



Universiteit
Leiden
The Netherlands

Protectiveness of NAM-based hazard assessment: which testing scope is required?

Zobl, W.; Bitsch, A.; Blum, J.; Boei, J.J.W.A.; Capinha, L.; Carta, G.; ... ; Escher, S.E.

Citation

Zobl, W., Bitsch, A., Blum, J., Boei, J. J. W. A., Capinha, L., Carta, G., ... Escher, S. E.
(2024). Protectiveness of NAM-based hazard assessment: which testing scope is required?
Altex - Alternatives To Animal Experimentation, 41(2), 302-319.
doi:10.14573/altex.2309081

Version: Publisher's Version
License: [Creative Commons CC BY 4.0 license](https://creativecommons.org/licenses/by/4.0/)
Downloaded from: <https://hdl.handle.net/1887/3765765>

Note: To cite this publication please use the final published version (if applicable).

Research Article

Protectiveness of NAM-Based Hazard Assessment – Which Testing Scope is Required?

Walter Zobl¹, Annette Bitsch¹, Jonathan Blum², Jan J. W. A. Boei³, Liliana Capinha⁴, Giada Carta⁴, Jose V. Castell⁵, Enrico Davoli⁶, Christina Drake¹, Ciaran P. Fisher⁷, Muriel M. Heldring⁸, Barira Islam⁷, Paul Jennings⁴, Marcel Leist², Damiano Pellegrino-Coppola⁸, Johannes P. Schimming⁸, Kirsten E. Snijders⁸, Laia Tolosa⁵, Bob van de Water⁸, Barbara M. A. van Vugt-Lussenburg⁹, Paul Walker¹⁰, Matthias M. Wehr¹, Lukas S. Wijaya⁸ and Sylvia E. Escher¹

¹Fraunhofer Institute for Toxicology and Experimental Medicine, Hannover, Germany; ²University of Konstanz, Konstanz, Germany; ³Leiden University Medical Center, Leiden, The Netherlands; ⁴Vrije Universiteit Amsterdam, Amsterdam, The Netherlands; ⁵Instituto de Investigación Sanitaria La Fe, CIBEREHD, CIBERBBN, Valencia, Spain; ⁶Istituto di Ricerche Farmacologiche Mario Negri IRCCS, Milan, Italy; ⁷Certara – Simcyp Division, Sheffield, United Kingdom; ⁸Leiden University, Leiden Academic Centre for Drug Research, Leiden, The Netherlands; ⁹BioDetection Systems, Amsterdam, The Netherlands; ¹⁰Cyprotex Discovery Ltd UK, Macclesfield, United Kingdom

Received September 8, 2023;
Accepted November 24, 2023;
Epub December 4, 2023;
© The Authors, 2023.

Correspondence:
Walter Zobl, MSc,
In Silico Toxicology,
Fraunhofer Institute for Toxicology and
Experimental Medicine,
Nikolai-Fuchs-Str. 1, 30625 Hannover,
Germany
(walter.zobl@item.fraunhofer.de)



ALTEX 41(2), 302-319.
doi:10.14573/altex.2309081

Abstract

Hazard assessment requires toxicity tests to allow deriving protective points of departure (PoDs) for risk assessment irrespective of a compound's mode of action (MoA). The scope of *in vitro* test batteries (ivTB) needed to assess systemic toxicity is still unclear. We explored the protectiveness regarding systemic toxicity of an ivTB with a scope that was guided by previous findings from rodent studies, where examining six main targets, including liver and kidney, was sufficient to predict the guideline scope-based PoD with high probability. The ivTB comprises human *in vitro* models representing liver, kidney, lung, and the neuronal system covering transcriptome, mitochondrial dysfunction, and neuronal outgrowth. Additionally, 32 CALUX®- and 10 HepG2 BAC-GFP reporters cover a broad range of disturbance mechanisms. Eight compounds were chosen for causing adverse effects such as immunotoxicity or anemia *in vivo*, i.e., effects not directly covered by assays in the ivTB. PoDs derived from the ivTB and from oral repeated dose studies in rodents were extrapolated to maximum unbound plasma concentrations for comparison. The ivTB-based PoDs were one to five orders of magnitude lower than *in vivo* PoDs for six of eight compounds, implying that they were protective. The extent of *in vitro* response varied across test compounds. Especially for hematotoxic substances, the ivTB showed either no response or only cytotoxicity. Assays better capturing this type of hazard would be needed to complement the ivTB. This study highlights the potentially broad applicability of ivTBs for deriving protective PoDs of compounds with unknown MoA.

Plain language summary

Animal tests are used to determine how much of a chemical is toxic ("threshold of toxicity") and which organs are affected. In principle, the threshold can also be derived solely from tests with cultured cells. However, only a limited number of cell types can practically be tested, so one challenge is to determine how many and which types shall be tested. In animal tests, only few organs including liver and kidney are regularly among those most sensitively affected. We explored whether a cell-based test battery representing these sensitive organs and covering important mechanisms of toxicity can be used to derive protective human thresholds. To challenge this approach, eight chemicals were tested that primarily cause effects in organs not directly represented in our test battery. Results provided protective thresholds for most of the investigated compounds and gave indications how to further improve the approach towards a full-fledged replacement of animal tests.

1 Introduction

Full replacement of animal testing in toxicology has been a long-term goal for decades for ethical, pragmatic, and scientific reasons. Regulatory agencies around the globe have expressed their vision to decrease the need for animal tests in support of regulatory human health risk assessment (RA) and to facilitate the

incorporation of new approach methodologies (NAMs) (EMA, 2020; EFSA, 2021; US EPA, 2021b; Escher et al., 2022b). This paradigm change will transform risk assessment by shifting the focus to mechanisms of action (US FDA, 2021). In addition to application of NAMs for hazard identification, efforts have been made to enable NAM-based hazard characterization. The principle of using NAM-based points of departure (PoDs) in a

screening context and comparing them with exposure estimates has been demonstrated for example by Paul-Friedmann and colleagues (2020). In their study, they compared NAM-based PoDs, which they had derived for more than 400 chemicals, to the corresponding threshold of toxicological concern (TTC) and *in vivo* PoDs from a range of study types. In a vast majority of cases, they found the NAM-based PoD to lie between the TTC value and the *in vivo* PoD, qualifying their approach as a tool facilitating prioritization for further testing of chemicals. Approaches for screening are in high demand, as next generation risk assessment (NGRA) is generally thought of as a hypothesis-driven, tiered, and iterative process often involving steps of prioritizing chemicals, which can be utilized even before regulatory hazard assessment (HA). In line with the idea of developing fit-for-purpose approaches in a hypothesis-driven way, case studies have been employed to modify and expand the set of NAMs used by Paul Friedman et al. (2020). Instead of limiting themselves to using *in chemico* and *in vitro* assays, where the latter could be phenotypic and receptor-binding assays, Delp et al. (2021) and Baltazar et al. (2020) added high-throughput transcriptomics as a less biased way of sensitively detecting effects for deriving PoDs regarding not-fully-understood neurological effects and systemic toxicity, respectively.

