines.pdf.

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

OECD. (2016. a). "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. (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. (2018). "OECD Harmonised Template 71: Genetic toxicity in vivo (Version [5.]-[August 2018]) ", from https://www.oecd.org/ehs/templates/harmonised-templates-health-effects.htm.

OECD. (2018.c). "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. (2018.a). "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. (2018.b). "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.

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).

Rodrigues, R. M., A. Heymans, et al. (2016). "Toxicogenomics-based prediction of acetaminophen-induced liver injury using human hepatic cell systems." Toxicol Lett **240**(1): 50-59.

Rooney, J. P., N. Ryan, et al. (2017). "From the Cover: Genomic Effects of Androstenedione and Sex-Specific Liver Cancer Susceptibility in Mice." Toxicol Sci **160**(1): 15-29.

Rowlands, J. C., R. Budinsky, et al. (2013). "A genomics-based analysis of relative potencies of dioxin-like compounds in primary rat hepatocytes." Toxicol Sci **136**(2): 595-604.

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.

Venkatratnam, A., J. S. House, et al. (2018). "Population-based dose-response analysis of liver transcriptional response to trichloroethylene in mouse." Mamm Genome **29**(1-2): 168-181.

Vomelova, I., Z. Vanickova, et al. (2009). "Methods of RNA purification. All ways (should) lead to Rome." Folia Biol (Praha) **55**(6): 243-251.

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/.

Wilfinger, W. W., K. Mackey, et al. (1997). "Effect of pH and ionic strength on the spectrophotometric assessment of nucleic acid purity." Biotechniques **22**(3): 474-476, 478-481.

Zhao, W., X. He, et al. (2014). "Comparison of RNA-Seq by poly (A) capture, ribosomal RNA depletion, and DNA microarray for expression profiling." BMC Genomics **15**: 419.

3. Data Acquisition and Processing Reporting Module (DAPRM) for Microarray Data

30.   Microarrays measure the abundance of a defined set of transcripts via labelling and hybridization to an array of complementary probes printed on a solid surface. The capacity of microarrays to detect simultaneously 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.

31.   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 hybridization), details of raw data acquisition and format and how normalisation, data filtering and outlier identification and removal was conducted.

3.1. Technology

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

33.   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 hybridization 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 hybridization that limits the accuracy of expression measurements, mainly for transcripts present in low abundance.

34.   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 targeting a region of a transcript used to detect the presence of nucleotide sequences through hybridization of complementary labeled target to single-stranded nucleic acid. 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.

35.   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.   Type and version of the platform, manufacturer's name (e.g., Affymetrix U133 Plus 2.0 Array)

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

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

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

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

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

3.7.   Control console operating system

3.8.   Any other relevant information

3.2. Transcriptomics Experimental Design

36.   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 hybridized to microarrays), and analytical processes applied to assess sample and hybridization data quality.

3.2.1. RNA Processing

37.   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 *in vitro* or *in vivo*. 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.

38.   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 indexes, 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, Bioanalyzer/ TapeStation readouts, and RT-PCR. (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:

Describe method used for preparation of RNA samples.

3.9.   RNA extraction

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

    b. Extraction and purification techniques

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

    d. Storage conditions

3.10.   Quantification and Qualification of RNA

    a. RNA quality

    b. RNA yield

    c. Performance metrics

- A260
- RNA/gr of tissue
- RNA/10^6 cells
- A260/280 ratio
- RIN for eukaryotic RNA
- PERM number (Chung et al. 2016) for eukaryotic RNA extracted from formalin-fixed paraffin embedded tissues
- Other….

3.2.2. Sample Pooling (if applicable)

39.   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:

3.11.   Whether any sample pooling has occurred (samples being combined into one hybridization) – Yes/No

3.12.   Reason for pooling of samples

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

3.14.   Any other relevant information

3.2.3. Batch processing of samples (if applicable)

40.   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 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 equal number of samples from each group in order to minimize batch effects and maximize biological differences. Please report the following parameters:

REPORT:

3.15.   Number of batches

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

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

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

3.2.4. Linear amplification, labelling, cDNA/cRNA preparation

41.   Prior to hybridization 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 hybridization. Labelling efficiency can be checked before hybridization. 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:

3.19.   Labelling protocol (i.e. whether it is manufacturer specific, modified or custom – describe the protocol in detail)

3.20.   Labeling description. Specify whether:

   a. the label is added during the reverse transcription step, or following amplification

   b. the label is fluorescent (one machine uses radiolabels)

   c. the labelling is direct or indirect

   d. the labelling occurs before hybridization (two-channel arrays) or after (single-channel arrays)

   e. Any other relevant information

3.2.5. *Hybridization*

42.   Hybridization is the process of complementary binding between a labeled 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 hybridization outcome. In addition, systematic errors in microarray production can lead to variability in hybridization (e.g. defective lot of microarrays). Conditions for microarray hybridization should be described as well as any experimental design features used to evaluate hybridization performance.

