P-A-R-C HORIZON-HLTH-2021-ENVLTH-03 CONTRACT N. 101057014

Partnership for the Assessment of Risks from Chemicals

Additional Deliverable AD5.4

Title: Inventory of omics data, biochemical data, other relevant readouts in public domain repositories for all PARC selected adverse outcomes and mapping of the available information sources and datasets related to human pathophysiology focusing on the prioritised endpoints

WP5 – T5.3

Technical reference

Work package	WP5 – Hazard Assessment
Task	Task 5.3 - Quantitative systems toxicology and development of new AOPs
Dissemination level 1	PU
Lead Beneficiary/ Responsible AE	UL-LACDR
Contributing Participants	Bob van de Water/UL-LACDR/water_b@lacdr.leidenuniv.nl
Co-authors	Giulia Callegaro/UL-LACDR/g.callegaro@lacdr.leidenuniv.nl Imke Bruns/UL-LACDR/i.b.bruns@lacdr.leidenuniv.nl Birgit Mertens/SCIENSANO/Birgit.Mertens@sciensano.be Philipp Antczak/UKK/pantczak@uni-koeln.de Olivier Taboureau/INSERM/olivier.taboureau@paris7.jussieu.fr Jörg Hackermüller/UFZ/joerg.hackermueller@ufz.de
Reviewers external to WP	Suzanne Fitzpatrick/US FDA/ Suzanne.Fitzpatrick@fda.hhs.gov Stefan Scholz/UFZ/ stefan.scholz@ufz.de
Reviewers internal to WP	Philip Marx-Stoelting/BfR/Philip.Marx-Stoelting@bfr.bund.de Gilles Rivière/ANSES/Gilles.RIVIERE@anses.fr
Due date of deliverable	30 April 2023
Actual submission date	29 November 2023

1 PU = Public

PP = Restricted to other programme participants (including the Commission Services)

RE = Restricted to a group specified by the consortium (including the Commission Services)

CO = Confidential, only for members of the consortium (including the Commission Services)

Document history

Version	Date	Reviewer name/Institutions	Short description of changes
V0	27 June 2023	Birgit Mertens/SCIENSANO & Philipp Antczak/UKK	Included RAx CS
V1	30 June 2023	Giulia Callegaro/UL-LACDR & Imke Bruns/UL-LACDR	Included ab initio CS and pathology CS
V2	4 July 2023	Olivier Taboureau/INSERM & Jörg Hackermüller/UFZ	Included temporal CS and first draft of introduction
V3	1 August 2023	Bob van de Water/UL-LACDR	Final version
V4	22 November 2023	All authors	Responses to reviewer's comments

Abstract

Systems toxicology is a research area that integrates molecular information from diverse sources for both mechanistic understanding and prediction of toxicological outcome that should be human relevant and reflect pathophysiology responses. Within PARC, the objective is to support the chemical risk assessment in the AOP framework, with the evaluation of mechanistic data for a quantitative hazard characterization through new approach methods (NAMs*). In this context, a workshop on data requirements for systems toxicology was organized with the aim to establish 4 case studies of relevance on this topic i.e.: i) application of omics for read across and grouping; ii) application of omics for hazard characterization for ab initio risk assessment; and iii) uncertainty of temporal omics responses for hazard characterization. The objective is to determine best practices and protocols in computational approaches that will be beneficial in the identification and integration of relevant modes of action covering a diversity of MIEs and KEs in AOPs

To reach this objective, first, an inventory of omics data available in public domain repositories for all PARC partners and of interest for the 3 case studies was discussed during the workshop and selected omics data are presented in this report.

Keywords

Systems toxicology, NAM, read across, ab initio, temporal analysis

“Funded by the European Union. Views and opinions expressed are, however, those of the author(s) only and do not necessarily reflect those of the European Union or the Health and Digital Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.”

* We here consider NAMs according to ECHA’s definition: "NAMs include in silico approaches, in chemico and in vitro assays, as well as the inclusion of information from the exposure of chemicals in the context of hazard assessment. They also include a variety of new testing tools, such as “high- throughput screening” and “high-content methods” e.g. genomics, proteomics, metabolomics; as well as some “conventional” methods that aim to improve understanding of toxic effects, either through improving toxicokinetic or toxicodynamic knowledge for substances.”

