Archives of Toxicology (2020) 94:371–388
https://doi.org/10.1007/s00204-020-02656-y

REVIEW ARTICLE

Prospects and challenges of multi-omics data integration in toxicology

Sebastian Canzler[1] · Jana Schor[1] · Wibke Busch[1] · Kristin Schubert[1] · Ulrike E. Rolle-Kampczyk[1] · Hervé Seitz[4] · Hennicke Kamp[3] · Martin von Bergen[1,2] · Roland Buesen[3] · Jörg Hackermüller[1]

Received: 8 November 2019 / Accepted: 29 January 2020 / Published online: 8 February 2020

Abstract

Exposure of cells or organisms to chemicals can trigger a series of effects at the regulatory pathway level, which involve changes of levels, interactions, and feedback loops of biomolecules of different types. A single-omics technique, e.g., transcriptomics, will detect biomolecules of one type and thus can only capture changes in a small subset of the biological cascade. Therefore, although applying single-omics analyses can lead to the identification of biomarkers for certain exposures, they cannot provide a systemic understanding of toxicity pathways or adverse outcome pathways. Integration of multiple omics data sets promises a substantial improvement in detecting this pathway response to a toxicant, by an increase of information as such and especially by a systemic understanding. Here, we report the findings of a thorough evaluation of the prospects and challenges of multi-omics data integration in toxicological research. We review the availability of such data, discuss options for experimental design, evaluate methods for integration and analysis of multi-omics data, discuss best practices, and identify knowledge gaps. Re-analyzing published data, we demonstrate that multi-omics data integration can considerably improve the confidence in detecting a pathway response. Finally, we argue that more data need to be generated from studies with a multi-omics-focused design, to define which omics layers contribute most to the identification of a pathway response to a toxicant.

Keywords Multi-omics · Toxicology · Chemical exposure · Risk assessment · Data integration

Introduction

Exposure of cells or organisms to chemicals triggers a series of effects at the molecular level. Regulatory pathways involved in such responses exhibit changes of levels, interactions, and feedback loops of biomolecules of different types that are active in complex networks. Omics technologies allow to interrogate a significant fraction or close to all biomolecules of a particular type in an untargeted manner. From a toxicological perspective, omics techniques, therefore, allow to efficiently and accurately generate relevant information on substance-induced molecular perturbations in cells and tissues that are associated with adverse outcomes. The European Centre for Ecotoxicology and Toxicology of Chemicals (ECETOC) has, therefore, conducted a series of workshops to evaluate the promises and challenges of omics techniques in chemical risk assessment (ECETOC 2008, 2010, 2013; Buesen et al. 2017). The most recent workshop concluded that omics techniques contribute to answering relevant questions in risk assessment including (i) the classification of substances and definition of similarity, (ii) the elucidation of the mode of action of substances, and (iii) the identification of species-specific effects and the demonstration of human health relevance (Sauer et al. 2017). However, participants also identified a need to increase the

Sebastian Canzler and Jana Schor contributed equally to this work.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00204-020-02656-y) contains supplementary material, which is available to authorized users.

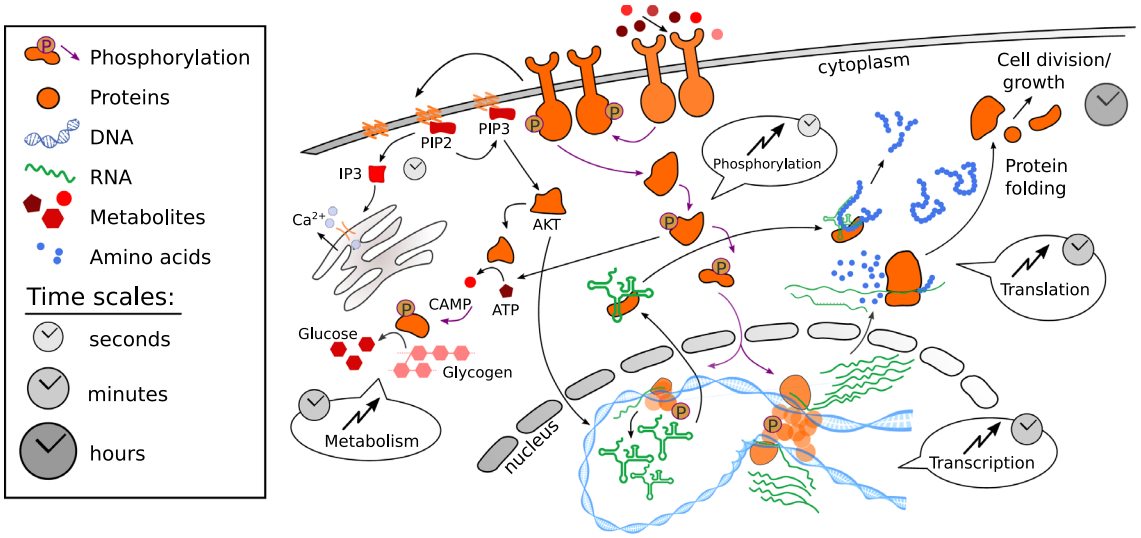
✉ Jörg Hackermüller
joerg.hackermueller@ufz.de

[1] Helmholtz Centre for Environmental Research - UFZ, 04318 Leipzig, Germany

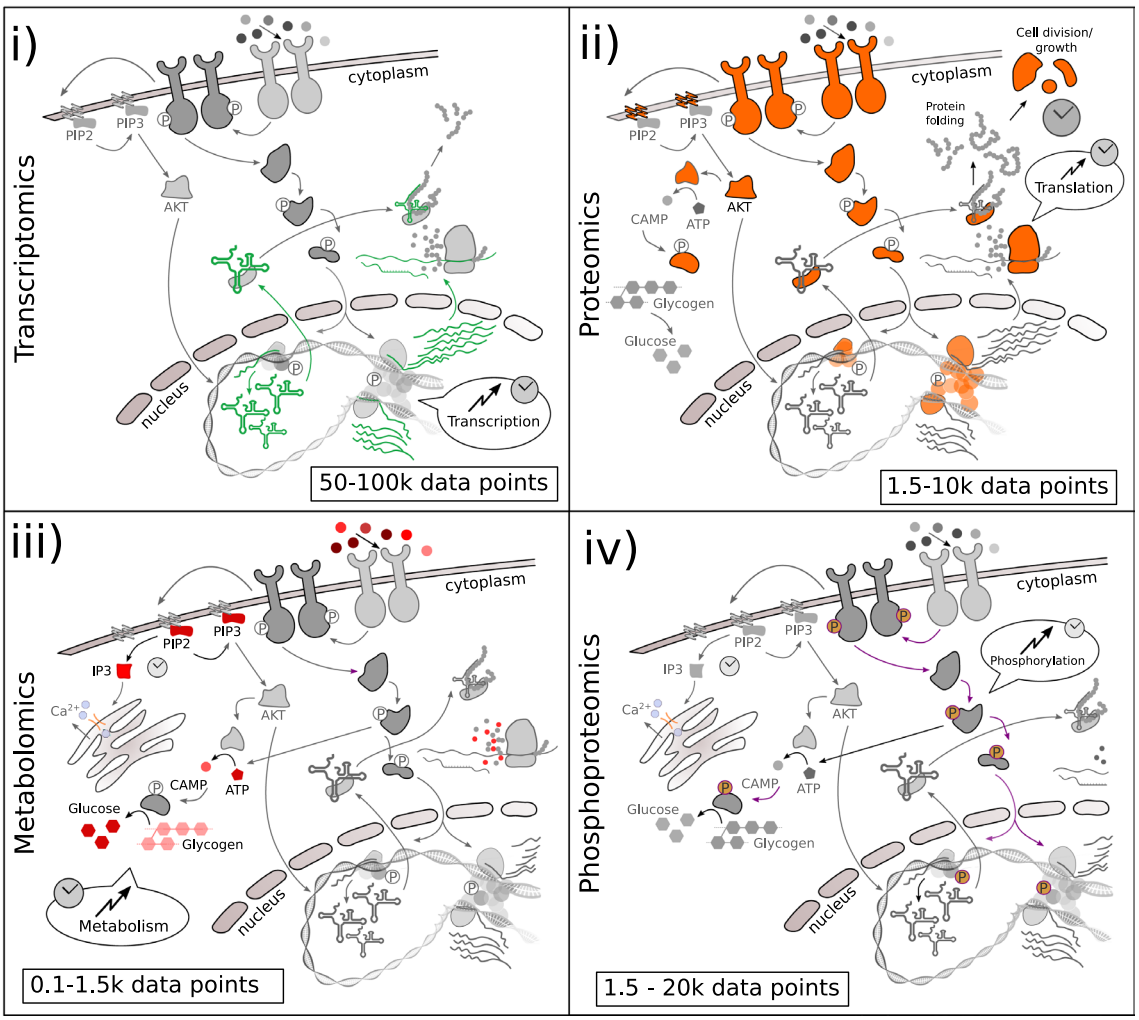
[2] University of Leipzig, Institute of Biochemistry, Brüderstraße 34, 04103 Leipzig, Germany

[3] Experimental Toxicology and Ecology, BASF SE, 67056 Ludwigshafen, Germany

[4] Institut de Génétique Humaine UMR 9002 CNRS-Université de Montpellier, 34396 Montpellier Cedex 5, France


Phosphorylation
Proteins
DNA
RNA
Metabolites
Amino acids
Time scales:
seconds
minutes
hours
cytoplasm
Cell division/
growth
PIP2
PIP3
IP3
Ca2+
AKT
Phosphorylation
Protein
folding
CAMP
ATP
Translation
Glucose
Glycogen
Metabolism
nucleus
Transcription

(a)


i)
Transcriptomics
cytoplasm
PIP2
PIP3
AKT
nucleus
Transcription
50-100k data points
ii)
Proteomics
Cell division/
growth
Protein
folding
Translation
CAMP
ATP
Glycogen
Glucose
1.5-10k data points
iii)
Metabolomics
IP3
Ca2+
Metabolism
0.1-1.5k data points
iv)
Phosphoproteomics
Phosphorylation
1.5 - 20k data points

(b)

◂**Fig. 1** **a** Sketch of a regulatory pathway illustrating the diversity of biomolecules involved. The flow of information as a response to an event is not confined to a single-omics layer. Response times vary considerably within and between omics layers as indicated by the clock symbols (details, see Fig. 2a). **b** Subset of the components of a pathway that are detected by a specific omics approach and number of data points returned during measurements for (i) phosphoproteomics, (ii) transcriptomics, (iii) proteomics and (iv) metabolomics

reproducibility of omics data acquisition and data analysis, and to define best practices (Buesen et al. 2017).

A single-omics technique, e.g., transcriptomics, will detect biomolecules of one type, also called one layer, and thus captures changes only for a small subset of the components of a particular pathway. Therefore, applying single-omics analyses in response to a toxicant in a non-continuous design led to the identification of biomarkers for certain exposures but not to a systemic understanding of toxicity pathways or adverse outcome pathways (AOPs) in the past. A necessity to evaluate the integration of different omics layers for defining AOPs has, therefore, also been identified at the ECETOC workshop (Buesen et al. 2017). Consequently, we hypothesize that a substantial improvement in detecting the pathway response to a toxicant can be achieved by using multi-omics data in a time- and concentration-resolved design [see also Escher et al. (2017)].

The CEFIC LRI-funded project "XomeTox - evaluating multi-omics integration for assessing rodent thyroid toxicity" investigates the utility of multi-omics data integration for applications in toxicology and develops best practices aiming at regulatory application.[1] Here, we report the findings of the first phase of XomeTox, which has evaluated the current status of multi-omics data in toxicological research, the availability of such data, of repositories for exchanging these data, and of methods for integration and analysis of multi-omics data to derive where possible best practices and identify knowledge gaps. Based on published integration methods and publicly available data with a toxicological research question, we evaluate in case studies whether integration of previously separately analyzed multi-omics data improves the detection of a molecular response at the pathway level. We summarize challenges for study design and sampling and argue how to select omics layers. Finally, we discuss pitfalls and knowledge gaps and the perspective for using multi-omics data in a regulatory application.