43.   Microarray platforms can be designed in a variety of ways. Multiple arrays may be present on a glass slide and samples can be labeled in 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 labeled on specific arrays, including dates of hybridization. This information is also essential for understanding whether the hybridization design can introduce potential confounding variables (e.g., lack of randomized assignment of samples across batches, microarray slides, hybridization dates, etc.). The association of samples and label should be reported.

REPORT:

3.21.   Hybridization protocol (i.e. whether it is manufacturer specific, modified or custom – describe the protocol in detail)

3.22.   In addition to the specific protocol used in the experiment, a full description of the design must be provided,

   a. Dates of hybridization for slides/microarrays

   b. Association of physical unit identifiers with sample IDs

   c. Hybridization oven temperature

   d. Hybridization time (e.g. 12 hours)

3.2.6. Scanning of Microarray

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

REPORT:

     a. Scanning protocol

     b. Name of instrument manufacturer

     c. Type of instrument (Scanner type)

     d. Scan rate

     e. Scanning time

     f. Other pertinent settings

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

     h. Any other relevant information

3.2.7. Array Quality Control Metrics/Criteria

45.   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.

46.   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 hybridization, uniformity of background hybridization, dynamic range of gene expression and percentage of detectable genes (> 25%) and (if applicable) linearity of signal for spike-in RNAs.

47.   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 fluid are often used to assess sample quality. For example, adequate expression of these housekeeping genes reflects suitable hybridizable 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 normalize the quantity of target RNAs.

48.	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 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 quantitation. 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 visualize drift or shift. In addition, a "no template" control can be used to evaluate background signal and contamination by stray nucleic acids.

49.	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 auto fluorescence). Finally, combinations of spiked molecules have been proposed as a way of tracking sample identity through specimen preparation and analysis.

50.	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:

3.23.	Quality Control Approach / Sample Type(s) (e.g. housekeeping genes, spiked controls, exogenous controls, no template controls, etc.)

   a.	Type

   b.	Source

3.24.	Quality Control Applicability

   a.	What aspects of sample processing are they designed to evaluate? (e.g. efficiency of amplification or labeling, hybridization, etc.)

   b.	What level of the experimental design are they intended to address ? (e.g. individual sample quality, sample batch quality)

3.25.	Quality Control Performance Metrics

3.26.	Quality Control Accept / Reject Criteria

3.27. Technical and Experimental Replicates

    a. Intended use for quality control

    b. Summary / aggregation strategy across replicates

3.28. Evaluation Metrics for Spike-In Controls

3.29. Reproducibility for Replicated Probes (e.g. the median CV of replicate probes)

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

3.31. Quality Control Results

    a. Pass / Fail score for each sample for each quality control metric

    b. Indicate which samples were included or excluded based on quality control results.

3.32. Any other relevant information

3.3. Specification of Raw Data

51.   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 hybridized 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.

3.3.1. Feature Extraction Software and Outputs

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

REPORT:

3.33. Feature Extraction Software:

    a. Name and version of feature extraction software.

    b. Name of grid template/array design file.

    c. Name of protocol used.

    d.   Other pertinent settings.

3.34.   Output files:

    a.   Type of files generated (e.g. raw intensity files, QC metrics, QC report);

    b.   File extensions (.txt, .pdf);

    c.   Naming convention;

    d.   Description of association of quantification matrices with raw data;

    e.   Experimental metadata file;

    f.   Retention as part of experimental record? (YES / NO);

    g.   Archiving location

    h.   Any other relevant information

3.3.2. Description of Raw Data

53.   Different companies use their proprietary feature extraction algorithm to generate different types of raw data. The quantification of raw data starting from analyzing the pixels of colored 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 features if an outlier 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. Any other relevant information.

3.3.3. Availability of Raw Data

54.   Raw unprocessed gene expression data should be accessible to 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 (for example, .txt, .xls, etc.)