The development of integrated approaches to testing and assessment (IATAs) is particularly challenging for complex toxicological endpoints such as systemic toxicity after repeated administration, as a variety of mechanisms can lead to the same/different toxicological effects and phenotypes. Further, kinetic processes (absorption, distribution, metabolism, and excretion) influence the biologically effective dose of a substance. The uncertainty associated with the lacking representation of organism-level processes, cellular diversity, and interaction remains a challenge. Accordingly, as far as complex or mechanistically not fully understood endpoints are concerned, a number of questions are still under debate around the required scope of testing, read-outs, and interpretation thereof, especially with omics data (Dent et al., 2018).

If we limit a test battery for human health risk assessment to human cell-based assays, an obvious advantage over *in vivo* assays is that inter-species differences become obsolete, whereas the most striking disadvantage is that the biology of only a limited number of organs and MoAs can be represented. However, if we shift our perspective to a slightly more probabilistic angle, we can see that we may not be too bad off with such a limited test battery. Liver and kidney are very often among the most sensitive targets in preclinical *in vivo* studies with oral exposure (Batke et al., 2013). Further, the upper and lower respiratory tracts are of-

ten among the most sensitive targets after repeated inhalation exposure (Escher et al., 2010). For some classes of compounds, other targets are more sensitive compared to lung effects, evidenced *in vivo* as effects observed in less frequent target organs at lower dose levels. However, even if liver and kidney are not affected at study LOEL (lowest observed effect level), there is a very high probability that liver or kidney effects are observed at the next higher dose (Batke et al., 2013). Conversely, if one only investigated liver and kidney effects, next to clinical chemistry and body weight changes, as may be the case in legacy studies with limited investigative scope, one could still derive quantitatively correct PoDs in the vast majority of cases, and the remaining uncertainty caused by the limited study scope could be controlled using an assessment factor.

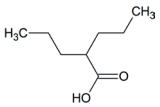
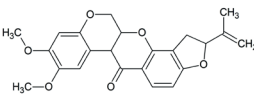
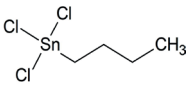
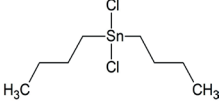
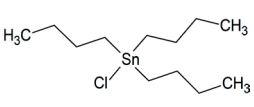
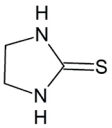
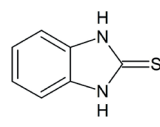
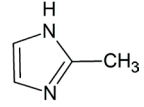
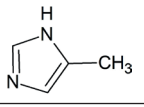
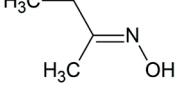
These findings were taken as a starting point for us to study more closely the required scope of testing for NAM-based derivation of PoDs and to challenge the approach with chemicals with rare modes of action (MoA). The present case study investigated the applicability of a defined and therefore limited *in vitro* test battery (ivTB) of predominantly high-throughput methods for deriving PoDs for repeated dose systemic toxicity as an exemplary complex endpoint. The ivTB applies high-throughput methods to derive *in vitro* benchmark concentrations (BMCs). Nominal media concentrations *in vitro* are corrected to the free unbound concentration using *in vitro* biokinetic modelling. For hazard characterization an *in-vitro-to-in-vivo* extrapolation is carried out to estimate the human equivalent plasma concentration as PoD for RA. As suggested earlier by Baltazar et al. (2020), the *in vitro* test battery includes both a set of nontargeted transcriptomics assays and assays targeting a broad spectrum of known molecular initiating events (MIE) and known phenotypic effects. Testing of transcriptome data for organ toxicity considered human hepatocytes, human renal proximal tubule epithelial cells (RPTEC/TERT1), as well as human primary bronchial epithelial cells (PBECS), and neuronal cells (Lund human mesencephalic (LUHMES) cells). The latter two cell types were chosen to address inhalation toxicity and developmental neurotoxicity.

Unlike previous case studies in NAM-based screening approaches such as Baltazar et al. (2020) and Farmahin et al. (2017), we used a small set of test substances. However, these test substances were selected to provide the highest conceivable potential of having their potency underestimated by the current approach, because available *in vivo* evidence suggests that they do not primarily target liver, kidneys, lung, or the neuronal system, i.e., the organs directly represented by suitable *in vitro* models in the ivTB. In other words, we challenged our *in vitro* testing approach with chemicals with rare MoA to explore its degree of

Abbreviations: 2-IT, 2-imidazolidinethione; 2-MBI, 2-mercaptobenzimidazole; 2-MI, 2-methylimidazole; 4-MI, 4-methylimidazole; AOP, adverse outcome pathway; Bioav, bioavailability; BMC, benchmark concentration; BMC_g, gene's BMC; BMC_{CL}, lower bound of the 95% confidence interval of the benchmark concentration; BMC_{pw}, median of a pathway's hits' BMCs; BMC_{g50}, median of BMC_g values; BMC_{pw50}, median of BMC_{pw} values; BMD, benchmark dose; BMR, benchmark response; CL, clearance; C_{nom}, nominal concentration; DBTC, dibutyltin dichloride; DEG, differentially expressed genes; F_a, fraction absorbed; FBS, fetal bovine serum; F_g, first pass gut metabolism; F_h, first pass hepatic metabolism; GFR, glomerular filtration rate; HA, hazard assessment; IATA, integrated approach to testing and assessment; ivTB, *in vitro* test battery; k_a, absorption rate constant; k_e, elimination rate constant; KE, key event; LOEL, lowest observed effect level; LUHMES, Lund human mesencephalic; MBTC, (mono-)butyltin trichloride; MIE, molecular initiating event; MoA, mode of action; NAM, new approach method; NDEG, number of DEGs; NDEG_{cr}, number of concentration-responsively expressed DEGs; NDEG_{high}, number of DEGs at the highest tested concentration; NGRA, next generation risk assessment; N_{pw}, number of pathways enriched by concentration-responsively expressed DEGs; PBEC, primary bronchial epithelial cells; PHH, primary human hepatocytes; PI, propidium iodide; PoD, point of departure; RA, risk assessment; RPTEC, renal proximal tubule epithelial cells; TBTC, tributyltin chloride; t_{max}, time at maximum concentration; TTC, threshold of toxicological concern; VPA, valproic acid.

Tab. 1: Overview of the critical shared *in vivo* effects of the test compounds applied in this case study

Lowest observed effect levels (LOELs) were scaled to subchronic equivalent using a scaling factor of 1.5 for subacute* and chronic** studies (as suggested by Escher et al., 2020). If more than one study was available, the lowest scaled LOEL was used.

Chemical group	Compound name	CAS	Chemical structure	Critical shared effects	LOEL (mmol/kg bw/d)	References
Positive control	Valproic acid (VPA)	99-66-1		Hepatotoxicity (liver steatosis)	0.7	Zhang et al., 2014
	Rotenone	83-79-4		Mitochondrial toxicity (complex-I-inhibitor)	0.007**	Tisdell, 1985
Butyltin	Butyltin trichloride (MBTC)	1118-46-3		Immunotoxicity, endocrine activity	1.9	Parametrix Inc., 2006f, as cited in Anonymous, 2012, p.20.
	Dibutyltin dichloride (DBTC)	683-18-1			0.01	Gaunt et al., 1968, as cited in Anonymous, 2012, p.20.
	Tributyltin chloride (TBTC)	1461-22-9			0.003*	Bressa et al., 1991, as cited in Anonymous, 2012, p.22.
Thiourea	2-Imidazolidinethione (ethylene thiourea; 2-IT)	96-45-7		Thyroid hyperplasia and hypertrophy	0.04*	Anonymous, 1992; National Institute of Health Sciences (Japan), 2004, as cited in HESS DB; Sakuratani et al., 2013
	2-Mercaptobenzimidazole (o-phenylene-thiourea; 2-MBI)	583-39-1			0.005*	Kawasaki et al., 1998; National Institute of Health Sciences (Japan), 1994, as cited in HESS DB; Sakuratani et al., 2013
Imidazole	2-Methylimidazole (2-MI)	693-98-1		Anemia (with secondary extramedullary hematopoiesis ^a)	0.5	Chan, 2004
	4-Methylimidazole (4-MI)	822-36-6			0.5	Chan, 2004
Oxime	Butanone oxime	96-29-7		Formation of methemoglobinemia ^a leading to secondary extramedullary hematopoiesis	0.5	Anonymous, 1999

^a In available studies the effect was found only at higher doses, i.e., not at LOEL; wt, weight.