Multi-omics in toxicology

Why relying on multi-omics in toxicology

Regulatory pathways of the cell involve a range of different (bio)molecules, which comprise very different physicochemical properties and which exhibit complex non-linear interactions (Fig. 1a). A single-omics technique will measure biomolecules of a specific type, e.g., ribonucleic acids (RNAs) in case of transcriptomics. Frequently, a single-omics technique will not even detect the entirety of biomolecules of one type but a smaller subset thereof of more similar physicochemical properties. The detection of short and long RNAs, for example, or of short peptides and proteins requires different transcriptomic or proteomic approaches, respectively. Only the use of multi-omics, also called cross-omics approaches, allows to directly detect a significant fraction of a pathways' response to chemical exposure. Employing multi-omics strategies for toxicological research questions has, therefore, repeatedly been called for in the literature (Prot and Leclerc 2012; Marx-Stoelting et al. 2015; Buesen et al. 2017; Escher et al. 2017; Dellafiora and Dall'Asta 2017), despite a cautious stand on this matter from a regulatory point of view (Tralau et al. 2015; Tralau and Luch 2015).

Which omics layers to use for toxicological research questions

Figure 1b illustrates which components of a pathway and how perturbations of a pathway, respectively, are detected by different omics layers. For detecting the pathway response to chemical exposure, we propose to focus on transcriptomics, proteomics, phosphoproteomics, and metabolomics and to consider epigenomics for specific cases and give a rational for this selection below.

Transcriptomics. Transcriptomics aims at a comprehensive detection of ribonucleic acids in the cell. A pathway response in transcriptomics is mainly detected via a known set of target genes of the pathway that are found differentially expressed. The information on association with a particular pathway is mainly available for protein-coding RNAs (mRNAs). However, a significant portion of the transcriptional response controlled by a specific pathway can consist of non-protein-coding RNAs (ncRNAs), e.g., Hackermüller et al. (2014). NcRNAs have been shown to participate in a range of different cellular pathways and seem to be particularly relevant for fine tuning the gene expression of eukaryotic cells (Cech and Steitz 2014; Dhanoa et al. 2018). For more than a decade, transcriptomics mainly relied on microarrays as a measurement technique. With the exception of tiling arrays, e.g., Otto et al. (2012), microarrays constitute a targeted detection approach, i.e., they require prior selection and knowledge of the sequence of the interrogated RNAs. Since about 10 years, transcriptomics has increasingly been relying on transcriptome sequencing (RNA-seq), which allows the simultaneous identification of transcripts, detection of isoforms, and their quantification with high dynamic range (Wang et al. 2009). RNA-seq is

[1] http://cefic-lri.org/projects/c5-xometox-evaluating-multi-omics-integration-for-assessing-rodent-thyroid-toxicity/.

per se un-targeted, but can be run in targeted mode to reduce costs and increase throughput. Compared to the other omics layers discussed in this section, transcriptomics stands out by its comprehensiveness in terms of coverage, i.e., it is estimated that 90–95% of genes are detected in mammals (García-Ortega and Martínez 2015). Depending on the research question, short-read sequencing methods (typically providing read lengths between 50 and several hundred bases) or long-read sequencing (allowing for direct detection of full-length transcripts of up to 20 to 30 kilobases), or a combination of both can be applied. Short (smaller than 200 nucleotides) and long RNAs are typically measured using separate approaches. Methods for a simultaneous detection have been described, e.g., Xiao et al. (2018), but are so far not routinely employed. For long RNA sequencing, different sample preparation strategies to avoid sequencing of ribosomal RNAs can be chosen in dependence of the biological question: Methods that capture or preferentially amplify polyadenylated transcripts, provide mainly information on messenger RNAs, while ribosomal RNA depletion-based approaches are capable of including a broader range of long ncRNAs. In general, RNA-seq experiments can provide a throughput of several thousand gigabases per run (Lowe et al. 2017).

Proteomics. Proteomics detects a pathway response highly analogously to transcriptomics via an enrichment of proteins allocated to a given pathway among differentially expressed proteins. However, beside the transcriptionally controlled effects on protein abundance, there are translational and also post-translational effects that affect the stability and are not reflected on the transcriptome level. Untargeted proteomic measurements currently mainly rely on liquid chromatography (LC), coupled to mass spectrometry (MS). Most frequently, proteins are extracted, optionally separated, digested and the proteolytic peptides analyzed by LC-MS/MS, which typically allows for the detection of 3000–5000 proteins per experiment, e.g., Schmidt et al. (2018). Quantitative detection can be improved in terms of reliability using labeling approaches such as stable isotope labeling (Ong et al. 2002). However, for animal experiments, which would require fully isotopically labeled animals as reference samples, or experiments involving primary cells that cannot be kept in culture sufficiently long, stable isotope labeling may not be feasible and label-free quantification the preferred approach. In the last years, TMT labelling became popular since it allows up to tenfold multiplexing and at the same time an isotope labelling-based quantification (McAlister et al. 2012). In contrast to the global approaches, targeted proteomics can be used by spiking selective reference peptides into a sample for the quantification of proteins of interest (Picotti and Aebersold 2012). Prior to the use of LC-MS/MS, proteomics frequently relied on 2D-gel electrophoresis, which separated proteins by mass and pI, subsequently detected differentially expressed spots using the DIGE method and identified these proteins using MS (Unlü et al. 1997). However, apart from the high manual effort and limited throughput, DIGE-based proteomics data sets are of limited use for multi-omics approaches, as quantitative information is only collected for strongly differentially expressed proteins and missing for the larger part of the proteome, which hampers integration with the other omics layers.

Metabolomics. The metabolome is different to the other *omes* discussed here, as it is a collection of chemically highly heterogeneous molecules. The metabolome is typically defined as the complete complement of all small molecule metabolites (< 1500Da) found in a specific cell, organ or organism (Wishart 2007). Thanks to initiatives like the Human Metabolome Project, the endogenous metabolome of human and animal model organisms has been mapped to a large extent. However, entries in databases for food- or environment-derived metabolites detected, e.g., in human serum are continuously growing, depending on the ever increasing sensitivity of mass spectrometry and improvements in databases. Metabolome measurements rely either on nuclear magnetic resonance (NMR)-based detection or on gas or liquid chromatography hyphenated to MS. NMR-based approaches have the advantage of the inherent capacity of quantification and constitute the gold standard for the structural elucidation of unknown metabolites. MS-based approaches outperform NMR in terms of sensitivity and can be run in an untargeted or targeted fashion. The latter interrogates between a handful and several hundred metabolites in many in-house established assay and commercial kits, respectively.

Metabolomics differs from proteomics and transcriptomics in (i) detecting the response at very different levels of a pathway and (ii) in simultaneously capturing molecules of a pathway that respond on extremely different time scales including, e.g., almost instantaneously formed second messengers like Phosphatidylinositol (3,4,5)- trisphosphate (PIP3) versus changes in components of the cell membrane with very long half lives.

Post-translational modifications of proteins. In addition to effects mediated by proteins based on their abundance, post-translational modifications (PTMs) are mostly relevant for regulating the activity of proteins. There are many different types of PTMs but the most abundant PTM is phosphorylation, a reversible, covalent modification, which is tightly regulated and particularly relevant for intracellular signaling (von Stechow et al. 2015). Phosphoproteomics aims at mapping and quantifying protein phosphorylation throughout the proteome. A major challenge of phosphoproteomics is the low stoichiometry of biologically relevant phosphorylations. Phosphopeptides thus need to be enriched before measurement. This results in a rather high demand for material compared to proteomics, which can hinder the

employment of phosphoproteomics. For quantification, metabolic or chemical labeling is frequently employed, but label-free approaches are available as well (Arrington et al. 2017). Current phosphoproteomics studies identified up to 20000 phosphorylation sites (von Stechow et al. 2015). Many signaling pathways include phosphorylation reactions and an association of these phosphorylation sites to the pathways is readily possible. Other phosphorylation sites require additional experiments to associate them to a specific kinase and thus to a pathway (Arrington et al. 2017).

Epigenomics. Chemical modifications of the DNA and of the proteins organizing the three-dimensional structure of genomic DNA are interrogated by epigenomics. In an untargeted manner, DNA methylation is today assessed by a modified genome sequencing approach (Methylome-seq). Due to the sometimes subtle effects of chemical exposure on the epigenome, Methylome-seq requires high coverage associated with still significant sequencing costs. Targeted microarray-based approaches are thus still frequently used. DNA methylation subsumes two different chemical modifications with different functional consequences—5-methylcytosine and 5-hydroxy-methylcytosine (Branco et al. 2011; Zhang et al. 2018). Most published studies do, however, not discriminate between both modifications. Histone modifications are interrogated using Chromatin-immuno-precipitation using antibodies specific for the targeted modification followed by sequencing (ChIP-seq) (Robertson et al. 2007). Typically, several modifications need to be detected to get a comprehensive picture, which is associated with several parallel ChIP-seq experiments. As ChIP-seq requires more sample material than the other epigenomic approaches discussed here, such a strategy may be hard to implement in toxicological studies where, e.g., material of particular organs is limiting. ATAC-seq is an alternative readout that does not probe chemical modifications directly, but rather one of the modifications' main consequence, DNA accessibility (Buenrostro et al. 2013).

Knowledge on the association of regulatory pathways and specific epigenetic modifications is still limited. Epigenomics is thus of limited value when a pathway-based data integration approach is used (see "Multi-omics data analysis"). Statistical integration approaches can manage epigenomic data but the biological interpretation remains challenging if detecting immediate effects and pathway responses are pivotal to the study [see also the discussion in Escher et al. (2017)]. However, epigenomic data have proven highly valuable for trans-generational studies (Jahreis et al. 2018) or when aiming to predict long-term effects of chemical exposure based on omics data of short-term exposure.

Single-cell omics. RNA-seq, methylome-seq, and ATAC-seq can also be applied to thousands of single cells in parallel (Ziegenhain et al. 2017; Hu et al. 2018). This may be a further direction for increasing the information for AOP development to advance from the cellular to tissue and organ response. Remarkable developments have recently been made for single-cell proteomics and metabolomics (Yang et al. 2019; Duncan et al. 2019). However, compared to sequencing-based approaches, the applicability of these methods is still limited by the number of cells that can be profiled, or the number of biomolecules detected in parallel. Several protocols have been published that interrogate multiple omics layers from the same single cell. These methods mainly combine genome sequencing and RNA-seq, methylome-seq and RNA-seq, or all three layers—see (Hu et al. 2018) for a recent review.

Which omics layers to chose. Selecting omics layers strongly depends on the research question and the model system. For addressing the question in focus of this review, i.e., to facilitate the detection of pathway responses to chemical exposure, omics layers should be chosen to optimally interrogate the suspected pathways of interest, in case these are known. A tool like `MOD-Finder` might be used to assemble prior knowledge for a chemical of interest (see section "Multi-omics data for toxicological research questions"). The XomeTox model problem of thyroid toxicity calls for an investigation using a multi-omics approach as the complete AOP can only be addressed at different omics levels: thyroid hormones are detected by metabolomics, while liver enzyme levels are most directly assessed by proteomics. Tumorigenesis in the thyroid or neurodevelopmental effects are potentially indicated by changes on the transcriptome level and manifest in altered protein activity level. Important switches in many pathways are determined by phosphoproteomics. Finally, non-coding RNAs, which have been shown to be important regulators in tumorigenesis [e.g, Boll et al. (2013)] can only be detected by transcriptomics, but require integration with other omics data sets for their functional annotation.