3.4. Data Normalisation

55.   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 quantitation 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 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. In particular, if a single gene is used for the normalization, 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 analysis.

3.4.1. Normalisation by manufacturer process

56.   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 data need to be provided.

REPORT:

a. Manufacturer of the product used for normalization

b. Manufacturer's normalisation method used

c. Any deviance from the manufacturer method

d. Software package used and date/version

3.4.2. Data normalisation by mean and median centering within sample data sets (non scaling)

57.   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 centered by dividing by this value such that the

mean or median of the sample data set is 1 (or 0 if the data are log2 transformed). The process does not change the distribution of the data. When plotted on a box plot if median centered, then the centers 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.

58.   This process can be summarised:

For each column (j) of data (X) where the columns are the data from each individual sample; for X1 to n, where n is the total number of data columns, compute mean or median called Xj. For each data element of Xj divide this by the mean or median of Xj.

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 procedure 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 log transformed, 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. Any other relevant information

3.4.3. Data normalisation by mean centering across sample data sets (scaling)

59.   This process differs from that used in 3.4.2 in that one sample data set is chosen as the comparator, and the mean 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 as the reference sample data set chosen. This method only applies to use of the mean; for the median, it is automatic as all the data sets are adjusted to a median of 1.

60.   This process can be summarised:

For each column (j) of data (X) where the columns are the data from each individual sample; for X1 to n where n is the total number of data columns compute mean for one sample data set Xj. For each of the other data sets calculate their mean and divide by mean of the sample data set Xj. Multiply each of the data elements by this ratio.

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 log transformed, 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 (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.   Any other relevant information

3.4.4. LOWESS normalisation

61.   This process of normalisation takes into account signal intensity in the normalisation, which is different from 3.4.2 or 3.4.3 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 hybridization.

62.   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 would then be 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.   Any other relevant information

3.4.5. Quantile normalisation

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

64.   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 procedure that were applied

    e. If the data were log transformed, 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. Any other relevant information

3.4.6. Robust Multiarray Averaging (RMA)

65.   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 log2-transformed of the perfect match values. RMA normalizes 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 procedure that was 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 log transformed, 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. Any other relevant information

3.4.7. Other normalisation methods

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

3.4.8. Availability of normalized data

Normalized gene expression data should be accessible to public/researchers to facilitate reproduction of data analysis steps and final results. Normalized data is often housed in public repositories such as GEO or ArrayExpress along side 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 normalized data to facilitate data retrieval.

   c.   Format of submitted data (for example, .txt, .xls, etc.)

   d.   Description of raw data table(s).

3.5. Data Filtering

67.   This module specifies the information a regulatory scientist needs to understand the types of filtering that were applied to a 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. Reporting requirements for two methods that are the most commonly used have been included in the TRF, as well as an option to report on other approaches.

3.5.1. Filtering by Signal Intensity

68.   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:

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

   a.   Background distribution calculation

   b.   Local background from the quantification software

   c.   Negative control spots

    d. Signal intensity distribution

    e. Any other relevant information

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

    a. Statistical test between the probe foreground and background intensity

    b. A statistic estimated from the negative control spots

    c. Quantile from the signal intensity distribution

    d. Any other relevant information

3.5.2. Filtering by Probe Level Variability

69.   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:

3.37.   How the probe level variability was measured

    a. Variance or coefficient of variation of the technical replicates

3.38.   How the probe level variability cut-off was determined and applied.

    a. Using a quantile estimate from the distribution of all probe level estimates

3.39.   Any other relevant information

3.5.3. Other filtering methods

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

3.6. Identification and Removal of low quality or outlying data sets

71.   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 identification of effects due to test article treatment can be maximized. 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 than when compared to rest of the samples within a group. Outliers can result from variability in experimental steps, poor hybridizations, data acquisition or scanning errors such as misaligned grids or unique biological response. The primary purpose of

removing an outlier is to decrease the leverage of the sample on any downstream analyses and the within group variance so that the statistical difference between two comparison groups can be maximized. 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:

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. Any other relevant information

3.7. 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, Vaníčková 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.

4. Data Acquisition and Processing Reporting Module (DAPRM) for RNA Sequencing and Targeted RNA Sequencing Data

72.   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 (e.g., 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 (both for wet and dry lab) needs to be fully documented and reported.

73.   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.

4.1. Technology

74.   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.

75.   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., polyA enrichment, rRNA depletion...)
- 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

76.   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.

77.