Tab. 2: Overview of assay types applied in the present study as well as read-outs and cellular mechanisms and processes covered

Main category	Subcategory 1	Subcategory 2	Assay/cell type	Readouts
Transcriptome			HepG2, PHH, RPTEC/TERT1, PBEC and LUHMES (templated oligo-sequencing)	Differentially expressed genes (DEGs), enriched pathways
Cell signaling (= reporter assays)	Stress response signaling	Oxidative stress	HepG2 BAC-GFP	AKR1B10, HMOX1, SRXN1
			CALUX	Nrf2
		DNA damage	HepG2 BAC-GFP	BTG2, P21
			CALUX	p21, p53 (U2OS), p53 (+S9)(U2OS), p53 (HepG2), p53 (+S9)(HepG2)
		Stress signaling (other)	HepG2 BAC-GFP	BIP, CHOP, TRIB3 (endoplasmic reticulum stress), HSPA1B (heat shock response), ICAM1 (inflammatory cytokines)
			CALUX	Hif1a (hypoxic stress), TCF, AP1, ESRE
	Disturbance physiological cell signaling	Endocrine activity	CALUX	ERα, ERα anti, AR, AR anti, PR, PR anti, GR, GR anti, TRβ, TRβ anti
		Cell function modification	CALUX	PPARα, PPARα-anti, PPARγ, PPARγ-anti, PPARδ, PPARδ-anti, AhR (H4IIE), AhR (HepG2), RAR, LXR, PXR
Cell function perturbation		Neuronal endpoints	NeuriTox (LUHMES)	Neurite outgrowth (NO)
		Mitochondrial dysfunction	HepG2	Mitochondrial membrane potential (MMP), formation of reactive oxygen species (ROS), intracellular calcium concentration (Ca ²⁺); all ± CYPs

“MoA-agnosticism”, i.e., whether the approach is protective irrespective of the MoA of a chemical. Two reference compounds targeting the liver and other organs (valproic acid (VPA) and rotenone) were added as positive controls, for which the present ivTB has already proven to provide target organ-specific data (Escher et al., 2022a; van der Stel et al., 2020).

2 Materials and methods

2.1 Selection and characterization of test compounds and biological control compounds

The selection of test compounds was based on effect data from high-quality repeated dose *in vivo* studies with oral administration available from the REPDOSE database (Bitsch et al., 2006). REPDOSE classifies study quality adopting reliability categories similar to those described by Klimisch et al. (1997). Only studies with acceptable scope of investigations and sufficient information regarding study design were considered. Eight test compounds were selected, for which sensitive and clearly adverse

effects were found exclusively in organs other than those directly represented by our assays (liver, kidneys, lung, and nervous system). Effects in liver, kidneys, lung, or nervous system were limited to weight changes and extramedullary hematopoiesis in the liver, the latter being clearly secondary to the formation of anemia. VPA and rotenone acted as positive controls in this respect with mechanistically well-characterized liver and mitochondrial toxicity, respectively. Test compounds include three groups of structurally similar compounds, namely butyltin, thiourea, and imidazole compounds, each in principle qualifying as read-across groups, and butanone oxime.

Table 1 details critical adverse effects and the LOEL observed in preclinical *in vivo* studies. Further effects found in these studies are summarized in Table S1¹ and further preclinical data is described in supplementary file 2².

2.2 Chemicals

Chemicals were purchased from Sigma Aldrich at the highest available purity: valproic acid (certified reference material for analytical application, PHR1061), rotenone (R8875, purity

¹ doi:10.14573/altex.2309081s1

² doi:10.14573/altex.2309081s2



$\geq 95\%$), butyltin trichloride (MBTC, 201057, purity 95%), dibutyltin dichloride (DBTC, 205494, purity 95%), tributyltin chloride (TBTC, T50202, purity 96%), 2-imidazolidinethione (2-IT, I504, purity 98%), 2-mercaptobenzimidazole (2-MBI, M3205, purity 98%), 2-methylimidazole (2-MI, M50850, purity 99%), 4-methylimidazole (4-MI, 199885, purity 98%), and butanone oxime (332828, purity 99%).

2.3 In vitro models

The ivTB applied in the present study includes assays for detecting changes in the transcriptome of cell types representing organs commonly affected by xenobiotics, high-content imaging reporter gene assays detecting cell signaling related to different stresses and disturbance of physiological cell signaling, and functional assays detecting disturbance of mitochondrial functions and neuronal phenotypic changes (Tab. 2).

Cells were regularly tested for mycoplasma contamination. At the time of the experiments, the use of fetal bovine serum (FBS) was indispensable in some assays. All assays were run at sub-cytotoxic test compound concentrations. Cells were treated with test compounds for 24 h and analyses were performed in technical triplicates unless specified otherwise for specific assays. The extracellular and intracellular approaches to representing compound metabolism (Thomas et al., 2019) were carried out in this case study through addition of S9 mixture to the cell medium in CALUX p53 assays and simultaneous transfection of HepG2 cells with recombinant adenoviruses encoding CYP1A2, CYP2C9 and CYP3A4 (Tolosa et al., 2012) in assays addressing mitochondrial function, respectively. These methods were run in parallel with the respective assays without metabolic activation.

2.3.1 Reporter assays and functional assays

CALUX® reporter assays

From the CALUX® (BioDetection Systems bv) battery of *in vitro* reporter gene assays a panel of 32 human cell-based assays was used, each able to measure chemical interactions between a test chemical and a specific nuclear receptor or cell signaling pathway (van der Burg et al., 2013). Exposure to the test compounds, dissolved at 0.06 M in DMSO, was performed for 24 h and at 0.1% or 1% (v/v) according to the assay procedure as described in DB-ALM protocol 197 “Automated CALUX reporter gene assay procedure”. The concentration range used was 20 pM to 630 μ M in 0.5 log unit increments. The analysis consisted of technical triplicates and was performed twice as independent biological replicates. Minimum effective concentration (MEC) values were derived per assay based on the background responses. For nuclear receptor agonist assays, the MEC was defined as the PC10 concentration, which is the concentration where the test compound causes an activation effect equal to 10% of the maximum effect elicited by the test’s reference compound. For nuclear receptor antagonist assays, the MEC was defined as the PC20 concentration, which is the concentration where the test compound causes an antagonist effect equal to 20% of the maximum antagonist effect elicited by the test’s reference compound. For the stress pathway-related assays, which typically

do not show sigmoidal concentration-response curves, the MEC was defined as the FI 1.5 concentration, which is the concentration where the test compound elicits pathway activation 1.5-fold above baseline.