Table of contents

Authors and Acknowledgments

Bob van de Water, Giulia Callegaro and Imke Bruns (Leiden University)

Birgit Mertens (Sciensano)

Philipp Antczak (UKK)

Olivier Taboureau (INSERM)

Joerg Hackermuller (UFZ)

Carsten Weiss (KIT)

Background

Systems toxicology is a research area that integrates mechanistic information from diverse sources for both mechanistic understanding and prediction of toxicological outcome. Systems toxicology takes advantage of the enormous amount of toxicology data from both *in vitro* and *in vivo* data. Such information has been used to establish prediction tools for mode-of-action using e.g. QSAR modelling. In addition, systems toxicology has integrated large amounts of omics data to establish gene networks that are co-regulated and associated with pathological outcomes. Moreover, integration of functional studies in systems toxicology has allowed underpinning of mechanistic relationships between critical events (i.e. MIE and KE) that determine the adverse outcome.

So far, systems toxicology has been feasible for adverse outcomes with a high data density, such as for liver and kidney toxicity. However, for the PARC areas *i.e.* immunotoxicology, neurotoxicology, non-genotoxic carcinogenesis, EDCs, metabolic disruption, these datasets are sparser making it less feasible to establish systems toxicology tools that can contribute to chemical risk assessment. In this project, we will focus on the data gathering and data gap filling for these PARC areas and establish interactive systems toxicology tools to facilitate mechanism-based chemical risk assessment, with a focus on omics data. While OECD guidelines have been proposed for the description of omics experiments and bioinformatics analysis (Harrill JA, et al., 2021), no guidelines are available for interpretation of omics data. Omics interpretation is largely based on the bioinformatics tools and approaches established in the different labs and the qualitative and quantitative interpretation of omics data is likely to be different based on different analysis steps. A systematic comparison of the validity of different interpretation methods is essential to provide confidence in the application of omics data in hazard characterization and full risk assessment. To reach this objective, we will focus on 4 case studies that will have direct impact on omics integration in safety assessment, particularly we will focus on: i) application of omics for read across and grouping approaches in safety assessment; ii) application of omics for hazard characterization for ab initio risk assessment; iii) uncertainty of temporal differences of omics responses with relevance for hazard characterization; and iv) how does in vitro transcriptomics data translate to human biology in vivo. Comparison of methodologies, protocols, computational approaches will be essential to establish best practices and solutions for the interpretation of omics data. We will particularly focus on transcriptomics. It should allow to harmonize the outcomes from these analyses. Given that WP5 Task 5.3 of PARC is also focusing on AOPs, we will complement activities in Task 5.3. by conducting appropriate experiments on transcriptome responses in the context of these case studies to establish if and how we can relate omics data to mode of action and can map it qualitatively and quantitatively (in)directly to AOP MIEs and KEs. In this

report we introduce the objective of the four case studies and describe the inventory of omics data available in public domain repositories that will be used in the three different case studies.

Results

Case study 1: Application of transcriptomics for read across and grouping approaches.

Introduction to the case study:
'Grouping of substances and read-across' is one of the most important approaches for filling data gaps, particularly under the REACH regulation. Within this approach, the properties of 'target' substances are predicted based on the properties of analogous ('source') substances, assuming that similar structural characteristics lead to similar hazards. This circumvents the need to test every target substance resulting in a significant reduction in animal testing. Under REACH, structural/physicochemical similarity is a pre-requisite for any grouping and read-across approach including a hypothesis that justifies why the toxicokinetic and toxicodynamic properties of the grouped compounds are expected to be similar. However, many read-across cases fail to demonstrate toxicokinetic and toxicodynamic similarities. Data collected by NAMs can contribute to the substantiation of read-across hypotheses by complementing the read-across based on structural/physicochemical properties with bioactivity profiling, helping to define molecular boundaries and characterize the toxicological (dis)similarities between source and target substances, optimizing grouping of chemicals and, thus, improving regulatory action. NAMs allow to characterize highly specific molecular responses which indicate a more in-depth impact of the substance on the organism. In the case of omics technologies this is even done on an unbiased scale allowing for a highly granular approach to grouping. This approach, unlike current methodologies, does not fully have to rely on a direct link to toxicity but can also be applied to simultaneously evaluate the effect of different toxicological endpoints. To make full use of such techniques in regulation, systems toxicology approaches are required to be developed and systematically evaluated.