REPORT:

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

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

4.3.   Flowcell used (type and catalogue number)

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

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

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

4.7.   Any other relevant information

4.2. Transcriptomics Experimental Design

78.   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.

4.2.1. RNA Processing

79.   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 utilize 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/tapestation or RNA Quality Score (RQS, LabChip).

80.   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.

81.   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 polyA enrichment to isolate the messenger RNA 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 do 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 polyA selection) and miRNA in one run.

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

83.   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 tissues 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.

84.

REPORT:

Describe method used for preparation of RNA samples if relevant.

4.8. RNA extraction

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

    b. Extraction and purification techniques

    c. Procedures for specific RNA enrichment or depletion procedure (e.g., polyA enrichment for mRNA, Ribosomal RNA depletion...)

    d. Storage conditions

4.9. Quantification and Qualification of RNA

    a. Tool for RNA assessment (e.g, Nanodrop, QuBit, Bioanalyzer…)

    b. RNA quality

- A260

- A260/280 ratio

- RIN for eukaryotic RNA

- PERM number (Chung et al. 2016) for eukaryotic RNA extracted from formalin-fixed paraffin embedded tissues

- Other….

    c. RNA quantity

- RNA/gr of tissue

- RNA/10^6 cells

- Other…

4.2.2. 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 TRF 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 polyA selection strategy has been applied to low RIN samples).

4.10.   Library preparation applied (full name, manufacturer and catalogue number of the kit used)

     a.   manual library preparation or automated system (if yes, which automation system)

     b.   fragmentation strategy (if applicable)

     c.   probe manifest (if applicable)

     d.   number of PCR amplification cycles

     e.   targeting probe annealing, ligation and PCR amplification steps (if applicable)

     f.   any other information

4.2.3. Sample Pooling

85.   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 range of strategies for indexing, with indexes 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.

86.   When considering an indexing strategy, the decision will depend on 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.

87.

REPORT:

4.11.   Whether any sample pooling has occurred (samples being combined into one sequencing run) – Yes/No

4.12.   Indexing Strategy (e.g. single/dual barcodes, sizes)

4.13.   Number of samples pooled per sequencing unit (lane or flowcell).

4.14.   Table of barcodes assigned to each sample

4.15.   Any other relevant information

4.2.4. Batch processing of samples (if applicable)

88.   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 minimize batch effects and maximize biological differences (or to carry out batches in replicate sets). Please report the following parameters:

REPORT:

4.16.   Number of batches

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

4.18.   Method of sample assignment to different batches (for example, simple randomization, randomized block design etc.)

4.2.5. Wet Lab quality control

89.   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.

90.   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.

91.   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 minimized. 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.

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

REPORT:

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

4.20.   Use of spiked-in control

    i.   Type of spike-in used (sequences)

    ii. Amount of sequence used

    iii. Source of spike-in (manufacturer, cat. Number…)

    iv. Step of introduction of the spiked control in the protocol

4.21.   Use of Unique Molecular Identifiers (UMIs)

    i.   Type of UMIs

    ii. Source of UMIs (manufacturer, cat. Number…)

4.22.   Any other relevant information

4.2.6. Sequencing Quality Control Metrics/Criteria

93.

94.   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. Others 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 flowcell and the % PF, together with a high number of duplicated rate, could indicate a clustering issue (under or over clustering).

95.

REPORT:

4.23.   Quality control standard (e.g. PhiX genome)

    c.   Type of standard

    d.   Source of standard (manufacturer, cat. Number, etc.)

    e.   Quantity used

4.24.   Sequencing quality metrics

    f.   Number of clusters passing filtering

    g.   Average quality score

    h.   Any other relevant information

4.3. Analysis of raw data

96.   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 step, such as demultiplexing with error correction, automated adapter removal, splitting into quality scores, etc).

97.   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.

4.3.1. Base calling

98.   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 samples (R1 and R2, or forward and reverse). It is also common to run a single sample on different sequencing lanes of a flowcell, which will then produce a file (or two if paired-end reads) 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.

99.

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

REPORT:

4.25.   Base calling software:

    e.   Name and version of base calling software.

    f.   Quality score version (e.g., Phred33)

    g.   Other pertinent information.

4.26.   Output files:

    i.   Naming convention (or sample ID metadata)

    j.   Experimental metadata file;

    k.   Any other relevant information

4.3.2. Availability of Raw Data

101.   Raw unprocessed gene expression data should be accessible to public/researchers 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:

4.27.   Output files:

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

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

   f. Format of submitted data (for example: .fastq, .bam, etc.)