In case no prior information is available, we propose to include the four layers discussed above and to consider additionally including epigenomics if the study aims at trans-generational or long-term effects. Transcriptomics using long RNA-seq and LC-MS/MS based proteomics may convey similar information on the overall pathway response. However, transcriptomics includes information on regulatory RNAs and is much more comprehensive than proteomics. Proteomics on the other hand is much closer to the phenotype and subsumes several regulatory principles that lead to changes on the protein level, without remarkable changes on the transcriptome level. Also, even drastic alterations on the transcriptome level do not necessarily result in detectable changes on the level of the corresponding proteins. Phosphoproteomics in turn allows to directly observe changes in, e.g., kinase cascades that directly control transcript and protein abundance independent activities.

Study design and sampling for multi-omics studies

Study design. From a data analysis point of view, multi-omics studies should rely on paired samples, i.e., on samples where all omics layers per replicate are generated from one individual. Non-paired samples require any method that operates on correlations or related measures to compute these not over the individual samples but over aggregates like mean expression or fold-changes per treatment group. This aggregation generates shorter vectors compared to the data on the individual replicates and approaches like correlation network inference or many dimensionality reduction methods cannot be feasibly applied. Paired samples correspond to what Cavill and colleagues call the *split-sample study*, where a tissue sample is split for different omics layers and the *source-matched study*, where different omics layers are generated from different tissues or cells originating from the same individual. They discriminate these from the *repeated study*, which is a repetition of the experiment for the different omics layers, and the *replicate-matched study*, which generates different omics layers from different replicates of the same experiment (Cavill et al. 2016). Another argument for split-sample or source-matched designs, which is particularly relevant when animal studies are involved, is the number of individuals. Repeated study and replicate-matched study designs are clearly disadvantageous from an 3R perspective, as they multiply the number of animals required by the number of omics layers included.

Timing. Given the central dogma of gene expression, one might argue that one omics layer that is downstream of another, e.g., proteome and transcriptome should be sampled at different time points to avoid the accumulation of noise due to time biases. However, as illustrated by the clock symbols in Fig. 1a and summarized in Fig. 2a, there is not only a timing variance between omics layers but also particularly within layers. In the metabolome, for example, second messengers, like PIP3 are formed within seconds or only a few minutes (Falkenburger et al. 2010), while changes in the energy metabolism occur on time scales of minutes and hours (Wang et al. 2011) and those in lipids forming the cell membrane are affected in weeks.

Consequently, there is no optimal time point per omics layer, or optimal time-distance between omics layers for sampling. For this reason, and the reasons that different sampling time points per omics layer would (i) result in non-paired samples, associated with the difficulties in data analysis discussed above, and (ii) multiply the number of animals required by the number of omics layers, as only one time point can be obtained from one individual, we argue that the samples for the different omics layers should be generated at identical time points.

Nonetheless, the diverging time scales in the different omics layers may increase noise in the integrative analysis. One option to address this issue is to generate dense time series data starting immediately with the beginning of exposure and to explicitly model the individual temporal behavior. This drastically increases the number of individuals in an animal experiment and the effort for omics. The other option is to ignore the dynamic response subsequent to the start of exposure, use a repeated dosing scheme and sample after a time period which is significantly longer than the expected time scale of immediate pathway effects and thus look at the steady state of a chronic signaling change. For most questions in chemical risk assessment, the latter option will be preferable. Also, it integrates well with established toxicity study designs. However, focusing on a steady state implies looking at the end of an adaptation process. Thus, dense temporal modeling may be advantageous for specific questions like unraveling pathway–pathway interactions between early key events of AOPs or when bound to a single-dosing scheme.

Tissue distribution bias. Distributing sample material to different omics layers in a split sample design can be trivial in case of a cell culture-based study or blood cells provided that enough cells are available. However, when solid tissues from an animal study are included, splitting the tissue can introduce additional bias particularly due to regional differences in tissue composition, see Fig. 2b: (i) To avoid this bias, homogenized tissue can be distributed to the different omics layers. However, some omics layers are frequently used with sample pre-treatment to stabilize, e.g., labile RNAs. Pre-treatments are usually not compatible among different omics layers. Protocols for the simultaneous extraction of RNA, DNA, proteins, and metabolites exist, but are of significantly lower efficiency than specific protocols and are, therefore, not an option for situations where tissue mass is critically low (Vorreiter et al. 2016). (ii) Alternatively, cryosections of embedded tissue can be evenly distributed to the omics layers. (iii) Finally, the consistent distribution of defined organ or tissue segments to the omics layers controls bias. In a treatment–control design, inherent differences in tissue composition will be controlled for in each omics layer.

Multi-omics data for toxicological research questions

Several multi-omics studies, particularly dual-omics studies, for toxicological research questions have been published, e.g., on nanomaterials (Scala et al. 2018), PPAR agonists and antagonists (Acharjee et al. 2016), Valproic acid (van Breda et al. 2018), Pregnenolone Carbonitrile (Nagahori et al. 2017), or investigating Benzo[a]pyrene exposure (Kalkhof et al. 2015). The diXA project has assembled a set of toxicogenomics studies from diverse sources, which in part comprise multi-omics data (Hendrickx et al. 2015).

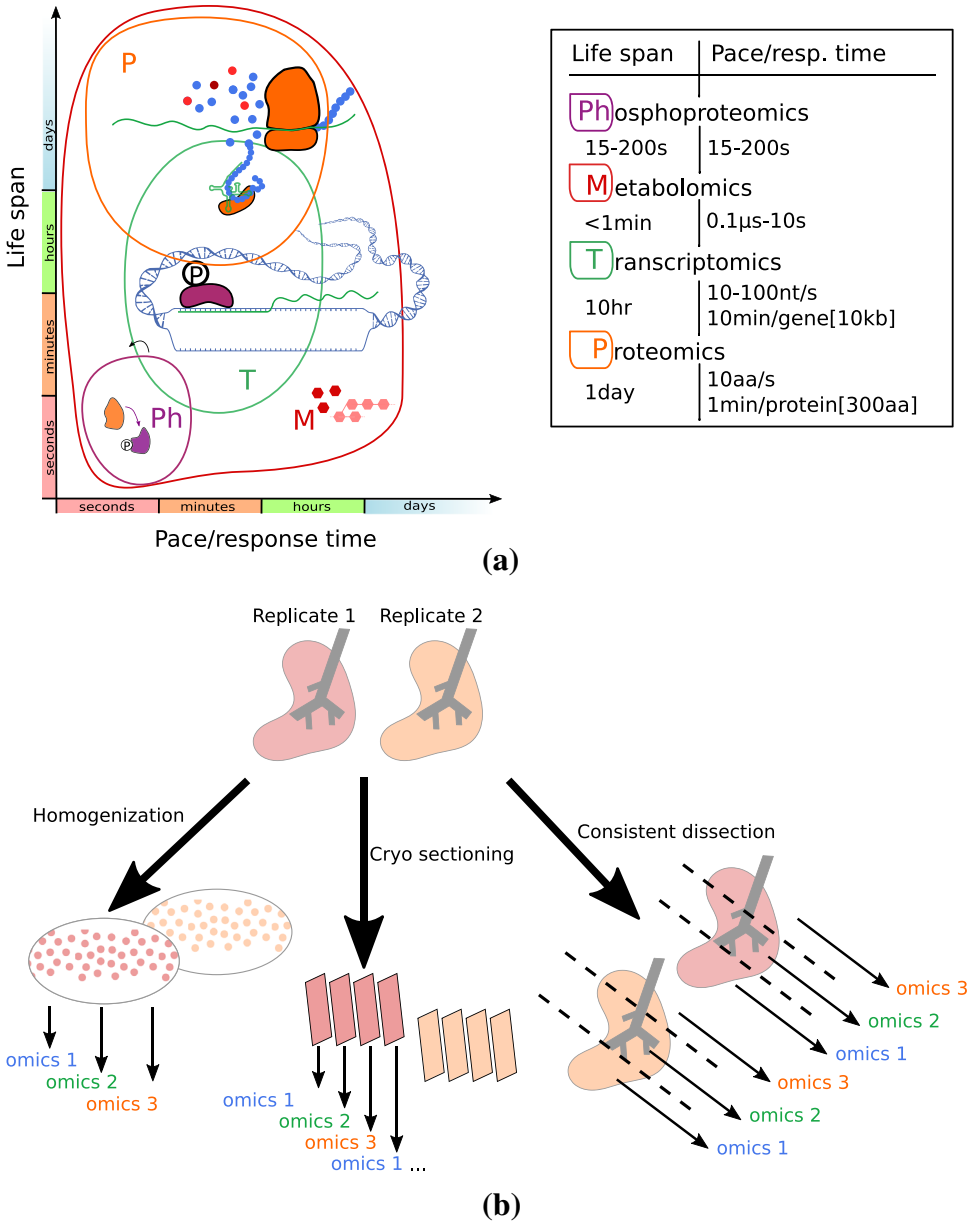


Fig. 2 **a** The response time and the life-span of biomolecules within and between single omics varies substantially, making timing a serious issue to consider when deciding on the sampling strategy during study design. **b** Different options for avoiding a bias in distributing sample material to different omics layers. A split sample design requires that replicates, e.g two organs in this figure are consistently distributed to the different omics experiments. Tissue may either be homogenized and the homogenate distributed to the omics experiments. Alternatively, tissue can be sectioned on a cryo-microtome and each *n*th section is distributed to the *n* different omics experiments. Finally, the tissue can be consistently dissected, so that one type of omics experiment is always based on the same segment of an organ

Omics data are deposited in repositories that are specific for one omics layer and several repositories may be used by the community for one layer. See Supplemental Text and Figures, Section S1, for a review of repositories and the current status and differences of data sharing for transcriptomics, proteomics and metabolomics data. Multi-omics data are usually not cross-referenced between repositories and in cases where data sets have been published separately, no corresponding manuscript has been published, or the deposited multi-omics data has not been referenced properly in the publication, identifying corresponding omics data sets is laborious. We have, therefore, developed the web application `MOD-Finder` which searches for multi-omics data sets related to a user-defined chemical of interest (Canzler et al. 2019). `MOD-Finder` is publicly available.[2]

Using `MOD-Finder` and an additional manual collection, we compiled a series of data sets for a subsequent evaluation of integration methods and derivation of best practices and pitfalls (Supplementary Table S1). However, none of the identified data sets for chemical exposure comprises paired samples in an *in vivo* setting, provides all four omics layers, we consider most relevant for detecting a pathway response, and was generated with state-of-the-art omics techniques.

[2] https://webapp.ufz.de/mod_finder.

Multi-omics data analysis

Multi-omics data integration strategies. A range of different approaches is available for the integrative analysis of multi-omics data. Here, we focus particularly on the application of these methods for detecting a response at the pathway level in a toxicological setting. For comprehensive reviews of integration methods, see (Ebbels and Cavill 2009; Buescher and Driggers 2016; Cavill et al. 2016; Bersanelli et al. 2016; Huang et al. 2017; Tarazona et al. 2018).

Several attempts have been made to categorize integration methods. Ebbels and Cavill (2009) discriminated (i) *conceptual integration* where each omics layer is analyzed separately and the researchers combine single-omics inferences to gain a comprehensive conclusion from (ii) *statistical integration* that aims at identifying statistical associations between features of different samples and omics layers, and (iii) *model-based* data integration strategies that use predefined models of a system to predict levels of molecular associations and organization. Cavill et al. (2016) further subdivided *statistical integration* into (iia) *correlation-based* approaches that infer correlative associations between features of different omics layers, (iib) *concatenation-based* methods that transform multiple layers into a single data frame prior to the analysis, (iic) *multivariate-based* methods as adaptations of standard multivariate approaches, and (iid) *pathway-based* methods that incorporate pathway annotations from various sources.