Under REACH, grouping and read-across can be used in different contexts: (i) read-across/grouping as adaptation (Annex XI) to fill information requirements under REACH; (ii) read-across/grouping as part of risk management measures; and (iii) grouping as part of ECHA family grouping to assess the regulatory needs. The first scenario is the most challenging because it requires an exploration of the added value of using molecular data, as there is a clear legal framework requiring high-quality data to fulfill information requirements (i.e. validated NAMs). The two other scenarios are more flexible and, therefore, more suited as a first step to explore how molecular data can support the grouping and read-across approach under REACH.

Based on the outcome of discussions with ECHA, it was decided to focus in the first case study on bisphenol (BPA) alternatives, which fits in the overall objectives of PARC. Many bisphenols are widely used and are known endocrine disruptors affecting human health and environmental organisms. Moreover, they also have reprotoxic properties (ECHA, 2021). In 2022, ECHA and the member states decided that a group restriction would be the best way to manage the risks of bisphenols. Germany is currently preparing a proposal to restrict BPA and other bisphenols (ECHA, 2022). One approach that has been proposed is to have a restriction for the bisphenols that already have a classification. However, for many 'unclassified' compounds only limited data are available which could result in regrettable substitution, i.e. replacement with other hazardous compounds. It is also important to ensure that, in case BPA is substituted by alternatives that are not part of the group of bisphenols, these compounds are also safe. Within this context, we will

investigate how both structural and molecular data can be used to evaluate the toxicological (dis)similarities between BPA alternatives with respect to biological responses and potency.

Overall objective of the case study:
The overall objective of the case study is to develop a workflow that is easily applicable for users in order to support the grouping of chemicals based on multiple relevant sets of features including molecular responses (e.g. omics), chemical structural information, and other chemical properties. Different existing molecular datasets on BPA alternatives will be collected. Afterwards, the following methods will be applied:

- Selection of features on the molecular and structural level improving accuracy in read-across (regularised regression, genetic algorithms, Random Forests, etc);
- Feature set enrichment analysis (i.e. on the gene level) to identify similar compounds suitable for read-across (i.e. GSEA or rank based statistics);
- Clustering approaches partitioning compounds using above mentioned selected features or response statistics (hierarchical, multidimensional, or fuzzy clustering).

The workflow will be developed for the grouping of BPA alternatives. Afterwards, its application for other groups of chemicals will be investigated.

Requirements for the datasets:

Publicly available or proprietary datasets owned by PARC partners.

Table with the datasets:

Partner/institute that generated the data	Assay Name	Organism	Target Tissue	Toxicity Domain	Cell Format	Cell Short Name	ZXCV	Exposure Time	Compounds Tested	Public Domain Source
AUTH	Metabolomics	Human	Pancreas	Metabolic diseases	3D	EndoC-βH1	10 different compounds in 3 different concentrations	48,72h (based on the compound)	BP, BPA, BPF, BPS, CdCl2, DBP, DDE, DEHP, PFOA, PFOS	Not yet published (dataset generated in the framework of OBERON)
University of Debrecen	RNA-seq	Human	Ovarian tissue	Metabolic diseases	2D	PEO1	1 concentration	8 h	BPA, zearalenone	Published (10.3390/toxins15020140)
Kings College London	RNA-seq	Human	Mammary gland epithelium	Metabolic diseases	2D	MCF-7, MCF-10A, MCF-12A	Several concentrations	6 days	BPA, BPAF, BPAP, BPB, BPF, BPS, BPZ, pendimethalin, pendimethalin formulation	Published (GEO: GSE182963; https://doi.org/ 10.18332/pht/155263)
Health Canada	TempO-Seq	Human	Mammary gland epithelium	Metabolic diseases	2D	MCF-7	10 concentrations	48h	BPA; 4,4'-BPF; BPAP; BPS; BADGE; DDS; BPAF; DMSO; E2; Dex; BPS-MAE; BTUM; D-8; Pergafast201; 2,4'-BPS; TGSA; BPC; 2,4'-BPF; BPS-MPE	Published (GSE211183; 10.1093/toxsci/kfac127)
Health Canada	RNA-seq	Human	Embryonic stem cells	Embryonic development	2D	WA09	5 concentrations	72h (3x24h)	BPA, BPS, BPF, TBBPA	Published (GSE153320 ; https://doi.org/ 10.1016/j.tiv.2021.105097

Within T5.1 and T5.2, additional datasets for BPA alternatives will be generated that could also be included in the case study in a later stage.