HepG2 BAC-GFP reporter assays

Human hepatoma (HepG2) BAC-GFP (short: GFP) reporter lines have been described previously (Wink et al., 2014, 2017; Callegaro et al., 2023). A set of 10 GFP-reporter lines was selected and maintained in DMEM high glucose (Fisher Scientific – Bleiswijk, The Netherlands) supplemented with 10% (v/v) FBS (Fisher Scientific- Bleiswijk, The Netherlands), 250 U/mL penicillin, and 25 μ g/mL streptomycin (Fisher Scientific – Bleiswijk, The Netherlands) in humidified atmosphere at 37°C and 5% CO₂ (Schimming et al., 2019). All cell lines were used between passage 14 and 20 (until 25 for GFP-ICAM). The cells were seeded in Greiner black μ -clear 384-well plates at 8000 cells per well. Cells were stained overnight 24 h after seeding with 100 ng/mL live Hoechst 33342 in complete DMEM high glucose. At the day of exposure, the medium containing Hoechst 33342 was refreshed with complete DMEM containing 0.2 μ M propidium iodide (PI) (Sigma, P4170). Exposure to the test compounds, dissolved at 0.06 M in DMSO, was performed for 24, 48, and 72 h and at 0.1% or 1% (v/v) according to the assay procedure as described (Wink et al., 2017). The concentration range used was 20 pM to 630 μ M in 0.5 log unit increments. The positive controls were as follows: CDDO-Me (Cayman chemical, 11883) for the oxidative stress reporters SRXN1, HMOX1, and AKR1B10 in a concentration range from 0.001 to 0.1 μ M (dilution factor 2); etoposide (Merck, E1383-25MG) for the DNA damage reporters BTG2 and P21 in a concentration range from 0.33 to 200 μ M (dilution factor 2.5); tunicamycin (Merck, T7765) for the unfolded protein reporters BIP, CHOP, and TRIB3 in a concentration range from 0.15 to 44.4 μ M (dilution factor 2.25); TNF α (R&D System-BioTechne, 210-TA-100) for the inflammatory reporter ICAM1 at 10 ng/mL; CdCl₂ (Merck, 202908-10G) for the heat shock response reporter HSPA1B in a concentration range from 0.1 to 100 μ M. Mitomycin (Selleckchem, S8146) at 150 μ M was included as a positive cell death control. For the inflammatory reporter ICAM1, TNF α at a final concentration of 10 ng/mL was added to all wells 8 h after the compound exposures. Plates were sealed after exposure with gas-permeable seals (VWR international, 731-0622). The experiments were performed as biological triplicates.

Imaging: The plates were imaged at 24, 48 and 72 h after compound exposure. The imaging was performed using a Nikon TiE2000 confocal laser microscope (laser: 647 nm, 540 nm, 488 nm, and 408 nm), equipped with automated stage and perfect focus system. During the imaging, the plates were maintained in humidified atmosphere at 37°C and 5% CO₂. The imaging was done with 20x magnification objective.

Image and data analysis: The quantitative image analysis was performed with CellProfiler version 2.2.0 (Kamentsky et al., 2011). Firstly, the nuclei per image were segmented with a watershed algorithm (Wink et al., 2017). The segmented nuclei were used to identify the cytoplasm (distance of 10 pixels around

the nucleus). For identification of PI-positive cells (i.e., dead cells), the segmented nuclei were laid on top of the segmented PI objects derived in the same manner from the image of the PI channel. The results were stored as HDF5 files. Data analysis, quality control, and graphics were performed using the in-house developed R package h5CellProfiler. For each reporter, the nuclear Hoechst 33342 intensity levels, GFP intensity (in the nucleus and the cytoplasm), and PI area were measured at the single cell level. To quantify the fraction PI positive, the PI images were masked with 2 pixel dilated nuclei based on nuclear segmentation to exclude the background staining noise. The area of PI objects was divided by the area of these nuclei to obtain a PI/nuclei ratio. PI positive was defined as a cell with more than 10% PI area. The GFP intensity from cell population means of each image was calculated based on the single cell results. GFP intensities were min-max scaled to the ranges [-1,1] for ICAM (to account for the up-and-down regulation of the TNF α -modulated ICAM1 and A20 regulation upon exposure of the compounds) and [0,1] for all other reporters. In addition, for each plate, the GFP intensity of DMSO control was calculated to determine the background. A GFP-positive cell was defined as a cell with an intensity > 2 times the mean GFP intensity of the DMSO control.

Benchmark response modeling: BMCs for all readouts were obtained using BMDEExpress version 2 (Sciome). The presence of any trend was assessed using a Williams trend test. A p-value of 0.1 (adjusted for multiple testing using the Benjamini-Hochberg method) was used as cut-off. Next, concentration-response models including hill, linear, poly 2, and exponential 2,3,4 and 5 were fitted. The best model was determined based on the (lowest) Akaike information criterion. In addition, absolute max-fold change ≥ 2 was required, and the BMC was required to be within the tested concentration range. GFP-based BMCs at cytotoxic levels were excluded from further analyses. Per reporter the lowest threshold (i.e., across read-out (GFP fraction positive, GFP intensity) and measurement timepoint) was used for further analyses.

HepG2 mitochondrial dysfunction assays

HepG2 cells (ECACC No.85011430) were cultured in DMEM supplemented with 7% FBS (Hyclone Research Grade FBS, South American Origin. Lot: RAB35926), 50 U penicillin/mL, and 50 μ g streptomycin/mL. For subculturing purposes, cells were detached by treatment with 0.25% trypsin/0.02% EDTA at 37°C. For toxicity studies, cells were seeded in 96-well plates (5000 cells/well, passage < 20) and were allowed to grow and equilibrate for 24 h in medium with lipid-depleted serum. Following treatment, cells were simultaneously loaded with 1.5 μ g/mL Hoechst 33342 and 1.5 μ g/mL PI (Merck). After a 30-min incubation at 37°C with the culture medium containing fluorescent probes, cells were imaged. After incubating with dyes, cells were imaged using the INCELL6000 Analyser (GE Healthcare, USA) as previously described (Tolosa et al., 2012). The cell count was generated from the number of Hoechst 33342-stained nuclei. Cell viability was determined by PI exclusion. Since PI is not permeant to live cells, it is also commonly used to detect dead cells in a population. This allows not only the direct quantification of cytotoxicity, but also the exclusion of dead cells from the analysis, thus restricting fur-

ther functional determinations to the live cell population in each sample. EC10 values were calculated using GraphPad software version 8. The analysis consisted of biological triplicates.

LUHMES neurite outgrowth assay

LUHMES cells were handled as described before (Krug et al., 2013; Scholz et al., 2011). In brief, cells were maintained with 1 μ g/mL fibronectin and 50 μ g/mL poly-L-ornithine (Sigma-Aldrich) pre-coated T75 flasks (Sarstedt, Nümbrecht, Germany). Cells were kept at 37°C and 5% CO₂ and were split every second or third day (at confluency of 80%) in freshly prepared proliferation medium consisting of advanced DMEM/F12 supplemented with 2 mM L-glutamine (both Gibco, Rockville, MD, USA), 1 $\times$ N2 supplement (Invitrogen/Thermo Fisher Scientific, Waltham, MA, USA), and 40 ng/mL recombinant human basic fibroblast growth factor (Bio-Techne, Minneapolis, MN, USA).