Bersanelli et al. (2016) classified integration methods into *sequential*, which largely corresponds to conceptual above, and *simultaneous* strategies, which correspond to statistical integration. Based on the methodological approach, they further discriminated *network-based versus network-free* whether or not the algorithm employs networks to model variable interactions, and *bayesian vs non-bayesian* whether or not the algorithms incorporate an *a priori* assumption about the data to compute the posterior probability distribution on the measured omics data making use of Bayes' rule.

Finally, Tarazona et al. (2018) categorized integration methods based on the incorporation of additional biological information and on using a *supervised* or an *unsupervised* approach. The first aims at predicting a certain response variable using the omics features as predictors, or at modeling a regulatory network of the underlying molecular system. The latter is mainly used to retrieve an exploratory overview of the data and to unravel potential relationships between omics layers. An overview of different integration principles and how different classification approaches relate to each other is given in Supplementary Figure S1.

Pathway integration. Few integration tools are available that incorporate pathway knowledge to interpret high-throughput datasets, e.g., via term over-representation or gene set enrichment analysis like `PaintOmics`[3] (Hernández-de Diego et al. 2018) or `IMPaLA`[4] (Kamburov et al. 2011). Approaches like `SPIA`, `Enrichnet`, or `SAFE` integrate pathway topologies, they do, however, not provide an explicit support for multi-omics data (Tarca et al. 2009; Glaab et al. 2012; Baryshnikova 2016). `PARADIGM` is a computational approach that integrates multiple genomic alterations such as copy number, DNA methylation, somatic mutations, mRNA expression and microRNA expression with the `REACTOME` database (Vaske et al. 2010). While this is a valuable selection of omics layers for oncological questions, it is of limited applicability in toxicology.

Choice of the integration approach depends on study design. As discussed above, most available multi-omics data sets are based on non-paired samples. These designs also lack an inherent, consistent ordering of samples between the omics layers, which prohibits the application of statistical integration methods. Cavill et al. (2016) argued that in this case, conceptual integration should be performed, which can provide important insights but misses associations that can only be found in a joint analysis and thus their interpretation has to be carried out with caution. Similarly, Tarazona et al. (2018) suggested to use a sequential integrative analysis or a meta analysis which is, however, associated with a substantial decrease in the amount of retrievable information.

Time series data. Due to the variability in toxicodynamics, time-resolved study designs are frequently employed in toxicology. Most integration methods treat time-resolved data as different experimental conditions, but will not take advantage of the continuous nature of this factor (Tarazona et al. 2018). Few computational approaches have recently become available that explicitly handle time in these data, like `MORE`[5] and `MathIOmica` (Mias et al. 2016), which is so far only available for `Mathematica`.

Flexibility of integration methods. Many integration methods have been developed for a very specific purpose and can only operate on few types of omics layers. Tools like `MCD` or `Conexic` focus on integrating data on genetic variation with gene expression or DNA-methylation data (Chari et al. 2010; Akavia et al. 2010). However, correlation- and concatenation-based approaches, or approaches for dimensionality reduction and clustering are flexible regarding the type of input layers; these include `moCluster` (Meng et al. 2016), `mixOmics` (Lê Cao et al. 2009; González et al. 2012), `MCIA` (Meng et al. 2014) and `MOFA` (Argelaguet et al. 2018).

Publicly available web-based tools such as `PaintOmics`, `IMPaLA`, or `OmicsNet`[6] (Zhou and Xia 2018)

[3] http://www.paintomics.org/.

[4] http://impala.molgen.mpg.de/.

[5] https://bitbucket.org/ConesaLab/more.

[6] https://www.omicsnet.ca/.

integrate biological knowledge from pathway databases. However, if only one or two pathway databases are used this can be limiting for the analysis (see *Case Studies* in Section "Case studies" for an example). Another issue that is more prominent with web-based tools is the support of different gene or protein identifiers and the mapping to the used biological knowledge base.

Reproducible computational research. Integrative analysis starting from raw multi-omics data is a complex, multi-step process. Bioinformatic tools need to be selected from a range of options and each software comes with a set of adjustable parameters which may impact the outcome substantially. Recently, a "reproducibility crisis" has been diagnosed (Baker 2016), criticizing the degree of reproducibility in biomedical research (Bustin 2014). Considering the complexity of multi-omics data analysis, independent replication of the results specifically depends on detailed reporting of the details of the analysis. Linking executable analysis code with intermediate and final resulting data is promoted as the gold standard of reproducibility by Peng (2011), while Sandve and colleagues suggest to follow the "ten simple rules for reproducible computational research" (Sandve et al. 2013). Workflow management systems, like `Galaxy`, `KNIME`, or `uap` promote the use of reproducible research principles in bioinformatic analyses (Goecks et al. 2010; Berthold et al. 2007; Di Tommaso et al. 2017; Kämpf et al. 2019). None of the currently available integration methods specifically supports reproducible research principles. However, many standalone tools could easily be wrapped for use in one of the workflow management systems. Web-based tools are particularly prone to reproducibility problems since the user is usually not able to keep track of versions of tools and databases and parameters used.

In summary, many powerful integration methods have been published. However, no single approach fulfills all requirements from a toxicological perspective, to integrate the four omics layers, we defined most informative, support time- and dose-resolved experiments and support reproducible research principles.

Case studies

We use published data to evaluate pathway-based versus statistical integration approaches in the following case studies. The *murine hepatocyte data set* was derived from two separately published data sets that used exactly the same experimental conditions: Omics data were generated from murine hepatocyte cell lines after 2h, 4h, 12h, and 24h of exposure to Benzo[a]pyrene (B[a]P) at two different concentrations (50nM/5μM). Transcriptome data were measured with Affymetrix GeneChips by Michaelson et al. (2011). SILAC-based proteomics and targeted metabolomics data were conducted by Kalkhof et al. (2015). All experiments were generated from different replicates. For statistical integration, we, therefore, relied on fold changes, aggregated for each concentration and time point versus their respective control samples. The transcriptome, proteome, and metabolome layer consisted of 8699, 958, and 163 features, respectively. The *mitochondrial stress response data set* interrogated HeLa cells treated for 24h with Doxycycline (30μ/ml), Actinonin (50μM), Carbonyl cyanide-p-trifluoromethoxyphenyl-hydrazone (FCCP, 10μM), and MitoBloCK-6 (MB, 50μM), respectively (Quirós et al. 2017). Transcriptome (Ion Torrent RNA-seq), proteome (tandem mass tag labeling), and metabolome (iFunnel Q-TOF) data were generated from distinct replicates yielding 15174, 8269, and 1021 features, respectively. For more information about the data generation and processing, please see the Supplementary Material in Section S3.

Pathway analysis

Based on the overview of current pathway-based approaches in Supplementary Text and Figures Section S3.1, we selected two tools for further evaluation. `IMPaLA` employs several well-known pathway databases, such as KEGG, Reactome, or WikiPathways combining more than 5000 pathway definitions. However, it is restricted to either transcriptome or proteome analysis versus the metabolome layer (Kamburov et al. 2011) meaning that `IMPaLA` always integrates two omics layer. Additionally, it is restricted to human pathways only. `PaintOmics` allows to include several different omics layers into its pathway enrichment analysis, provides feature mapping to KEGG IDs, but solely relies on KEGG pathways (Hernández-de Diego et al. 2018). `PaintOmics` supports the interpretation of time series experiments and uses fold change interpolations (i.e., genes, proteins, or metabolites over time) to cluster significantly altered pathways in an overall network. It is also possible to install and run `PaintOmics` locally.

Mitochondrial response data. `PaintOmics` mapped 12546 (transcriptome), 8060 (proteome), and 918 (metabolome) features to KEGG IDs, corresponding to a 17%, 3%, and 10% of unmapped features, respectively. Compared to single-omics approaches, the multi-omics approach yielded the highest amount of significantly enriched pathways for each of the four toxicants (Fig. 3). Also, adjusted p values were lower, sometimes by several orders of magnitude, for a combined multi-omics pathway enrichment compared to each single-layer analyses. These pathways include expected response pathways like oxidative phosphorylation, biosynthesis of amino acids, or serine metabolism. For a more in-depth analysis and direct pairwise comparison of p values between each single- and multi-omics approach, please see Supplementary Section S3.1.

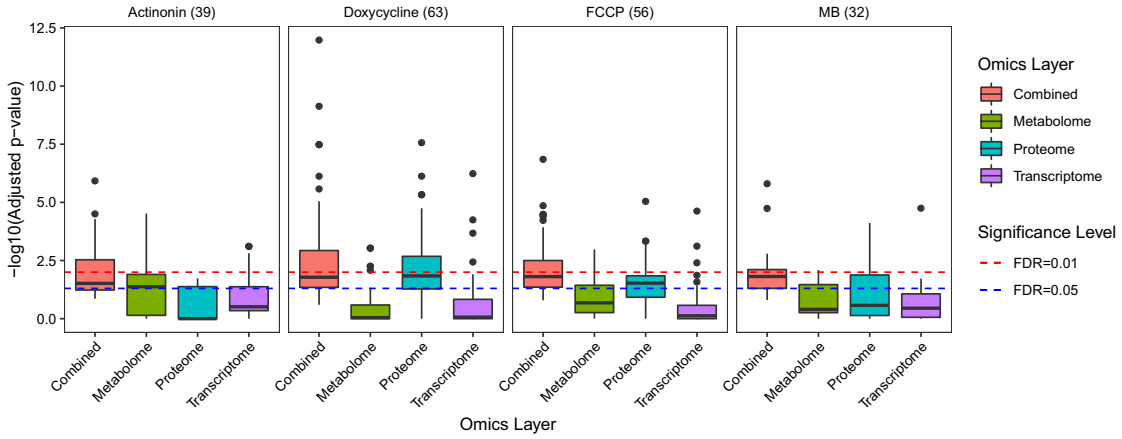


Fig. 3 Comparison of single- and multi-omics pathway enrichment analyses. Adjusted *p* values were derived by `PaintOmics` on HeLa cells treated with Doxycycline, Actinonin, FCCP, and MB. 'Combined' means that all three layers were taken together and a combined adjusted *p* value was calculated by `PaintOmics` using Fisher's method. For each toxicant, all pathways that have been found to be significantly enriched at least once were collected and their adjusted *p* values were compared. The number of pathways found significantly enriched at least once per toxicant is stated in parentheses at the top of each plot

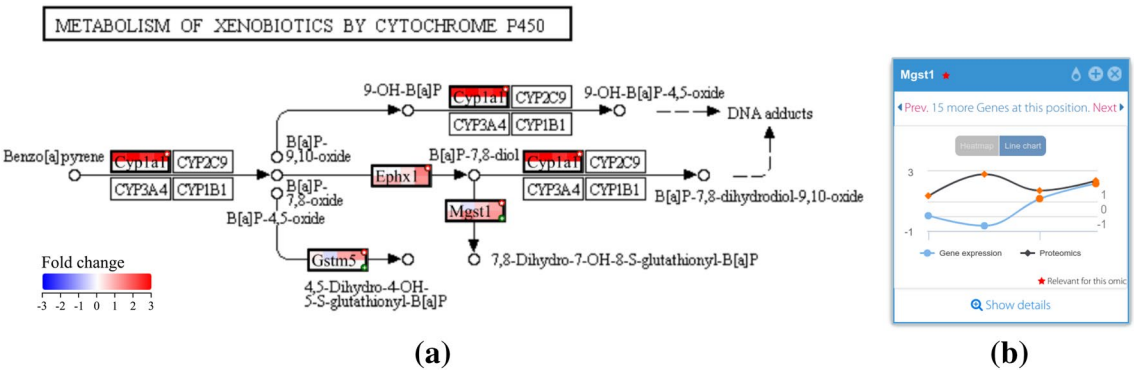


Fig. 4 **a** `PaintOmics` visualization of the enriched B[a]P metabolism pathway that is part of the KEGG xenobiotics metabolism pathway by cytochrome P450. Combined adjusted *p* value: 0.079, transcriptome: 0.0013, proteome: 1, metabolome: NA. **b** Time-resolved fold changes for selected protein and transcript Mgst1 are shown

When applying `IMPaLA` to this data set, we achieved similar results. The number of significantly enriched pathways was always highest when all omics layers were combined. In case of Doxycycline and MB, even more pathways were found to be significantly enriched than by simply adding both two-layer approaches (transcriptome/metabolome + proteome/metabolome). Again, this gain in the amount of detected pathways was accompanied with an increase in confidence, i.e., lower adjusted *p* values (Supplementary Section S3.1, Supplementary Figures S7, S8).