Rationale for the selected dataset as a starting point for the CS:
First, the inventory of molecular datasets available among the partners of the 5.3.1.a_Y1_SystemsToxicology_UL-LACDR project was consulted to search for BPA-related datasets. Although a few datasets containing information on one or more BPA alternatives were found, these were considered too limited for the purpose of our case study. Consequently, a wider literature search for BPA datasets was performed. Several datasets were identified, among which the dataset of Matteo et al. (2023) that became very recently publicly available, was considered of high interest for the case study. This TempO-Seq Human Whole Transcriptome dataset collected in MCF-7 cells with 10 concentrations of BPA and 15 data-poor BPA alternatives will be used as a starting point for the case study.

The details of the dataset are provided below.

Short Description of the datasets:

Dataset Matteo et al. (2023)

- Cell type: MCF-7 cells (human-derived breast cancer cell line)
- Compounds tested: BPA; 4,4'-BPF; BPAP; BPS; BADGE; DDS; BPAF; DMSO; E2; Dex; BPS-MAE; BTUM; D-8; Pergafast201; 2,4'-BPS; TGSA; BPC; 2,4'-BPF; BPS-MPE
- Concentrations: 10 concentrations in the range 0.0005–100 µM
- Exposure time: 48h
- Methodology: TempO-Seq Human Whole Transcriptome v2.0
- Data processing and QC (R-ODAF pipeline):
  - Processing of reads into FASTQ files using bcl2fastq v2.20.0.422.
  - Processing of FASTQ files with TempO-SeqR analysis pipeline in R (version 3.1) supplied by BioSpyder
  - Study-wide QC on the count matrix using several methods to measure consistency and remove low-quality samples, using the methods in Harrill et al. (2021) as a guideline
  - Gene filtering according to Omics Data Analysis Frameworks for Regulatory application (R-ODAF) guidelines

Contact was taken with the Canadian research groups that generated this dataset and the possibilities for collaboration are being further explored. Since BPA alternatives have been selected as priority compounds for WP5 in PARC, additional datasets that might be relevant for the read-across case study will be generated under T5.1 and T5.2.

Expectations for the case study:
The outcome of this case study will be several fold: 1) we will provide a defined strategy how to evaluate omics data for read across and grouping approaches by applying and comparing different bioinformatics approaches; 2) we will get insight in biological similarity and potency of BPA alternatives, priority compounds of the PARC project; 3) we will have direct regulatory impact as our insights can provide guidance on which BPA alternatives need to be covered by the restriction and which not. ECHA has been involved in the design of the case study and will be regularly consulted during the execution of the case study to ensure the regulatory relevance of the outcome.

Case study 2: Application of transcriptomics for *ab initio* hazard characterization and risk assessment.

Introduction to the case study:

When read-across cannot be easily applied, e.g., when no suitable groups are identified or insufficient information on group compounds is available, the suggested framework to implement a NGRA approach is via the so-called *ab initio* approach, where in vitro and in silico methods are employed to reach hazard identification and assessment (Berggren E, et al., Comput Toxicol. 2017). In this context, in vitro and in silico methods should help to 1) narrow the hypothesis of toxicity and 2) derive quantitative dose-dependent data to reach a Point of Departure (PoD) definition for the identified hazards.

Omics technologies have several advantages that align with the requirements in the *ab initio* approach. First, they inherently provide a wide view of the molecular events underlying compound exposure, allowing to evaluate, at once, many possible mechanisms of action (Harrill, Viant, et al., Toxicological Sciences 2021; Krewski et al., Arch Toxicol 2020). Second, omics approaches traditionally provide a quantitative output that can be easily adapted to reach a determination of the dose at which effects are triggered (Cote et al., Environmental Health Perspectives 2016; Farmahin et al., Archives of Toxicology 2017). Additionally, technological innovations have moved omics forward in terms of cost and feasibility. For example, transcriptomics has progressed from using arrays and full genome RNA sequencing to more cost-efficient high-throughput transcriptomics using targeted RNA sequencing, such as the TempO-Seq technology.