To initiate differentiation, cells were passaged to T175 flasks (Sarstedt, Nümbrecht, Germany) and the medium was changed to differentiation medium consisting of advanced DMEM/F12 supplemented with 2 mM L-glutamine (both Gibco, Rockville, MD, USA), N2 supplement (Invitrogen/Thermo Fisher Scientific, Waltham, MA, USA), 1 mM N6,2'-O-dibutyl 3',5'-cyclic adenosine monophosphate (cAMP), 1 μ g/mL tetracycline (both Sigma-Aldrich, Merck, Darmstadt, Germany), and 2 ng/mL recombinant human glial cell-derived neurotrophic factor (GDNF) (Bio-Techne, Minneapolis, MN, USA).

LUHMES neurite outgrowth measurements (UKN4 assay) were performed as described previously (Krug et al., 2013; Stiegler et al., 2011). Briefly, after 2 days of differentiation, cells were plated into 96-well plates (Sarstedt, Nümbrecht, Germany) pre-coated with 1 μ g/mL fibronectin and 50 μ g/mL poly-L-ornithine (Sigma-Aldrich) at a cell density of 100,000 cells/cm² in differentiation medium (without cAMP and GDNF). After 1 h of attachment, cells were treated for 24 h with compounds in 1:3 dilutions spanning 9 concentrations with a highest test concentration of 180 μ M. Exceptions included DBTC and TBTC, for which the highest test concentration was 30 μ M. Plates treated with these 2 compounds were also sealed with sealing tape, and a row with “medium only” was introduced on the plates between the compounds to avoid inter-well transfer. Cells were stained with Hoechst 33342 (1 μ g/mL) and calcein-AM (1 μ M), and image acquisition was performed with an ArrayScan VTI HCS microscope (Cellomics, Waltham, MA, USA). Cell viability and neurite area were assessed in parallel using an automated algorithm as described previously (Krug et al., 2013; Stiegler et al., 2011). The experiments were performed as 2-5 biological replicates, all consisting of 3 technical replicates. A different LUHMES cell passage was used for each biological replicate (passage $\leq$ 20).

2.3.2 High-throughput transcriptomics assays

HepG2 cells, primary human hepatocytes (PHH), RPTEC/TERT1 (Wieser et al., 2008), PBECs, and LUHMES cells were handled as described in the dedicated paragraphs below and processed as required by BioClavis (Glasgow, UK) for sequencing. In brief, cells were washed with PBS after 24 h of test compound



exposure and then the medium was replaced with TempO-Seq Lysis Buffer (BioClavis, Glasgow, UK). Cells were kept in the lysis buffer at room temperature for 15 min. Then, plates were sealed and immediately frozen at -80°C . Targeted transcriptome sequencing using the Templated Oligo-Sequencing (TempO-Seq) technology and data pre-processing including probe alignment was conducted at Bioclavis (Biospyder Tech., Glasgow, UK) (Yeakley et al., 2017). Sequencing covered the EU-ToxRisk 2.2 gene set (Tab. S2¹). The broader Human Whole Transcriptome gene set (Biospyder Technologies, Inc.), which was available only for LUHMES experiments, was subset during data analysis to the EU-ToxRisk 2.2 gene set for comparability. As we aimed to establish concentration-response relationships for early transcriptional response, test compound concentrations ranged from just below cytotoxic levels to orders of magnitude lower (Tab. S3¹) and included up to seven concentration levels.

LUHMES

LUHMES cells were handled as described for the LUHMES neurite outgrowth assay described above. The experiments were performed as 3 biological replicates, with controls consisting of 3 technical replicates each.

RPTEC/TERT1

The human RPTEC/TERT1 cells were routinely cultured in a 1:1 mix of DMEM (Gibco 11966-025) and Ham's F12 (Gibco 21765-029) (final concentration of 5 mM glucose) and supplemented with 2 mM Glutamax, 10 ng/mL epidermal growth factor, 36 ng/mL hydrocortisone, 5 $\mu\text{g/mL}$ insulin, 5 $\mu\text{g/mL}$ transferrin, 5 ng/mL selenium, 100 U/mL penicillin, 100 $\mu\text{g/mL}$ streptomycin, and supplemented with a final concentration of 0.5 % FBS. Cells were cultured in a controlled humidified 37°C , 5% CO_2 environment and differentiated by contact-inhibition after reaching confluency (Aschauer et al., 2013). Passages used were between 75 and 95. The chemical exposures to RPTEC/TERT1 were performed at least 7 days after confluency to guarantee cell differentiation.

PBEC

PBECs were cultured in Keratinocyte Serum Free Medium (KS-FM, Life-technologies 17005-059) supplemented with Pen/Strep (Lonza, DE17-602), epidermal growth factor (EGF, 0.2 ng/mL, Life-technologies 37000-015), bovine pituitary extract (BPE, 25 $\mu\text{g/mL}$, 13028-014) and isoproterenol (1 μM , Sigma I-6504). Prior to seeding of the PBECs, culture surfaces were coated with a mixture of Purecol (30 $\mu\text{g/mL}$, Advanced BioMatrix, 5005-B), bovine serum albumin (10 $\mu\text{g/mL}$, Sigma, A-7030), and fibronectin (5 $\mu\text{g/mL}$, Nalgene, C-43060) in phosphate-buffered saline (Gibco, 10010-015) for 2 h at 37°C . Early passage (p3) PBECs were seeded in pre-coated 96-well plates. Per well, 40,000 cells were seeded in 100 μL culture medium. After 24 h, 100 μL culture medium containing the test compounds at double the final concentration was added to the wells followed by a 24 h incubation period. Next, 100 μL medium was collected for cytotoxicity testing using the LDH cytotoxicity detection kit (Roche 11644793001, according to the manufacturer's protocol).

PHH

Upon thawing, PHHs (LiverPool™ 10-donor mixed gender pooled cryoplatable human hepatocytes, X008001-P; BioreclamationIVT) were diluted in warm thawing/seeding medium (William's E medium, phenol red-free; Sigma, ref number: W1878) supplemented with Thawing Cocktail (Thermo Fisher Scientific, ref no: CM3000) and centrifuged for 5 min at 100 g. Cells were resuspended in fresh medium and plated in 384-well microplates (Corning™ BioCoat™ Collagen Type I-Treated Flat-Bottom Microplate, ref. number: 354667) at 10,000 cells per well, and plates were sealed. 4–6 hours after plating, 25 of 30 μL of the medium was replaced with culturing medium containing William's E Medium supplemented with Cell Maintenance Cocktail (Thermo Fisher Scientific, ref no: CM4000). Plates were sealed and incubated overnight before dosing. Cells were exposed in sealed 384-well plates for 24 h in biological triplicates. Cytotoxicity was determined using a cellular ATP kit (Promega).

HepG2

HepG2 cells (wild type) were purchased from ATCC, Germany (clone HB8065) and maintained the same as the HepG2 BAC-GFP cells. The plates were incubated in a humidified atmosphere at 37°C and 5% CO_2 . All exposures were performed three times independently to cover biological variability.