4.3.3. Raw data filtering

102.   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).

103.   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.

104.   Please report the following items:

REPORT:

4.28.   Data filtering and trimming

   a. Minimum read count

   b. List of samples excluded and rationale (thresholds) for exclusion

   c. Trimming algorithm/software (e.g., FastQC v0.11.8, Trimmomatic v0.39)

   d. Trimming parameters (e.g. leading 3, trailing 3, sliding window 4:15)

   e. Any other relevant information

4.3.4. Alignment of sequence

105.   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:

4.29.   Alignment

    f. 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…).

    g. Mapping software (e.g., STAR v2.7.0a)

    h. Mapping parameters (gene or transcript level, number of mismatch, gap penalty…)

    i. Any other relevant information

4.3.5. Gene quantification

106.   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 one run (e.g., "Salmon" for transcript level), and should then be reported in both section (1.3.4. Alignment and 1.3.5. quantification).

107.   Please report the following items:

REPORT:

4.30.   Quantification:

    a. Source and version of the software (e.g. Rsem v1.1.17).

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

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

    d. Description of method for summarizing transcript or probe counts to gene counts (if applicable)

    e. Other relevant information

4.4. Data Normalisation

108.   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.

109.   The normalisation is usually applied using an R package and some statistical methodologies incorporate the normalization and experimental design as part of the method (e.g., DeSeq2).".

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

REPORT:

4.31.   Normalisation of the raw count

e.   Normalisation method applied

f.   Package and version used. Provide link if possible.

g.   Factor used as design (if applicable)

h.   Any other information

4.5. Post-normalisation Data Filtering

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

111.   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 maximized. 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 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.

112.   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 outlying dataset:

REPORT:

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

i.   Threshold, if any

j.   Processing step where exclusion occurs

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

l.   Any other relevant information

113.

4.5.2. 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 (or even 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:

4.32.   Specific gene filtering

f.   Low read count filtering step (if any)

g.   High variation among replicate filtering step

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

e.   Any other relevant information

4.6. References

Chung, J.Y., Cho, H., et al. (2016) "The paraffin-embedded RNA metric (PERM) for RNA isolated from formalin-fixed paraffin-embedded tissue." <u>Biotechniques</u> 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.

5. Data Acquisition and Processing Reporting Module (DAPRM) for Quantitative Reverse-Transcription PCR Data

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 TRF.

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).

5.1. 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 qPCR amplification reaction. This includes the RNA extraction protocol, assessment of RNA quantity and quality, and the reverse transcription of RNA into cDNA.

5.1.1. RNA Extraction

The extraction of RNA for qPCR experiments requires many of the same considerations as other transcriptomic technologies. To review these considerations please refer to the RNA extraction section of microarray TRF module (Section 3.2.1).

      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)

5.1.2. Quantification and Qualification of RNA

    a. Description of instrument(s) used

    b. Protocol or kit and any modifications

    c. Purity (A260/A280, and optionally A260/230)

    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)

5.1.3. Reverse Transcription of RNA to cDNA

    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 hex, 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

5.2. qPCR Assay Design

This section describes which qPCR 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

5.2.1. qPCR Protocol

    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), Mg2+, 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 catalog 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)

5.2.2. Gene Target Information

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?

5.2.3. Oligonucleotides and Amplicon

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

5.2.4. Assay Validation

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

5.3. Data Analysis

This section describes the details that should be reported for the analysis of qPCR data. This includes information about how the Cq values are determined, quality control results, pre-processing of low/high-signal samples, and the data normalization method employed. The most common method for data normalization for qPCR experiments is to normalize to a set of so-called *reference genes* (often also referred to as *housekeeping genes*). As such, this TRF provides more details on the reference gene normalization method. However, as multigene qPCR 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 normalization results can be replicated (see microarray normalization section 3.4). XXXXXXX

5.3.1. Data Acquisition and Quality Control

a. qPCR 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 and stage (reverse transcription or qPCR) of technical replicates

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

ocessing

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 normalization (if any)

5.3.3. Data Normalization

a. Description of normalization 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 normalization calculations, or assumed to be equal across targets

e. **RG method:** formula used for normalization

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

5.4. References

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

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

6. Data Analysis Reporting Module (DARM) for Discovery of Differentially Abundant Molecules – Univariate Methods

114.   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 (DAM); i.e., for transcriptomics, to identify differentially expressed genes (DEG). Henceforth DEG are referred to as DAM.