The major challenge in utilizing omics-derived data lies in their interpretation and the relation with risk assessment. There is general consensus on the bioinformatics approaches to normalize and extract differential abundant molecules. Subsequently, systems biology employs an array of diverse and powerful approaches to identify and quantify the pathways and processes that are affected by chemical exposure. Frameworks to report omics experiments and bioinformatics approaches have been proposed by OECD working groups (OECD, 2022). Nevertheless, it is still unclear how the different interpretation approaches can contribute to hazard assessment, how their application should be reported, and what are the consequences of using different methods and approaches (e.g., in terms of hazards and PoDs identified).

Overall objective of the case study:

This case study aims at investigating the application of different systems toxicology methods available at case study partners for an *ab initio* approach for hazard assessment. In particular, we aim to:

- Compare the outcome of different interpretative tools and methods as provided by the partners and applied to the same dataset:
  - Do they lead to the same hazard identification?
  - If so, are comparable PoDs identified?
- Define how the investigated methods can be used to connect the OMICS responses to the AOP framework.
- Provide tools to define the human predictivity of omics applied to *in vitro* methods.

Among the tested interpretation methodologies we will evaluate and systematically compare Weighted Gene Correlation Network Analysis (WGCNA), pathway enrichment analysis, GSEA, IPA and BMD Express.

Requirements of the datasets:

For the selection of the omics data for the case study we proposed a two-step approach, based on the current availability of datasets within the project:

1. Data-rich and knowledge-rich compounds: to evaluate the different approaches against a known scenario.
2. Data-poor compounds, in line with PARC interest areas (EDC, immunosuppressants, DNT, Non-Genotoxic Carcinogens).

Table with datasets:

Partner	Assay Name	Organism	Target Tissue	Toxicity Domain	Cell Format	Cell Short Name	Assay Format Type	Omics type	Content Readout Type	ZXCV	Exposure Time	Endpoint Measurement	Endpoint Readout	Compounds Tested
UL-LACDR	TempO-seq whole transcriptome	Human	Kidney	Nephro-toxicity	2D	RPTEC/ TERT1	96w	Transcrip-tomics	mRNA	9 concentr-ations	4h, 8h, 16h, 24h, 48h, 72h	fluorescence	Propidium iodide	cisplatin, cyclosporine
UL-LACDR	TempO-seq whole transcriptome	Human	Kidney	Nephro-toxicity	2D TransWell	Human primary PTEC	96w	Transcrip-tomics	mRNA	6 concentr-ations	8h, 24h, 72h	NA	NA	cisplatin, cyclosporine
UL-LACDR	TempO-seq whole transcriptome	Rat	Kidney	Nephro-toxicity	2D TransWell	Ratprimary PTEC	96w	Transcrip-tomics	mRNA	6 concentr-ations	8h, 24h, 72h	NA	NA	cisplatin, cyclosporine

Rationale for the selected datasets as a starting point for the CS:

1. Data-rich and knowledge-rich compounds: We suggest starting with a data-rich (whole-genome, multiple concentrations, several test systems) and knowledge-rich (mode of action known) compound in order to be able to robustly benchmark the methods proposed. The datasets identified as complying with the requirements are described below.

2. Data-poor compounds: We have not selected a dataset and we will expect new datasets from other PARC projects in T5.1 and T5.2, or collaborative projects outside PARC (e.g. RISK-HUNT3R or PrecisionTox).

Short description of the datasets:

Cisplatin and Cyclosporine A exposure, RNA-seq whole transcriptome:

- RPTEC/TERT1, 9 concentrations, 7 time points
- human primary PTEC, 6 concentrations, 3 time points
- rat primary PTEC, 6 concentrations, 3 time points

Expectations for the case study:

We expect this case study will: 1) define the impact of different transcriptomics interpretation methods on the qualification and quantification of the hazard after chemical exposure; 2) define the impact of different transcriptomic interpretation on the predictions for hazard assessment within the scope of the identified exemplar datasets; and 3) contribute to the expansion of the reporting omics framework to define a guideline for application of bioinformatic tools and omics data for ab initio testing.