2.3.3 Transcriptomics data analysis

Probe alignment was performed by BioClavis. Briefly, FASTQ files were aligned using Bowtie, allowing for up to 2 mismatches in the target sequence. This pipeline applies several quality controls with mapped/unmapped reads, replicate clustering, and sample clustering (Yeakley et al., 2017). Count tables were returned by BioClavis as probe counts per sample, with genes being associated to multiple probes. An in-house R pipeline was subsequently used for the following data analysis: 1) probe counts were summed by gene, and genes with no counts in any sample were filtered out (i.e., no information); 2) library size thresholds (Tab. S4¹) were optimized per model system in order to offer a balance between samples discarded and retained, and samples with a lower library size were filtered out (i.e., low information quality); 3) sample-specific normalization size factors for counts per million (CPM) were computed and passed along with count table and matching metadata to DESeq2 (Love et al., 2014) for differential gene expression taking into account treatment, concentration, and timepoint. Plate and solvent differences were also taken into consideration to protect from batch effects.

Along the lines of earlier studies (Farmahin et al., 2017; Webster et al., 2015), genes were filtered based on significance adjusted for multiple testing using the Benjamini-Hochberg method ($p\text{-adj} < 0.05$) and $\log_2$ foldchange ($|\log_2\text{FC}| > 1.5$). Only genes that met both the significance and foldchange criteria in at least one test concentration were considered for BMC modelling.

Transcriptomics BMC modelling and pathway analysis

Figure 1 outlines the overall approach for deriving HA values based on individual genes' concentration-response curves and on concentration-responsive gene expression in enriched pathways.

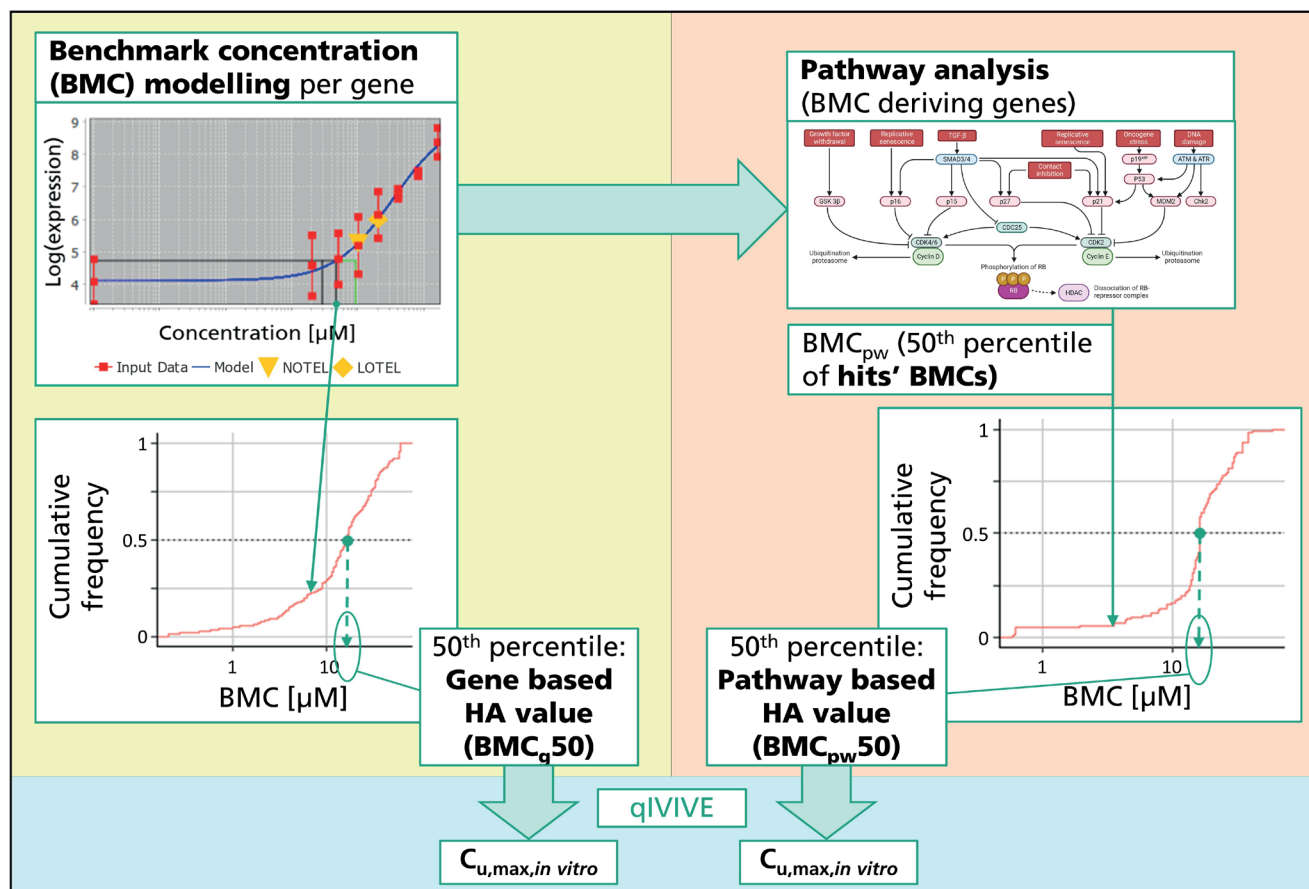


Fig. 1: Benchmark concentration modelling for transcriptomics-based toxicity thresholds

In the scope of this study, the term HA value was defined as the threshold of activation (i.e., a concentration or dose) derived from an individual assay which shall be integrated with other thresholds of activation following the rationale of an IATA to derive a PoD for risk assessment. The terms “gene” and “transcript” are used synonymously in this context. DEG expression data were processed in BMDExpress 2.3 (Sciome) (Phillips et al., 2019) along the lines of Ramaiahgari et al. (2019) and NTP (2018) to derive a BMC estimate for each DEG. In brief, the Williams trend test with a 0.05 p-value cutoff with Benjamini-Hochberg correction was used to prefilter DEGs. Next, eight mathematical models (hill, power, linear, polynomial 2, exponential 2°-5°) were fitted to these prefiltered data. Best fitting models with a goodness-of-fit p-value greater than 0.1 were used to determine the genes’ BMCs (BMC_g) with a BMR of one standard deviation from vehicle control level. Uncertain BMC_g values ($BMC/BMCL \geq 20$; $BMC \leq \text{lowest tested concentration}/10$ or $BMC \geq \text{highest tested concentration}$) were excluded.

The gene-based HA value corresponds to the median of BMC_g values (BMC_{g50}). For pathway-level derivation of HA values, we adapted the approach described by the NTP (2018) for determining gene-set level potencies. In brief, the functional classification method implemented in BMDExpress 2.3 was used to enrich genes with valid BMCs in Wikipathways, KEGG, and Reactome

pathways, which had been retrieved from ConsensusPathDB’s human biological pathways dataset (Kamburov et al. (2012), release 35; with EntrezID annotations). Pathways in which at least three genes met the aforementioned criteria for valid gene-level BMC and these genes (“hits”) made up at least 10% of the pathway population of possible hits were considered biologically significantly enriched. The median BMC (BMC_{pw}) of hits was calculated for each enriched pathway. As for the gene-based HA value, the pathway-based HA value corresponds to the median of BMC_{pw} values (BMC_{pw50}), which was suggested earlier for PoD selection for compounds with unknown MoA (Farmahin et al., 2017; Webster et al., 2015).