Murine hepatocyte data. For transcriptomics and proteomics, approximately 1% of features could not be mapped to KEGG-identifiers using `PaintOmics`. For metabolomics, the 163 metabolites of the Biocrates AbsoluteIDQ™p150 kit resulted in 26 unique KEGG compounds, including 15 amino acids and mainly carnitines. This restricted number of KEGG-matched metabolites hampered the ability of `PaintOmics` to identify enriched pathways supported by all three omics layers. `PaintOmics` interpolates all given samples of one time series into one 'fold change curve', independent of other covariates, e.g., different concentrations. Hence, we had to compute on both concentrations separately, which weakened the power of the analysis.

For high-dose B[a]P exposure, `PaintOmics` found 6 significantly out of 240 total enriched pathways at the transcriptome level, while only two significant pathways were found for the multi-omics approach [FDR < 0.05, Supplementary Figure S4(a)]. This discrepancy to the findings of the mitochondrial data can be explained by the low coverage of the proteome and metabolome layer of this comparably old data set.

PaintOmics has decent visualization capabilities. Figure 4 displays the B[a]P metabolism pathway, which is part of the KEGG pathway on xenobiotics metabolism by cytochrome P450. Within this subnetwork of 20 total nodes, 17 and 8 nodes were covered by transcriptomic and proteomic data, respectively. Transcripts showed a clear upregulation starting between 4 and 12 h and continuing up to 24 h of exposure, which was not reflected at the proteome level.

The two main known response pathways of B[a]P exposure, the AHR and NRF2 pathways, were not detected by PaintOmics. The simple reason is that these pathways are not represented in the KEGG database. This clearly illustrates the need for a pathway-based analysis to utilize a large variety of high-quality pathway databases to increase the coverage of molecular response pathways or to include custom-defined pathways.

As IMPaLA is limited to human pathways, we mapped the murine transcript and protein IDs accordingly. When IMPaLA was applied to transcriptome and metabolome data, 3841 pathways were found to be enriched, 17 significantly (q value < 0.05). Alternatively, we manually combined transcriptome and proteome data prior to the IMPaLA-integration with the metabolome layer and yielded 3956 enriched and 16 significant pathways. We did not observe a consistent trend between both approaches. In some cases, the q values changed drastically. The q value for the NRF2 pathway, for example, dropped by nearly an order of magnitude in the latter approach compared to the two-layer one. See Supplemental Text and Figures, Section S3.1, for details on the IMPaLA-based analysis.

In summary, neither PaintOmics nor IMPaLA completely satisfy the needs for a pathway-based integration of toxicological multi-omics data. While IMPaLA provides a better integration of pathway databases, PaintOmics exclusively utilizes time-series data and supports more than two omics layers. Both lack support for species-specific or custom-defined pathways and the option to export charts in easily editable formats, like svg or pdf. Our results illustrate that the combination of comprehensive information on multiple molecular levels increases the explanatory power (mitochondrial response data). In contrast, combining omics layers with strongly varying degree of coverage in the different layers, as observed for the older murine hepatocyte data set, is not only not beneficial but also detrimental compared to single-omics analysis of the layer with the highest information content.

Multi-omics factor analysis - MOFA

As a member of the group of simultaneous, statistical integration tools we evaluated MOFA (multi-omics factor analysis), which is a framework for unsupervised integration of multi-omics data sets to discover the principal sources of variation (Argelaguet et al. 2018). The method implements a generalization of principle components analysis aiming at inferring a low-dimensional representation of the data using latent factors. A latent factor is analogous to a principle component but combines data over different omics layers, i.e. separate data matrices. Due to the non-paired nature of the datasets used, we had to aggregate these over treatment groups. The resulting low number of conditions may limit the results of the MOFA-analysis.

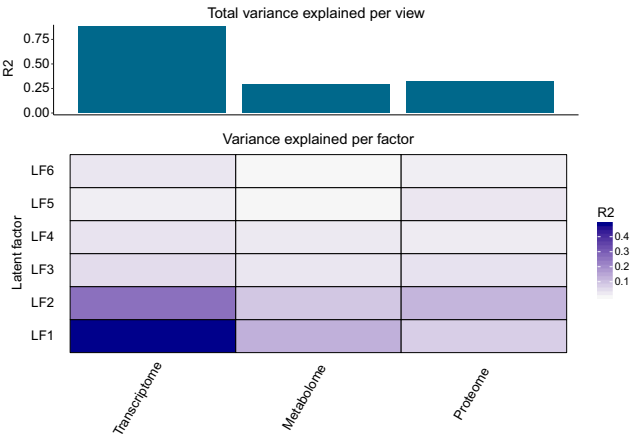


Fig. 5 Variance explained by the MOFA model trained on the B[a]P data set. On top, the total amount of explained variance within the different omics layers is shown. The vast majority of transcriptomic variance can be explained by this model, while merely up to 30% can be explained in the proteome and metabolome layer, respectively. The chart at the bottom displays the explained variance for each latent factor and omics layer. As usual with dimensionality reduction techniques, the first components (here: latent factors) capture more variance. MOFA was able to encapsulate variance of all three omics layer into one latent factor (see LF 1 and 2) highlighting co-regulation of certain transcripts, proteins, and metabolites

Figure 5 illustrates the explained variance per latent factor (LF) across all omics layers after training the model. LF1 and LF2 explained most variance across all three omics layers, reflecting co-variation between features of these layers, while the importance of the remaining factors decreased rapidly. A detailed description on model selection and most relevant transcripts, proteins, and metabolites in each LF is provided in Supplemental Text and Figures, Section S3.2.

For each LF, we conducted gene set enrichment analyses (GSEA) against MSigDB[7] (Subramanian et al. 2005; Liberzon et al. 2015). In brief, LF1 includes, gene sets for regulating gene expression and cellular responses, whereas LF2 covers gene sets for xenobiotic metabolism and apoptosis. In LF6, significantly enriched gene sets were mainly mitochondrial- or cancer-associated gene sets. No such clear

[7] http://software.broadinstitute.org/gsea/msigdb.

conclusion could be drawn from the remaining latent factors (see Supplementary Table S2).

MOFA is a powerful and easy-to-use tool for integrating multi-omics data sets of arbitrary layers. It is able to encapsulate elements of heterogeneity across omics layers and helps to infer their biological function. However, the latter investigation would strongly benefit from the option to jointly analyze multiple omics layers. Furthermore, MOFA is not able to explicitly interpret time series data; instead each time point is interpreted as an independent condition.

Joint gene set enrichment analysis

As an alternative to the omics-layer-specific gene set enrichment in MOFA we evaluated the results of a GSEA in a union of transcriptome and proteome data versus the individual omics layers relying on the 12 h and 24 h time points of exposure to the high B[a]P concentration. Overall, we observed lower adjusted *p* values when combining transcriptome and proteome information compared to the single-omics approaches (Supplementary Figure S12). Venn diagrams of the overlap between the enriched pathways indicate that a combination of omics layers mainly lead to a gain of additionally enriched gene sets, while only few were found only in single-omics experiments at 24 h, which was less pronounced at 12 h (Fig. 6). Apart from the increase in detected gene sets, the adjusted *p* values of sets found to be significantly enriched with a single transcriptomic *and* a combined approach were similar (12 h) or stronger (24 h) when omics layers were combined.

These findings suggest that the advantages of more identified gene sets with stronger *p* values, outweigh the disadvantages, of a small number of gene sets not found in the multi-omics approach and that an increase of noise that might have been provoked by the combination of multiple omics layer seems a minor issue.

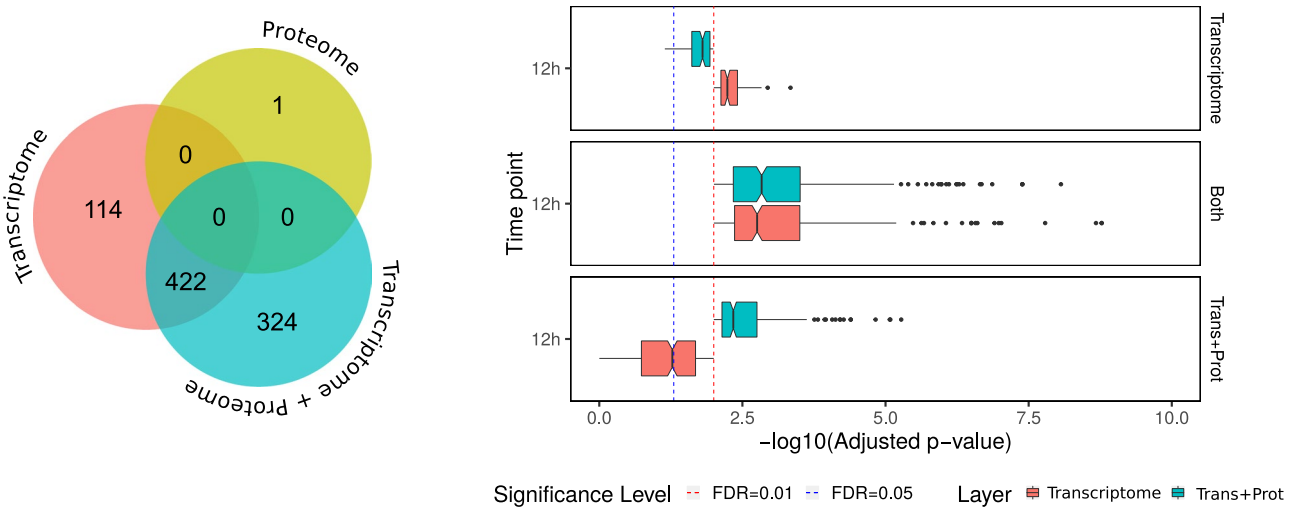
In summary, several approaches seem applicable for multi-omics data integration and pathway detection in a toxicological setting. However, all available methods have inherent disadvantages that need to be addressed to use them for risk assessment purposes. Furthermore, only a combination of different tools is currently able to draw a comprehensive picture of the triggered molecular responses in the cell.

Conclusions

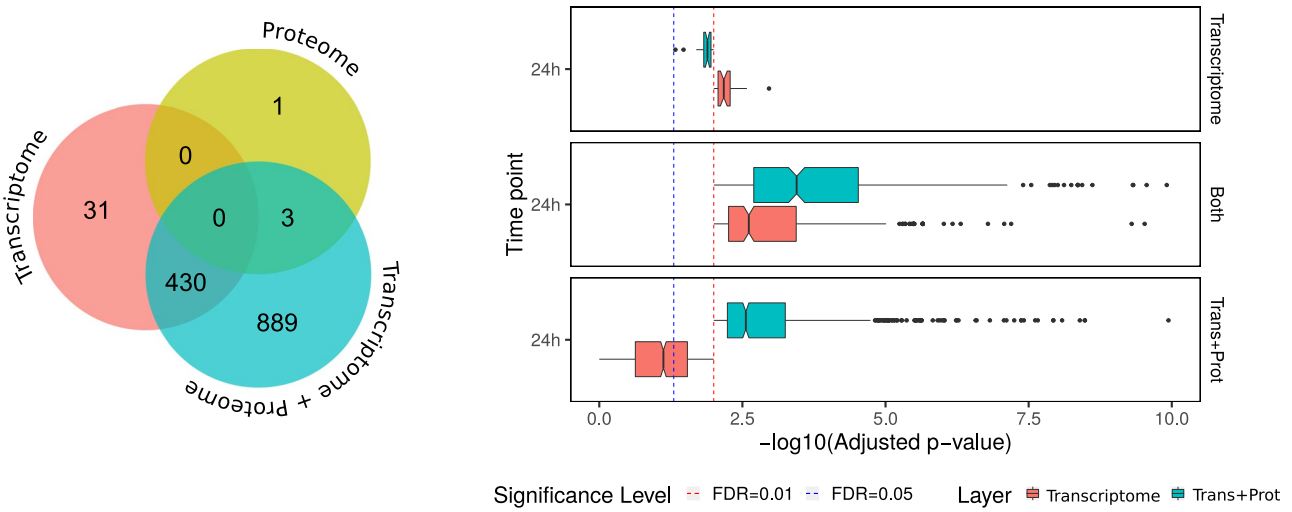
Here, we reviewed the state-of-the-art options for multi-omics data acquisition and data analysis and the prospects and challenges for employing multi-omics data integration in toxicology. In part, we base our conclusions on a retrospective analysis of available multi-omics data for toxicological research questions.