Case study 3: Assessment of temporal transcriptomics responses and the impact on hazard identification in safety assessment.

Introduction to the case study:

It is recognized that cellular responses to toxicants have a highly dynamic nature, and exhibit both temporal complexity and dose-response shifts. Most current gene enrichment or pathway analysis lack the recognition of the inherent correlation within time series data, and may potentially miss important pathways or yield biased and inconsistent results that ignore dynamic patterns and time-sensitivity [Gao C. et al. Env. Sci. technol. 2015]. With the accessibility to “omics” data performed with a set of times and doses, assessment of the influence of exposure duration and dose on chemical toxicity is a challenge to tackle.

In this case study, we will apply different computational approaches and protocols developed by the partners in PARC that are suitable for the analysis of such datasets and to propose best practices for time course experiments in systems toxicology. We expect to respond to some questions such as:

- How do different computational/statistical approaches compare?
- How stable are the results with respect to parametrization?
- Are there time-points which are particularly relevant?
- How to cover early and later KEs in an AOP with omics data?

Requirements for the datasets:

Based on datasets collected and proposed (owned) by the PARC partners, some requirements were suggested:

- Publicly available or proprietary datasets owned by PARC partners.
- At least two timepoints.
- For multi-omics data sets: paired samples, i.e., an inherent order of samples across omics layers.

Overall, 4 datasets were selected which are presented below.

Table with the datasets:

Partner	Assay Name	Organism	Target Tissue	Toxicity Domain	Cell Format	Omics type	Content Readout Type	ZXCV	Exposure Time	Compounds Tested	Public domain source
UFZ	Multi-omics	Rat	Liver, thyroid	Thyroid toxicity	tissue	multi-omics	RNA, protein, metabolites	2 concentrations	2w, 4w, 6w (recovery)	Phenytoin, PTU	Manuscript in preparation, not yet published
UFZ	Multi-omics	Human	Hepatic mircotissue	Hepatotoxicity	microtissue	multi-omics	RNA, miRNA, protein, metabolite, epigenomics	2 concentrations	0, 2, 8, 24, 72, 168, 240, 336h	Fluorouracil, Acetaminophen, Azathioprine, Cyclosporin A, Diclofenac, Isoniazid, Methotrexate, Phenytoin, Rifampicin, Valproic acid	https://doi.org/10.1038/s41597-022-01825-1
UFZ	Multi-omics	Human	Cardiac microtissue	Cardiotoxicity	microtissue	multi-omics	RNA, miRNA, protein, metabolite, epigenomics	2 concentrations	0, 2, 8, 24, 72, 168, 240, 336h	Fluorouracil, Doxorubicin, Epirubicin, Idarubicin, Daunorubicin, Amiodarone, Celecoxib, Docetaxel, Mitoxantrone, Paclitaxel	https://doi.org/10.1038/s41597-022-01825-1
UFZ	Multi-omics	Mouse	Liver	Hepatotoxicity	tissue	multi-omics	RNA, protein, metabolite, phosphosites	1 concentration	0, 60, 120, 180, 240min	Insulin	https://doi.org/10.1016/j.celrep.2021.109569

The rationale for the selected dataset as a starting point for the CS:
The rational of selecting these 4 datasets was to have a minimal number of time points for one or multiples compounds, accessible, and serve as a model for both human and eco-toxicology. For one dataset (Verheijen et al. 2021), the association of several omics data for the same set of compounds makes it quite interesting for a multi omics data analysis comparison.

Short description of individual datasets:

Wijaya L. et al (LACDR Cisplatin exposure (under revision – bioRxiv)

- Kidney tissue from in vivo rat experiment
- Compounds tested: cisplatin
- Concentration: 5 mg/kg i.v.
- Sampling time points: 1, 2, 4 24h and 3, 5, 8, 10, 12, 15, 20 and 28 days

Dataset Schüttler et al. (2019)

- Zebrafish embryo, whole organism
- Compounds tested: Diclofenac, Naproxen, Diuron, Mix of the three
- Concentration: from LC0.5 to LC25, respectively
- Sampling time points: 3,6,12,24,48,72 h (Exposure start at 24 hpf)
- Samples 262
- Methodology: Microarrays, RNA-Seq (transcriptomics)