For quality control, we investigated qualitative differences between assays and treatments that might indicate impaired robustness of the transcriptional response and of the HA values derived further in the data processing pipeline. Measures considered relevant include (1) the shape of the concentration-response curves in the number of DEGs (N_{DEG}), (2) the relationship between N_{DEG} , the number of concentration-responsively expressed DEGs ($N_{DEG,cr}$), and the number of enriched pathways. Along the lines of Baltazar et al. (2020), BMC_{pw50} s based on 20 enriched pathways or less ($N_{pw} \leq 20$) were considered uncertain. Equally, BMC_{g50} s based on 20 BMC_g s or fewer were not expected to be reliable.



Alongside BMC_{g50} and a BMC_{pw50} from each transcriptomics assay, HA values included the thresholds of activation of each reporter or functional assay as well as cytotoxicity thresholds. Their integration was exemplified by assigning the lowest HA value to be used as the PoD, thereby adhering to a worst-case approach.

2.4 Toxicokinetic modelling

Forward dosimetry applied to in vivo PoDs: *In vivo* LOELs (Tab. 1) were used to determine plasma $C_{u,max}$ of the test compounds using a one-compartment pharmacokinetic model. Physicochemical properties of the compounds (molecular weight, pK_a , $\log K_{ow}$) were sourced from ChEMBL (EBI) (Gaulton et al., 2017) or EPI Suite™ (US EPA, 2021a) databases. Fraction unbound in plasma (f_u) was predicted using the equation defined by Lobell and Sivarajah (2003). The number of hydrogen bond donors and the polar surface area of the compounds were used to predict the absorption rate constant (k_a , h^{-1}) and fraction absorbed (F_a) using the first-order absorption model available in the Simcyp simulator V20 (Jamei et al., 2009). As no transporter or *in vitro* metabolism data is available for these compounds, the fraction of drug escaping the first pass gut metabolism (F_g) and first pass hepatic metabolism (F_h) were assumed to be 1, giving a worst-case scenario for first pass metabolism/biliary clearance. The volume of distribution of the compounds (V_{ss} , L) was predicted in the Simcyp simulator using a modification of the published model described by Rodgers and co-workers (Rodgers et al., 2005; Rodgers and Rowland, 2006). This method considers the potential difference across the cell membrane and allows the ionic fraction of the drugs to permeate the cell membrane depending on the potential difference across the cell. Thus, the rate of permeation of ionized drugs into or out of the intracellular water depends on the inherent permeability of the ion, charge, concentration gradient, and the membrane potential (Fisher et al., 2019). Average values of glomerular filtration rate (GFR, L/h) and weight (kg) were based on parameters of the rat in Simcyp. The compounds were assumed to be cleared only through passive renal filtration of the unbound fraction in plasma as no information was available in the literature regarding intrinsic hepatic clearance of the compounds. The systemic clearance (CL) of the compounds was thus predicted as $f_u \cdot GFR$. The elimination rate constant (k_e , h^{-1}) was calculated as $k_e = CL/V_{ss}$. Time at maximum concentration (t_{max} , h) of the compounds was predicted in R using the following equation (1):

$$t_{max} = \frac{\ln \left(\frac{k_a \cdot (1 - e^{k_e \cdot \tau})}{(k_e \cdot (1 - e^{k_a \cdot \tau}))} \right)}{k_a - k_e} \quad (1)$$

where τ is the dosing interval of the compound and was taken as 24 h.

The bioavailability (Bioav) of each compound was calculated as $F_a \cdot F_g \cdot F_h$. Bioav was used in the following equation (2) to predict C_{max} (ng/mL or μM) of the compounds:

$$C_{max} = \frac{Bioav \cdot dose \cdot k_a}{V_{ss}(k_a - k_e)} \cdot (e^{-k_e \cdot t_{max}} - e^{-k_a \cdot t_{max}}) \quad (2)$$

Biokinetics modelling applied to in vitro HA values: The unbound fraction of C_{max} ($C_{u,max}$, μM) in culture medium in *in vitro* assays was derived from the nominal concentrations (C_{nom} , in media) by biokinetic modelling using the VIVD model available in Simcyp's SIVA toolkit (v4.0) (Fisher et al., 2019). The VIVD model considers the lipid and protein binding in culture medium, binding to cell culture plastic and air partitioning to predict free concentrations in culture medium.

Finally, $C_{u,max}$ in culture medium was compared to *in vivo* rat plasma $C_{u,max}$. This is a valid approach under the assumption that rat and human organs are equally sensitive to the test compound, steady-state distribution is achieved *in vitro*, and there is no permeability restriction on the distribution of the unbound fraction across the cell membrane *in vitro*, or between tissue and plasma *in vivo*.

3 Results

3.1 Transcriptomics analysis

A pronounced and generally concentration-responsive transcriptional response was elicited by VPA, rotenone, and the butyl-tin substances, but not by thioureas, imidazoles, and butanone oxime. This led us to differentiate between active and inactive compounds. The number of differentially expressed genes (N_{DEG} ; applying $p\text{-adj} < 0.05$ and $|\log 2FC| > 1.5$) (i.e., a representation of the activity level) surpassed 250 at the highest tested sub-cytotoxic concentration in at least three test systems for all the active substances, except for MBTC, which is the least toxic of the butyl-tin compounds as determined *in vivo*. N_{DEG} was largely consistent in HepG2, PHH, RPTEC/TERT1, and LUHMES, except for a relatively low response (N_{DEG}) of HepG2 to DBTC, of PHH to TBTC, and of LUHMES to rotenone (Fig. 2A). PBEC results were included in this comparison only under reserve due to their impaired reliability, which is a consequence of experimental irregularities: PBEC solvent control samples were missing on some plates, which would have been required to fully control batch and plate effects. Further, contamination with Triton X-100, which was used as positive control for the LDH cytotoxicity detection kit, cannot be excluded for test compound solutions.

Genes that are differentially expressed at some test substance concentrations do not necessarily exhibit a concentration-responsive expression, and concentration-responsive expression need not involve differential expression at every individual test concentration. However, for test substances eliciting a concentration-response in the number of DEGs, we expected the number of concentration-responsively expressed DEGs ($N_{DEG,cr}$) to roughly equal the number of DEGs at the highest tested concentration ($N_{DEG,high}$), as most of the concentration-responsively expressed DEGs were expected to be differentially expressed at the highest tested concentration of the sub-cytotoxic test concentration regime adhered to in this study. Indeed, in the group of active compounds, an average of 90% of the concentration-responsively expressed DEGs were differentially expressed at the highest tested concentration (data not shown) and $N_{DEG,cr}$ (Tab. 3) was generally similar to $N_{DEG,high}$ with the exception of PHHs ex-