Toxicologic multi-omics data sets beyond two omics layers are rare. Multi-omics data are distributed over several repositories and typically there are no links between corresponding data sets. Therefore, we implemented the publicly available web-tool MOD-Finder to facilitate the search for multi-omics data for a compound of interest Canzler et al. (2019). Using MOD-Finder and manual curation we compiled a table of multi-omics data sets for toxicological research questions (Supplementary Table S1). As this phase of the XomeTox project aimed at deriving best practices and guidelines for the multi-omics oral rat toxicity study in phase II of the project, we particularly aimed at identifying or compiling multi-omics data sets with three or more omics layers. From this analysis, we can conclude that toxicological multi-omics data sets with three and more layers are rare. Also, most data sets we compiled are of limited utility from a current perspective, as many are fairly dated and consequently employ out-dated methods for omics data acquisition. While this is not a serious problem for transcriptomics data, where microarrays may have their disadvantages compared to RNA-seq but still provide a broad coverage of the transcriptome layer, the situation is different for proteomics and metabolomics data. Older proteomics data sets are often based on the DIGE technique and comprise only data for comparably few strongly differentially expressed proteins, which hampers integration with the other omics layers. Metabolomics data frequently consist of targeted analyses of a small number of metabolites, which again generates difficulties for the integration, in particular if many of the detected metabolites cannot be properly mapped to mainstream data bases like KEGG.

None of the triple omics data sets we compiled that uses up-to-date methods follows a split-sample or at least a source matched study design. For those data sets, which we compiled from different studies this is expected. However, also for data sets where all omics layers where generated in one study, we did not obtain a paired-sample study. For older experiments, this is not surprising: in a split-sample design, the available sample material is often limiting and one of the features that changed dramatically during the technical development over the last couple of years is the sensitivity and thus material consumption of omics techniques. Non-paired samples, however, have an important consequence for data analysis: As there is no inherent order in the vector of experiments, e.g., correlations cannot be computed over the entire set of samples but only over an aggregate over experimental conditions. The consequence for methods like correlation network inference or many dimensionality reduction techniques often is that due to the limited number of experimental conditions, the aggregated data vectors are too short for a meaningful application.



(a) Significantly enriched gene sets with high B[a]P at 12h



(b) Significantly enriched gene sets with high B[a]P at 24h

Fig. 6 Comparison of the number of significantly enriched gene sets (FDR < 0.01) and the distribution of adjusted *p* values at 12 h (**a**) and 24 h (**b**) post-exposure to B[a]P. The Venn diagrams on the left side display the total number of significantly enriched gene sets that were either detected with a single transcriptome, single proteome, or a combined transcriptome/proteome approach. Overlapping sets indicate an enrichment with more than one method. The boxplots on the right side compare the distribution of adjusted *p* values of gene sets that were found to be significantly enriched. The top layer displays adjusted *p* values of gene sets found exclusively significantly enriched with the single transcriptome, the bottom layer of those exclusively found with the combined transcriptome/proteome method, and the layer in between of gene sets found with both approaches

A flexible pathway enrichment tool for triple omics data and beyond is missing. There is a broad range of multi-omics data integration and analysis tools available (see Section "Multi-omics data analysis"). However, many of these methods are exclusively available for dual omics data. Also, many approaches focus solely on dimensionality reduction and the selection of features for classification tasks. While the classification of tissue based on multi-omics data, e.g., into different tumor subtypes is an important challenge in oncology, classification is of less importance for toxicological applications. In our opinion of particular relevance for toxicology are (i) detecting a pathway response to exposure, to associate the molecular response with certain key events of an AOP and (ii) the refinement of AOPs using multi-omics data. A pathway response is typically detected using pathway or more formally term enrichment among a set of responding biomolecules. Only few such tools are available for multi-omics data. We have evaluated several pathway

enrichment tools in the case study section. While each of the evaluated tools had its strengths and weaknesses, none fully addressed a set of basic requirements, which we consider fundamental for multi-omics pathway analysis in toxicology: (i) The capability to deal with more than two omics layers, e.g., the four layers that we have selected for the second phase of XomeTox in Section "Multi-omics in toxicology" or ideally with an arbitrary number of omics layers. (ii) Interacting with a flexible set of pathway databases and permitting the definition of custom pathways. (iii) Providing at least rudimentary support (e.g., like `PaintOmics`) for time series data. (iv) Allowing to use concentration *and* time resolved data. (v) Providing capabilities to visualize enriched pathways and provide this visualization in a vector graphics format.

Several tools are available for network inference or pathway extension using multi-omics data. However, to the best of our knowledge, all available tools are focused on a specific combination of omics layers, many inspired by cancer research involving copy number variation (e.g., Zarayeneh et al. (2017), Yuan et al. (2018)) and none of them is geared towards the four omics layers we have selected.

Multi-omics data integration improves pathway detection. The main hypothesis XomeTox is based on is that multi-omics data improve the detection of a pathway response due to a more complete representation of the pathway in the data compared to single-omics data. However, integration of different omics layers may also add up noise, either due to limitations of the experimental design or the inherent variance in time scales within and between the omics layers (see Section "Multi-omics in toxicology"). To objectify whether multi-omics data improves detection of a pathway response, we proposed to select omics data for a well understood AOP or mode of action and compare the significance of pathway detection, e.g., via term enrichment for the multi-omics case versus the single-omics layers. A recently published data set on mitochondrial stress response served as an optimal setting for multi-omics data integration, as it provided high coverage across all omics layers. We observed a remarkable increase in confidence in the detection of triggered pathways, clearly demonstrating the power of multi-omics data integration. A data set on B[a]P exposure of hepatocytes demonstrated the challenges in multi-omics data analysis with older data sets that exhibit varying levels of coverage across the omics layers. While we detected enrichments for NRF2 and AHR pathways or B[a]P metabolism in all integration approaches these were often not significant and confidence improved only with some approaches of multi-omics integration compared to single-omics analysis. We would also expect a more pronounced effect for multi-omics data generated according to our proposed best practices in particular following a split-sample design. In summary, we argue that multi-omics data integration can significantly improve the detection of a pathway response if the quality and coverage of the individual omics experiments is high.

Best practices and pitfalls for multi-omics in toxicology. First, and probably most importantly, multi-omics needs to be considered in study design from the beginning. Adding another omics layer to existing omics data later on is strongly limiting the power of the approach. The experimental design should follow a split sample approach. If there is a biological reason to compare different tissues or sample materials, e.g., transcriptome in tissue and serum metabolome, one should adhere to a source matched design.

Sampling should be performed at identical time points for all omics layers to enable a split sample or source matched design. To account for the different time scales within and between omics layers, either a dense time resolution covering the different time scales from the beginning of exposure and paralleled by an appropriate modeling strategy could be used, which might however be difficult to realize in an animal experiment and also conflicts with RRR goals. Alternatively, we propose to adhere to a repeated dosing scheme, and start acquiring omics layers after a time span that guarantees to reach a steady state of the pathway response. To discriminate adaptive from adverse responses we expect that a recovery phase provides important insights.

Omics layers should be selected to maximize the coverage for the pathways in focus, if this information is available prior to the study. From a current perspective, we recommend to include both transcriptomics and proteomics, although some of the generated information is redundant. However, the case studies demonstrated that they mutually improved pathway detection. Both should be acquired in a non-targeted fashion. We expect phosphoproteomics to provide important contributions to pathway interrogation. Due to the lack of phosphoproteomics layers in any data sets we have compiled, we cannot currently judge on the merits of phosphoproteomics. The contribution of metabolomics was hard to assess in data we have used. Metabolomics should be acquired using a broad targeted approach that provides metabolites, which also can be mapped to pathways in standard databases, or ideally additionally including a non-targeted strategy. Epigenomics may provide important information for studies aiming at trans-generational or long-term effects. However, detecting a pathway response from epigenomic data is still limited.

Acquired data should be deposited in the available repositories. Meta data should follow the guidelines proposed by consortia like MAQC (Shi et al. 2017). To enable linking of the omics data, links to the other layers should be included in the meta data. Based on the experience made with `MOD-Finder`, we call for encoding a unique identifier for the substance used and the concentration and time point in the meta data. Frequently, this information can only be retrieved from an accompanying publication.

When assembling multi-omics data from independent experiments, e.g., using the `MOD-Finder` researches should be aware of the non-paired nature of the samples, which may require aggregation of the samples over experimental conditions or employing a meta analysis subsequent to an analysis of the individual omics layer as proposed in Tarazona et al. (2018). Also, even if time points and concentrations match, noise can still be introduced by diverging cell lines or strains even if identically named and differences in the substance formulation or the substance itself, regarding purity or contained byproducts, when obtained from different suppliers.

Prospects and challenges for a regulatory use of multi-omics data. In addition to the best practices discussed above, regulatory application demands more considerations, particularly regarding the reproducibility of data acquisition and data analysis.

Methods for omics data acquisition develop at a high pace. As we learned during the case studies, using up-to-date technology would have provided considerable advantages over the in part comparably old data we used. Consequently, any omics technology that would be used in a regulatory setting would be quickly outdated. How long would it be useful to stick to an outdated technology for consistency reasons, despite up-to-date approaches would provide advantages like better coverage and thus also more insight?

While animal experiments employing apical endpoints could be independently replicated probably after decades, as long as the used animal strain would be available, this is not the case for omics data. Commercial platforms are—sometimes abruptly—discontinued and replicating a microarray experiment with a platform used 15 years ago is simply impossible. Omics data for regulatory application could thus build on "open source kits" and materials, which could be custom manufactured later. However, for instrumentation this is not feasible. Regulatory application would thus need to build on optimal reproducibility of the analysis, which was called the second best alternative to full replication (Peng 2011).

We have summarized the main requirements of Reproducible Computational Research in Section "Multi-omics data analysis". One important constituent of reproducibility is reporting all versions, parameters, and data connections and linking analysis code and results. We provide a software for the analysis of omics data that supports reproducibility (Kämpf et al. 2019). Additionally, full reproducibility also requires to preserve the computational environment that was used for the analysis (Grüning et al. 2018). The authors of `Nextflow` provide a solution that strongly builds on using containers (Di Tommaso et al. 2017). The latter is important to prevent an effect termed numerical instability, the observation that exactly the same analysis code can lead to diverging results on different computational architectures, due the accumulation of tiny numerical errors.

Finally, regulatory application requires stringent quality control of the individual omics layers and the integrative analysis. For the individual omics layers several tools are available, typically depending on the technology or platform used. For the integrative analysis, developing quality criteria is an urgent task. As a start we propose to investigate the pathway associations of the individual versus the integrative analysis to monitor erroneous results due to accumulation of noise.

In summary, we propose to further evaluate the potential of multi-omics in toxicological research. We have presented promising results in the cases studies. However, we urgently need to generate a multi-layer omics experiment based on a multi-omics focused design using up-to-date omics techniques. This will allow to more thoroughly evaluate the promises of multi-omics, to employ different integration methods, and to further evaluate the contribution of individual omics layers.