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

6.1. Software Documentation

116.   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 DEGs / discover 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 DEGs / discovering 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:

REQUIRED:

6.1.   **Software:** The name of the software/analysis package along with version.

6.2.   **Operating System:** In some cases, modeling of data can be impacted by the operating system that the software is run on, therefore it is important to document this information.

6.3.   **Additional Libraries Used:** A list of libraries used for analysis along with version for each.

OPTIONAL:

6.4.   **Software Availability:** If the software is open source, a hyperlink to the software and source code (if available) is desireable.

6.2. Contrasts for Which Differentially Abundant Molecules were Identified

117.   To ensure clarity, the contrasts (or group comparisons) that were performed to identify the DEGs / discover 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 12hr vs

field control 12hr" is more clear than stating "treated time 12hr 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:

REQUIRED:

6.5.   **Contrasts :** A table or listing that must include each factor and level within each factor that are being considered in the DEG / DAM analysis.

6.3. Assay Experimental Design

118.   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 DEG / 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:

REQUIRED:

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

  a.   **Group Sizes:** A table or listing of the number of samples used for identifying DEGs / discovering DAMs 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 not 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).

6.4. Statistical Analysis to Identify Differentially Abundant Molecules

119.   Today there are a myriad ways for identifying DEGs / discovering 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:

 REQUIRED:

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

a. **Statistical Approach:** The name and description of statistical approach

b. **Data Transformation:** Clearly state all data transformations (e.g., log2 transformation) that are performed in the course of analysis. Alternatively, indicate if no data transformations were performed to ensure clarity.

c. **Effects Models:** If using a general linear model, general linear mixed model, ANOVA, or something similar, specify the effects being modeled including all fixed and random effects, as well as their interactions, if any, and any nesting.

d. **Modeling 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).

e. **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 DEGs / discover the DAMs.

f. **Decision Criteria:** if using a *p*-value criteria to identify DEGs / discover the DAMs then specify the nominal alpha value and the p-value threshold. If a multiple-testing correction 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.

g. **Other:** If using another approach for ranking and prioritization of DEGs / DAMs, specify the procedure clearly such that anyone could replicate the work should it be necessary.

6.5. Outputs

120. Because there are numerous approaches for the analysis and modeling of DEG / 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:

REQUIRED:

6.8. **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 DEG / 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.

6.6. 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.

7. Data Analysis Reporting Module (DARM) for Benchmark Dose Analysis and Quantification of Biological Potency

121.　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.).

122.　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. Judgment relating to the suitability of the analysis plan is left to the end-user.

7.1. 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:

REQUIRED:

7.1.　**Software:**　The name of the software/analysis package along with version.

7.2.　**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.

7.3.　**Additional Libraries Used:** A list of libraries used for analysis along with version for each.

OPTIONAL:

7.4.　**Software Availability:** If the software is open source, a hyperlink to the software and source code (if available) is desireable.

7.2. Description of the Data to be Modeled

In 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/normalized 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 maximize 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 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 emphasize 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:

REQUIRED:

7.5.   **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.

7.6.   **Annotation of 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 modeling 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.

7.7.   **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.

7.3. Prefiltering of Data Prior to Dose-Response Modeling

123.   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:

REQUIRED:

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

   a. **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 pair-wise 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)

   b. **Statistical Threshold Applied:** Indicate the statistical threshold that was applied when performing any statistical test to prefilter the data.

   c. **Statistical Multiple Testing Correction Method:** Because of the large number of replicate statistical tests that are performed when analyzing an omics data set, an adjusted statistical threshold is often applied that corrects for 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).

   d. **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.

   e. **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).

7.4. Benchmark Dose Modelling

124.   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 estimate. 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 modifies, 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:

REQUIRED:

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

   a. **BMD Modeling 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.

   b. **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.

   c. **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.

   d. **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.

   e. **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.

   f. **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.

   g. **Number of Fitting 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.

   h. **Confidence interval of the BMD (a.k.a., Confidence level):** This modeling 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.

   i. **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.

   j. **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.

k. **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)).

l. **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.

7.5. Determination of Biological Entity or Biological Set Level BMD Values

125.   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). Organizing the MSD into recognized biological entities helps in contextualizing 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.

126.

REPORT:

REQUIRED:

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

a. **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.

b. **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.

c. **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.

d. **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 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.

127.

7.6. Outputs

7.11. **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.

7.7. References