Dataset Matsuzaki et al. (2021)

- Liver tissue from *in vivo* mouse experiment
- Compounds tested: human insulin, phosphate-buffered saline (control)
- Concentration: 0.7U/kg
- Sampling time points: 0, 60, 120, 180, 240min
- Replicates: 3 replicates per condition
- Samples: 54
- Methodology: RNA-Seq (TempO-Seq transcriptomics)

Dataset Verheijen et al. (2021)

- *In vitro* human 3D microtissue – hepatic or cardiac
- Compounds tested hepatoxic: Fluorouracil, Acetaminophen, Azathioprine, Cyclosporin A, Diclofenac, Isoniazid, Methotrexate, Phenytoin, Rifampicin, Valproic acid
- Compounds tested cardiotoxic: Fluorouracil, Doxorubicin, Epirubicin, Idarubicin, Daunorubicin, Amiodarone, Celecoxib, Docetaxel, Mitoxantrone, Paclitaxel
- Concentration: 0, therapeutic (based on PBPK modeling), toxic (IC20 after 7 days)
- Sampling time points: 0, 2, 8, 24, 72, 168, 240, 336h (only therapeutic)
- Replicates: 3 replicates per condition
- Samples: 63 per compound
- Methodology: RNA-Seq (transcriptomics), short RNA-Seq (miRNAs) LC-MS (proteomics), LC-MS & NMR (metabolomics), MeDIP-Seq (Epigenomics)
- Paired data: yes

Expectations for the case study:
We will use the selected datasets to compare the results of the approaches of different team members. Ideally, at least two teams should work per dataset. We will primarily compare at the pathway level how different algorithmic approaches, or diverging parameter selections of

similar approaches determines the detection of a molecular response and propose best practices for time course experiments in systems toxicology and consequences of temporal omics data on qualitative and quantitative hazard characterization and risk assessment in relation to either read across or ab initio based safety assessment.

Case study 4: Translation of in vitro omics to human in vivo: omics datasets that inform on human pathophysiology.

Introduction to the case study:
Currently, the safety of drugs and chemical compounds is tested by the use of animal models. However, these models are inefficient and costly, and raise concerns about their translatability to humans. Animal-free methodologies such as human-based *in vitro* test systems provide a next step towards human relevance, but knowledge is incomplete with respect to their reproducibility of complex organs. Therefore, there is a need for methods able to reveal mechanistic detail of preserved responses between animal and human, and between *in vitro* test systems and human to improve next generation risk assessment.
It is generally known that transcriptomes organize themselves in co-expression clusters regulated by transcription factors, eventually incorporated into biological pathways. For this reason, analyzing the transcriptome can lead to fruitful information about the mechanisms triggered by an external stimulus, such as chemical compound exposure. Gene co-expression network analysis is a widely used method which relies on correlation patterns between genes across samples that are used to generate clusters of genes that behave similarly, called modules. For several animal *in vivo* and human-based *in vitro* test systems, co-regulated gene networks have been established. However, to adequately translate animal-based results to humans and to evaluate the *in vitro* test systems suitability to represent humans, establishing co-regulated networks in the human target organs is crucial.

Overall objective of the case study:
For this case study, transcriptomics data from human liver and kidney biopsies of patient cohorts will be analyzed to establish co-regulated gene networks. These gene networks will be associated with pathologies and used as a human reference of chemical-induced toxicities under the assumption that chemical-induced pathogenesis can be captured by disease-based cohorts. Subsequently, the preservation of these established co-regulated gene networks in humans will be tested for animals and human *in vitro* models, providing insights into the overall applicability of models used in preclinical safety assessment.

Requirements for the datasets
Previous findings have indicated that co-expression networks clearly benefit from increasing sample size, and that the performance of small sized datasets ($n < 100$) can be variable due to the sample composition. Additionally, including a high variety of conditions in the data is needed to establish robust gene co-expression networks covering as many biological processes as possible. For this case study, it is therefore essential to incorporate biopsy samples with a wide variety of diagnoses.

Table with the relevant human liver datasets.