Acknowledgements Open Access funding provided by Projekt DEAL. The authors are grateful to the members of the Research Liasion Team of the XomeTox project for helpful discussions.

Compliance with ethical standards

Funding This work was supported by CEFIC LRI through funding the project C5 - XomeTox. Martin von Bergen is grateful for funding by the DFG funded Collaborative Research Centre "Gut-Liver Axis" 1382.

Conflict of interest The authors declare that they have no potential conflict of interest.

References

Acharjee A, Ament Z, West JA, Stanley E, Griffin JL (2016) Integration of metabolomics, lipidomics and clinical data using a machine learning method. BMC Bioinform 17:440. https://doi.org/10.1186/s12859-016-1292-2

Akavia UD, Litvin O, Kim J, Sanchez-Garcia F, Kotliar D, Causton HC, Pochanard P, Mozes E, Garraway LA, Pe'er D (2010) An integrated approach to uncover drivers of cancer. Cell 143(6):1005–17. https://doi.org/10.1016/j.cell.2010.11.013

Argelaguet R, Velten B, Arnol D, Dietrich S, Zenz T, Marioni JC, Buettner F, Huber W, Stegle O (2018) Multi-omics factor analysis—a framework for unsupervised integration of multi-omics data sets. Mol Syst Biol 14(6):e8124. https://doi.org/10.15252/msb.20178124

Arrington JV, Hsu CC, Elder SG, Andy Tao W (2017) Recent advances in phosphoproteomics and application to neurological diseases. Analyst 142:4373–4387. https://doi.org/10.1039/c7an00985b

Baker M (2016) 1,500 scientists lift the lid on reproducibility. Nature 533:452–454. https://doi.org/10.1038/533452a

Baryshnikova A (2016) Systematic functional annotation and visualization of biological networks. Cell Syst 2(6):412–421. https://doi.org/10.1016/j.cels.2016.04.014, http://www.sciencedirect.com/science/article/pii/S240547121630148X

Bersanelli M, Mosca E, Remondini D, Giampieri E, Sala C, Castellani G, Milanesi L (2016) Methods for the integration of multi-omics data: mathematical aspects. BMC Bioinform 17(Suppl 2):15. https://doi.org/10.1186/s12859-015-0857-9

Berthold MR, Cebron N, Dill F, Gabriel TR, Kötter T, Meinl T, Ohl P, Sieb C, Thiel K, Wiswedel B (2007) KNIME: the Konstanz Information Miner. In: Studies in classification, data analysis, and knowledge organization (GfKL 2007). Springer, Berlin

Boll K, Reiche K, Kasack K, Mörbt N, Kretzschmar AK, Tomm JM, Verhaegh G, Schalken J, von Bergen M, Horn F, Hackermüller J (2013) MiR-130a, miR-203 and miR-205 jointly repress key oncogenic pathways and are downregulated in prostate carcinoma. Oncogene 32:277–285. https://doi.org/10.1038/onc.2012.55

Branco MR, Ficz G, Reik W (2011) Uncovering the role of 5-hydroxymethylcytosine in the epigenome. Nat Rev Genet 13:7–13. https://doi.org/10.1038/nrg3080

Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ (2013) Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, dna-binding proteins and nucleosome position. Nat Methods 10:1213–1218. https://doi.org/10.1038/nmeth.2688

Buescher JM, Driggers EM (2016) Integration of omics: more than the sum of its parts. Cancer Metab 4:4. https://doi.org/10.1186/s40170-016-0143-y

Buesen R, Chorley BN, da Silva Lima B, Daston G, Deferme L, Ebbels T, Gant TW, Goetz A, Greally J, Gribaldo L, Hackermüller J, Hubesch B, Jennen D, Johnson K, Kanno J, Kauffmann HM, Laffont M, McMullen P, Meehan R, Pemberton M, Perdichizzi S, Piersma AH, Sauer UG, Schmidt K, Seitz H, Sumida K, Tollefsen KE, Tong W, Tralau T, van Ravenzwaay B, Weber RJM, Worth A, Yauk C, Poole A (2017) Applying 'omics technologies in chemicals risk assessment: Report of an ECETOC workshop. Regul Toxicol Pharmacol: RTP 91(Suppl 1):S3–S13. https://doi.org/10.1016/j.yrtph.2017.09.002

Bustin SA (2014) The reproducibility of biomedical research: sleepers awake!. Biomol Detect Quantif 2:35–42. https://doi.org/10.1016/j.bdq.2015.01.002

Canzler S, Hackermüller J, Schor J (2019) MOD-Finder: identify multi-omics data sets related to defined chemical exposure. arXiv e-prints arXiv:1907.06346

Cavill R, Jennen D, Kleinjans J, Briedé JJ (2016) Transcriptomic and metabolomic data integration. Br Bioinform 17(5):891–901. https://doi.org/10.1093/bib/bbv090

Cech TR, Steitz JA (2014) The noncoding rna revolution-trashing old rules to forge new ones. Cell 157(1):77–94. https://doi.org/10.1016/j.cell.2014.03.008

Chari R, Coe BP, Vucic EA, Lockwood WW, Lam WL (2010) An integrative multi-dimensional genetic and epigenetic strategy to identify aberrant genes and pathways in cancer. BMC Syst Biol 4:67. https://doi.org/10.1186/1752-0509-4-67

Dellafiora L, Dall'Asta C (2017) Forthcoming challenges in mycotoxins toxicology research for safer food—a need for multi-omics approach. Toxins https://doi.org/10.3390/toxins9010018

Dhanoa JK, Sethi RS, Verma R, Arora JS, Mukhopadhyay CS (2018) Long non-coding rna: its evolutionary relics and biological implications in mammals: a review. J Anim Sci Technol 60:25. https://doi.org/10.1186/s40781-018-0183-7

Di Tommaso P, Chatzou M, Floden EW, Barja PP, Palumbo E, Notredame C (2017) Nextflow enables reproducible computational workflows. Nat biotechnol 35:316–319. https://doi.org/10.1038/nbt.3820

Duncan KD, Fyrestam J, Lanekoff I (2019) Advances in mass spectrometry based single-cell metabolomics. Analyst 144(3):782–793. https://doi.org/10.1039/c8an01581c

Ebbels TMD, Cavill R (2009) Bioinformatic methods in NMR-based metabolic profiling. Prog Nucl Magn Reson Spectrosc 55(4):361–374. https://doi.org/10.1016/j.pnmrs.2009.07.003

ECETOC (2008) European centre for ecotoxicology and toxicology of chemicals. The application of 'omic technologies in toxicology and ecotoxicology: case studies and risk assessment. In: Workshop report no. 11. ECETOC, Belgium

ECETOC (2010) European centre for ccotoxicology and toxicology of chemicals. 'Omics in (eco)toxicology: case studies and risk assessment. In: Workshop report no. 19. ECETOC, Belgium

ECETOC (2013) European centre for ecotoxicology and toxicology of chemicals. 'Omics and risk assessment science. In: Workshop report no. 25. ECETOC, Belgium

Escher BI, Hackermüller J, Polte T, Scholz S, Aigner A, Altenburger R, Böhme A, Bopp SK, Brack W, Busch W, Chadeau-Hyam M, Covaci A, Eisenträger A, Galligan JJ, Garcia-Reyero N, Hartung T, Hein M, Herberth G, Jahnke A, Kleinjans J, Klüver N, Krauss M, Lamoree M, Lehmann I, Luckenbach T, Miller GW, Müller A, Phillips DH, Reemtsma T, Rolle-Kampczyk U, Schüürmann G, Schwikowski B, Tan Y, Trump S, Walter-Rohde S, Wambaugh JF (2017) From the exposome to mechanistic understanding of chemical-induced adverse effects. Environ Int 99:97–106. https://doi.org/10.1016/j.envint.2016.11.029

Falkenburger BH, Jensen JB, Dickson EJ, Suh BC, Hille B (2010) Phosphoinositides: lipid regulators of membrane proteins. J Physiol 588(Pt 17):3179–85. https://doi.org/10.1113/jphysiol.2010.192153

García-Ortega LF, Martínez O (2015) How many genes are expressed in a transcriptome? Estimation and results for RNA-Seq. PLoS One 10(6):e0130262. https://doi.org/10.1371/journal.pone.0130262

Glaab E, Baudot A, Krasnogor N, Schneider R, Valencia A (2012) Enrichnet: network-based gene set enrichment analysis. Bioinformatics 28(18):i451

Goecks J, Nekrutenko A, Taylor J (2010) Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Genome Biol 11:R86. https://doi.org/10.1186/gb-2010-11-8-r86

González I, Cao KA, Davis MJ, Déjean S (2012) Visualising associations between paired 'omics' data sets. BioData Min 5(1):19. https://doi.org/10.1186/1756-0381-5-19

Grüning B, Chilton J, Köster J, Dale R, Soranzo N, van den Beek M, Goecks J, Backofen R, Nekrutenko A, Taylor J (2018) Practical computational reproducibility in the life sciences. Cell Syst 6:631–635. https://doi.org/10.1016/j.cels.2018.03.014

Hackermüller J, Reiche K, Otto C, Hösler N, Blumert C, Brocke-Heidrich K, Böhlig L, Nitsche A, Kasack K, Ahnert P, Krupp W, Engeland K, Stadler PF, Horn F (2014) Cell cycle, oncogenic and tumor suppressor pathways regulate numerous long and macro non-protein-coding rnas. Genome Biol 15:R48. https://doi.org/10.1186/gb-2014-15-3-r48

Hendrickx DM, Aerts HJWL, Caiment F, Clark D, Ebbels TMD, Evelo CT, Gmuender H, Hebels DGAJ, Herwig R, Hescheler J, Jennen DGJ, Jetten MJA, Kanterakis S, Keun HC, Matser V, Overington JP, Pilicheva E, Sarkans U, Segura-Lepe MP, Sotiriadou I, Wittenberger T, Wittwehr C, Zanzi A, Kleinjans JCS (2015) diXa: a data infrastructure for chemical safety assessment. Bioinformatics (Oxford, England) 31:1505–1507. https://doi.org/10.1093/bioinformatics/btu827

Hernández-de Diego R, Tarazona S, Martínez-Mira C, Balzano-Nogueira L, Furió-Tarí P, Pappas GJ Jr, Conesa A (2018) PaintOmics 3: a web resource for the pathway analysis and visualization of multi-omics data. Nucleic Acids Res 46(W1):W503–W509. https://doi.org/10.1093/nar/gky466

Hu Y, An Q, Sheu K, Trejo B, Fan S, Guo Y (2018) Single cell multi-omics technology: methodology and application. Front Cell Dev Biol 6:28. https://doi.org/10.3389/fcell.2018.00028

Huang S, Chaudhary K, Garmire LX (2017) More is better: recent progress in multi-omics data integration methods. Front Genet 8:84. https://doi.org/10.3389/fgene.2017.00084

Jahreis S, Trump S, Bauer M, Bauer T, Thürmann L, Feltens R, Wang Q, Gu L, Grützmann K, Röder S, Averbeck M, Weichenhan D, Plass C, Sack U, Borte M, Dubourg V, Schüürmann G, Simon JC, von Bergen M, Hackermüller J, Eils R, Lehmann I, Polte T (2018) Maternal phthalate exposure promotes allergic airway inflammation over 2 generations through epigenetic modifications. J Allergy Clin Immunol 141:741–753. https://doi.org/10.1016/j.jaci.2017.03.017

Kalkhof S, Dautel F, Loguercio S, Baumann S, Trump S, Jungnickel H, Otto W, Rudzok S, Potratz S, Luch A, Lehmann I, Beyer A, von Bergen M (2015) Pathway and time-resolved benzo[a]pyrene toxicity on hepa1c1c7 cells at toxic and subtoxic exposure. J Proteome Res 14(1):164–82. https://doi.org/10.1021/pr500957t