Target organ	Disease	#Individuals	Omics type	Study ID
Liver	Transplantation	10	Affymetrix-HGU-133 Plus 2.0	E-GEOD-14951
Liver	Diabetes Mellitus Type II	17	Affymetrix-HGU-133 Plus 2.0	E-GEOD-23343
Liver	Steatosis	15	Affymetrix-HGU-133 Plus 2.0	E-GEOD-37031
Liver	NAFLD	30	Affymetrix-HGU-133 Plus 2.0	E-GEOD-49541
Liver	Steatosis	18	Affymetrix-HGU-133 Plus 2.0	E-GEOD-63067
Liver	APAP induced injury	6	Affymetrix-HGU-133 Plus 2.0	E-GEOD-74000
Liver	Hepatitis B	128	Affymetrix-HGU-133 Plus 2.0	E-GEOD-83148
Liver	Hepatitis B	34	Affymetrix-HGU-133 Plus 2.0	E-GEOD-96851
Liver	Fibrosis	51	Affymetrix-HGU-133 Plus 2.0	E-GEOD-6764
Liver	Fibrosis	22	Affymetrix-HGU-133 Plus 2.0	E-MTAB-950
Kidney	Transplantation	323	TempO-seq	SP0278
Kidney	Transplantation	355	TempO-seq	SP0295

Rationale for the selected datasets as a starting point for the CS:
The selected datasets contain a total of 331 and 678 samples for liver and kidney, respectively. Moreover, it is noteworthy that both the liver and kidney datasets contain a diverse range of pathologies, suggesting their potential utility in generating co-expression networks covering a diversity of biological processes.

Short description of individual datasets:

Human liver (GEO)
A thorough search through the Gene Expression Omnibus (GEO) database provided 10 different datasets containing human liver biopsy samples with various conditions, namely: Transplantation (E-GEOD-14951), Diabetes Mellitus Type II (E-GEOD-23343), NAFLD (E-GEOD-49541), Steatosis (E-GEOD-63067 & E-GEOD-37031), APAP induced injury (E-GEOD-74000), Hepatitis B (E-GEOD-96851 & E-GEOD-83148), Fibrosis (E-MTAB-950 & E-GEOD-6764). The transcriptomics data was generated using microarray technology.

Human kidney (LUMC/AMC)
A total of 678 human kidney allograft biopsy samples were collected from biobanks in the Dutch academic hospitals, the Leiden University Medical Centre (LUMC) and the Academic Medical Center Amsterdam (AMC). All biopsy samples were taken between 2017 and 2022 and formalin fixed paraffin embedded (FFPE) before storing them in the biobank. To obtain the transcriptomics, a slice of 10 µm was sectioned from the FFPE tissue and sequenced using high throughput targeted RNA sequencing (TempO-seq).

Expectations for the case study:

We expect this case study will reveal mechanistic detail of preserved responses between animal and human and between *in vitro* test systems and human. This approach will facilitate the evaluation of the general applicability of preclinical safety assessment models in terms of their human relevance.

Conclusion

In the systems toxicology project we have now defined 4 case studies and related objectives. Through workshops and follow up discussions we have identified relevant omics datasets that will be fit for purpose for the different case studies. These datasets are either from project partners or available in the public domain. We expect that the systematic analysis and comparison of omics analysis approaches will contribute to assess the readiness level of omics data for integration in risk assessment related to read across and ab initio approaches as well as provide insight into the most relevant timepoints for omics analysis as well as translation to human pathophysiology.

References

ECHA, 2021. Assessment of regulatory needs for bisphenols. https://www.echa.europa.eu/documents/10162/1bd5525c-432c-495d-9dab-d7806bf34312

ECHA, 2022. Group assessment of bisphenols identifies need for restriction. https://echa.europa.eu/nl/-/group-assessment-of-bisphenols-identifies-need-for-restriction

Matteo G, Leingartner K, Rowan-Carroll A, Meier M, Williams A, Beal MA, Gagné M, Farmahin R, Wickramasuriya S, Reardon AJF, Barton-Maclaren T, Christopher Corton J, Yauk CL, Atlas E. In vitro transcriptomic analyses reveal pathway perturbations, estrogenic activities, and potencies of data-poor BPA alternative chemicals. Toxicol Sci. 2023 Feb 17;191(2):266-275. doi: 10.1093/toxsci/kfac127. PMID: 36534918; PMCID: PMC9936204.