Kamburov A, Cavill R, Ebbels TM, Herwig R, Keun HC (2011) Integrated pathway-level analysis of transcriptomics and metabolomics data with IMPaLA. Bioinformatics 27(20):2917–8. https://doi.org/10.1093/bioinformatics/btr499

Kämpf C, Specht M, Scholz A, Puppel S, Doose G, Reiche K, Schor J, Hackermüller J (2019) uap: Reproducible and robust HTS data analysis. bioRxiv https://doi.org/10.1101/690438

Lê Cao KA, González I, Déjean S (2009) integrOmics: an R package to unravel relationships between two omics datasets. Bioinformatics 25(21):2855–6. https://doi.org/10.1093/bioinformatics/btp515

Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P (2015) The molecular signatures database (MSigDB) hallmark gene set collection. Cell Syst 1(6):417–425. https://doi.org/10.1016/j.cels.2015.12.004

Lowe R, Shirley N, Bleackley M, Dolan S, Shafee T (2017) Transcriptomics technologies. PLoS Comput Biol 13(5):e1005457. https://doi.org/10.1371/journal.pcbi.1005457

Marx-Stoelting P, Braeuning A, Buhrke T, Lampen A, Niemann L, Oelgeschlaeger M, Rieke S, Schmidt F, Heise T, Pfeil R, Solecki R (2015) Application of omics data in regulatory toxicology: report of an international bfr expert workshop. Arch Toxicol 89(11):2177–84. https://doi.org/10.1007/s00204-015-1602-x

McAlister GC, Huttlin EL, Haas W, Ting L, Jedrychowski MP, Rogers JC, Kuhn K, Pike I, Grothe RA, Blethrow JD, Gygi SP (2012) Increasing the multiplexing capacity of TMTs using reporter ion isotopologues with isobaric masses. Anal Chem 84:7469–7478. https://doi.org/10.1021/ac301572t

Meng C, Kuster B, Culhane AC, Gholami AM (2014) A multivariate approach to the integration of multi-omics datasets. BMC Bioinform 15:162. https://doi.org/10.1186/1471-2105-15-162

Meng C, Helm D, Frejno M, Kuster B (2016) moCluster: identifying joint patterns across multiple omics data sets. J Proteome Res 15(3):755–65. https://doi.org/10.1021/acs.jproteome.5b00824

Mias GI, Yusufaly T, Roushangar R, Brooks LR, Singh VV, Christou C (2016) MathIOmica: an integrative platform for dynamic omics. Sci Rep 6:37237. https://doi.org/10.1038/srep37237

Michaelson JJ, Trump S, Rudzok S, Gräbsch C, Madureira DJ, Dautel F, Mai J, Attinger S, Schirmer K, von Bergen M, Lehmann I, Beyer A (2011) Transcriptional signatures of regulatory and toxic responses to benzo-[a]-pyrene exposure. BMC Genomics 12:502. https://doi.org/10.1186/1471-2164-12-502

Nagahori H, Nakamura K, Sumida K, Ito S, Ohtsuki S (2017) Combining genomics to edentify the pathways of post-transcriptional nongenotoxic signaling and energy homeostasis in livers of rats treated with the pregnane X receptor agonist, pregnenolone carbonitrile. J Proteome Res 16(10):3634–3645. https://doi.org/10.1021/acs.jproteome.7b00364

Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, Mann M (2002) Stable isotope labeling by amino acids in cell culture, silac, as a simple and accurate approach to expression proteomics. Mol Cell Proteom: MCP 1:376–386

Otto C, Reiche K, Hackermüller J (2012) Detection of differentially expressed segments in tiling array data. Bioinformatics (Oxford, England) 28:1471–1479. https://doi.org/10.1093/bioinformatics/bts142

Peng RD (2011) Reproducible research in computational science. Science (New York, NY) 334:1226–1227. https://doi.org/10.1126/science.1213847

Picotti P, Aebersold R (2012) Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. Nat Methods 9:555–566. https://doi.org/10.1038/nmeth.2015

Prot JM, Leclerc E (2012) The current status of alternatives to animal testing and predictive toxicology methods using liver microfluidic biochips. Ann Biomed Eng 40:1228–1243. https://doi.org/10.1007/s10439-011-0480-5

Quirós PM, Prado MA, Zamboni N, D'Amico D, Williams RW, Finley D, Gygi SP, Auwerx J (2017) Multi-omics analysis identifies ATF4 as a key regulator of the mitochondrial stress response in mammals. J Cell Biol 216(7):2027–2045. https://doi.org/10.1083/jcb.201702058

Robertson G, Hirst M, Bainbridge M, Bilenky M, Zhao Y, Zeng T, Euskirchen G, Bernier B, Varhol R, Delaney A, Thiessen N, Griffith OL, He A, Marra M, Snyder M, Jones S (2007) Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. Nat Methods 4:651–657. https://doi.org/10.1038/nmeth1068

Sandve GK, Nekrutenko A, Taylor J, Hovig E (2013) Ten simple rules for reproducible computational research. PLoS Comput Biol 9:e1003285. https://doi.org/10.1371/journal.pcbi.1003285

Sauer UG, Deferme L, Gribaldo L, Hackermüller J, Tralau T, van Ravenzwaay B, Yauk C, Poole A, Tong W, Gant TW (2017) The challenge of the application of 'omics technologies in chemicals risk assessment: background and outlook. Regulatory toxicology and pharmacology: RTP 91(Suppl 1):S14–S26. https://doi.org/10.1016/j.yrtph.2017.09.020

Scala G, Kinaret P, Marwah V, Sund J, Fortino V, Greco D (2018) Multi-omics analysis of ten carbon nanomaterials effects highlights cell type specific patterns of molecular regulation and adaptation. NanoImpact 11:99–108

Schmidt JR, Vogel S, Moeller S, Kalkhof S, Schubert K, von Bergen M, Hempel U (2018) Sulfated hyaluronic acid and dexamethasone possess a synergistic potential in the differentiation of osteoblasts from human bone marrow stromal cells. J Cell Biochem. https://doi.org/10.1002/jcb.28158

Shi L, Kusko R, Wolfinger RD, Haibe-Kains B, Fischer M, Sansone SA, Mason CE, Furlanello C, Jones WD, Ning B, Tong W (2017) The international MAQC society launches to enhance reproducibility of high-throughput technologies. Nat Biotechnol 35:1127–1128. https://doi.org/10.1038/nbt.4029

von Stechow L, Francavilla C, Olsen JV (2015) Recent findings and technological advances in phosphoproteomics for cells and tissues. Expert Rev Proteom 12:469–487. https://doi.org/10.1586/14789450.2015.1078730

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci USA 102(43):15545–50. https://doi.org/10.1073/pnas.0506580102

Tarazona S, Balzano-Nogueira L, Conesa A (2018) Chapter eighteen—multiomics data integration in time series experiments. In: Jaumot J, Bedia C, Tauler R (eds) Data analysis for omic sciences: methods and applications, comprehensive analytical chemistry, vol 82. Elsevier, Amsterdam, pp 505–532. https://doi.org/10.1016/bs.coac.2018.06.005

Tarca A, Draghici S, Khatri P, Hassan S, Mittal P, Kim J, Kim C, Kusanovic J, Romero R (2009) A novel signaling pathway impact analysis. Bioinformatics 25(1):75–82

Tralau T, Luch A (2015) Moving from rats to cellular omics in regulatory toxicology: great challenge toward sustainability or "up-shit-creek without a paddle"? Arch Toxicol 89(6):819–21. https://doi.org/10.1007/s00204-015-1511-z

Tralau T, Oelgeschläger M, Gürtler R, Heinemeyer G, Herzler M, Höfer T, Itter H, Kuhl T, Lange N, Lorenz N, Müller-Graf C, Pabel U, Pirow R, Ritz V, Schafft H, Schneider H, Schulz T, Schumacher D, Zellmer S, Fleur-Böl G, Greiner M, Lahrssen-Wiederholt M, Lampen A, Luch A, Schönfelder G, Solecki R, Wittkowski R, Hensel A (2015) Regulatory toxicology in the twenty-first century: challenges, perspectives and possible solutions. Arch Toxicol 89(6):823–50. https://doi.org/10.1007/s00204-015-1510-0

Unlü M, Morgan ME, Minden JS (1997) Difference gel electrophoresis: a single gel method for detecting changes in protein extracts. Electrophoresis 18:2071–2077. https://doi.org/10.1002/elps.1150181133

van Breda SGJ, Claessen SMH, van Herwijnen M, Theunissen DHJ, Jennen DGJ, de Kok TMCM, Kleinjans JCS (2018) Integrative omics data analyses of repeated dose toxicity of valproic acid in vitro reveal new mechanisms of steatosis induction. Toxicology 393:160–170. https://doi.org/10.1016/j.tox.2017.11.013

Vaske C, Benz S, Sanborn J, Earl D, Szeto C, Zhu J, Haussler D, Stuart J (2010) Inference of patient-specific pathway activities from multi-dimensional cancer genomics data using PARADIGM. Bioinformatics 26(12):i237–45

Vorreiter F, Richter S, Peter M, Baumann S, von Bergen M, Tomm JM (2016) Comparison and optimization of methods for the simultaneous extraction of DNA, RNA, proteins, and metabolites. Anal Biochem 508:25–33. https://doi.org/10.1016/j.ab.2016.05.011

Wang R, Dillon CP, Shi LZ, Milasta S, Carter R, Finkelstein D, McCormick LL, Fitzgerald P, Chi H, Munger J, Green DR (2011) The transcription factor Myc controls metabolic reprogramming upon t lymphocyte activation. Immunity 35(6):871–82. https://doi.org/10.1016/j.immuni.2011.09.021

Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 10:57–63. https://doi.org/10.1038/nrg2484

Wishart DS (2007) Human metabolome database: completing the 'human parts list'. Pharmacogenomics 8:683–686. https://doi.org/10.2217/14622416.8.7.683

Xiao Z, Cheng G, Jiao Y, Pan C, Li R, Jia D, Zhu J, Wu C, Zheng M, Jia J (2018) Holo-seq: single-cell sequencing of holo-transcriptome. Genome Biol 19:163. https://doi.org/10.1186/s13059-018-1553-7

Yang L, George J, Wang J (2019) Deep profiling of cellular heterogeneity by emerging single-cell proteomic technologies. Proteomics. https://doi.org/10.1002/pmic.201900226

Yuan L, Guo LH, Yuan CA, Zhang YH, Han K, Nandi A, Honig B, Huang DS (2018) Integration of multi-omics data for gene regulatory network inference and application to breast cancer. IEEE/ACM Transact Comput Biol Bioinform. https://doi.org/10.1109/TCBB.2018.2866836

Zarayeneh N, Ko E, Oh JH, Suh S, Liu C, Gao J, Kim D, Kang M (2017) Integration of multi-omics data for integrative gene regulatory network inference. Int J Data Min Bioinform 18:223–239. https://doi.org/10.1504/IJDMB.2017.10008266

Zhang X, Chen X, Weirauch MT, Zhang X, Burleson JD, Brandt EB, Ji H (2018) Diesel exhaust and house dust mite allergen lead to common changes in the airway methylome and hydroxymethylome. Environ Epigenet 4:dvy020. https://doi.org/10.1093/eep/dvy020

Zhou G, Xia J (2018) OmicsNet: a web-based tool for creation and visual analysis of biological networks in 3D space. Nucleic Acids Res 46(W1):W514–W522. https://doi.org/10.1093/nar/gky510

Ziegenhain C, Vieth B, Parekh S, Reinius B, Guillaumet-Adkins A, Smets M, Leonhardt H, Heyn H, Hellmann I, Enard W (2017) Comparative analysis of single-cell RNA sequencing methods. Mol cell 65:631–643.e4. https://doi.org/10.1016/j.molcel.2017.01.023