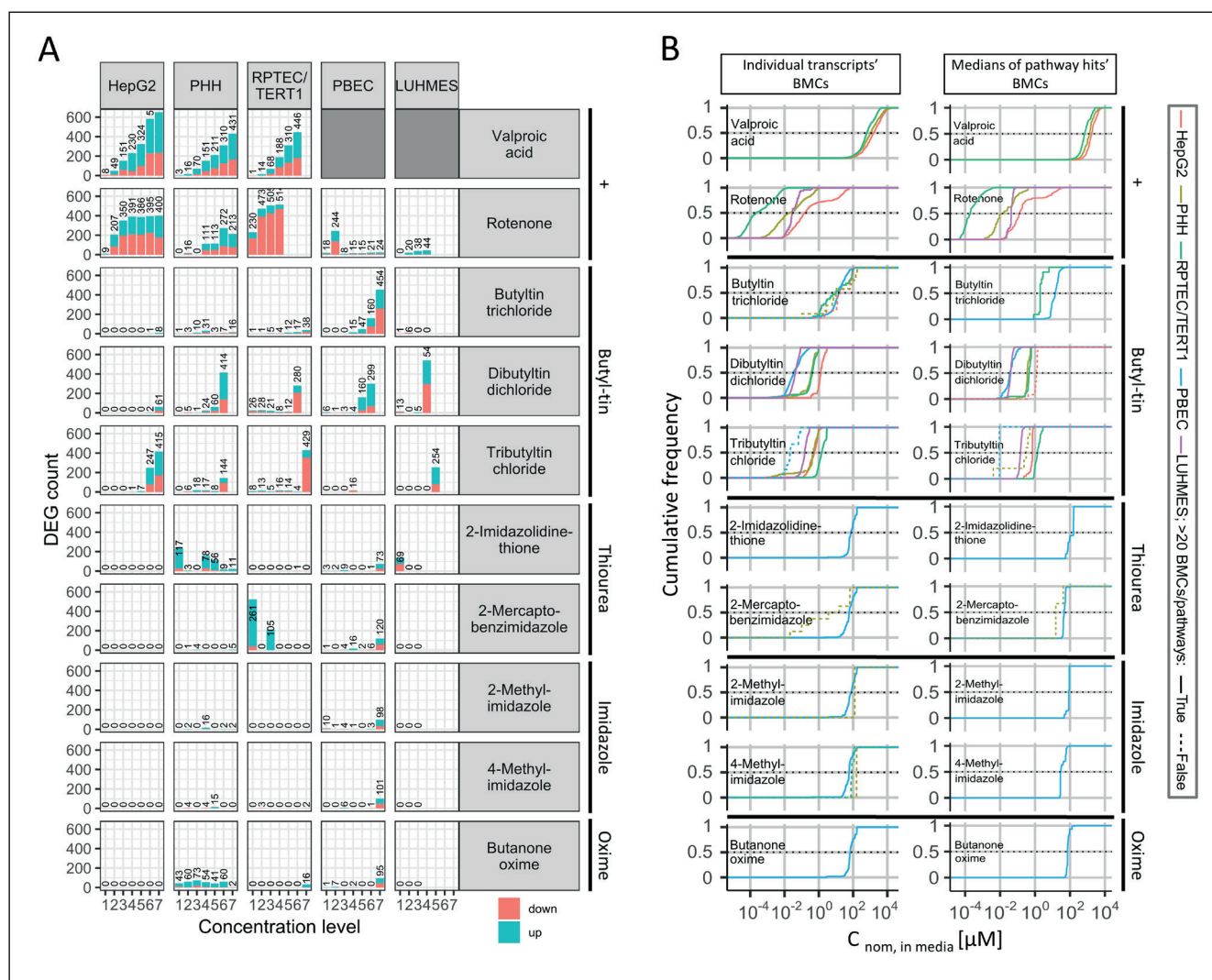


Fig. 2: Concentration-response assessment per test compound and transcriptional assay/cell type

(A) Number of DEGs ($p\text{-adj} = q < 0.05$, $\log_2FCI > 1.5$) per test compound, assay, and test concentration level. Numbers on top of stacked bars indicate the total count at the respective concentration level. Red, DEGs with $\log_2fc < -1.5$; green, DEGs with $\log_2fc > 1.5$; grey boxes, not tested. (B) Accumulation of benchmark concentrations (BMCs) of concentration-responsive genes, which were differentially expressed ($p\text{-adj} = q < 0.05$, $\text{abs}(\log_2fc) > 1.5$) at least at one tested concentration. Dotted lines indicate that no more than 20 gene-level BMCs or pathway-level median BMCs were found; $C_{nom, in media}$, nominal *in vitro* concentration.

posed to DBTC and TBTC as well as HepG2 cells exposed to TBTC, where the ratio of $N_{DEG, cr}/N_{DEG, high}$ was 0.4, 0.5, and 0.4, respectively.

For the active compounds including the two positive controls, also the number of pathways enriched by concentration-responsively expressed DEGs (N_{pw}) was in a similar range as $N_{DEG, cr}$ (and $N_{DEG, high}$). Accordingly, we found $N_{pw} \gg 100$ in at least two test systems for all active substances except for MBTC, where the response was generally weaker (Tab. 3). N_{pw} correlated with $N_{DEG, cr}$ across all substances and assays (Pearson = 0.95).

In contrast to the active substances, thiourea and imidazole compounds as well as butanone oxime evoked either little dif-

ferential expression ($N_{DEG} \ll 200$ at every concentration level) or one that lacked a positive correlation with increasing test compound concentration (Fig. 2A).

In the group of rather inactive compounds, the maximum $N_{DEG, cr}$ was much lower ($\ll 200$; Tab. 3). In fact, all $N_{DEG, cr}$ and N_{pw} were zero or close to zero in all assays but PBEC.

HA values were derived for all gene- and pathway level accumulations of BMCs (Fig. 2B). Marginally activated assays ($N_{DEG, cr} \leq 20$ and $N_{pw} \leq 20$) as well as PBEC results were considered to be of low reliability (as described above). Corrected and extrapolated HA values (plasma $C_{u, max}$) reveal an increasing potency among butyl-tin substances with an increasing number of butyl groups (MBTC $\ll$ DBTC $<$ TBTC) in all transcriptom-



Tab. 3: Number of concentration-responsively expressed DEGs ($N_{\text{DEG, cr}}$) per test compound and assay as determined by BMC modelling; and number of pathways (N_{pw}) enriched by the prior (in brackets) ($N_{\text{DEG, cr}}$ (N_{pw}))

	HepG2	PHH	RPTEC	PBEC	LUHMES
Valproic acid	735 (1099)	501 (684)	451 (702)	NA	NA
Rotenone	535 (658)	298 (135)	463 (587)	0 (0)	48 (60)
Butyltin trichloride	7 (0)	12 (0)	32 (22)	421 (409)	0 (0)
Dibutyltin dichloride	59 (12)	153 (88)	261 (162)	280 (129)	533 (734)
Tributyltin chloride	160 (79)	72 (5)	421 (361)	12 (1)	230 (298)
2-Imidazolidinethione	0 (0)	0 (0)	0 (0)	101 (65)	0 (0)
2-Mercaptobenzimidazole	0 (0)	8 (3)	0 (0)	138 (54)	0 (0)
2-Methylimidazole	0 (0)	1 (0)	0 (0)	124 (43)	0 (0)
4-Methylimidazole	0 (0)	2 (0)	3 (0)	109 (75)	0 (0)
Butanone oxime	0 (0)	0 (0)	0 (0)	112 (87)	0 (0)

NA, not applicable

ics assays (Tab. 4). Rotenone was approximately as potent as TBTC. VPA showed much response but only at high concentrations. Thiourea- and imidazole compounds as well as butanone oxime elicited only marginal transcriptional response ($N \leq 20$, i.e., no more than 20 genes or pathways) or response with decreased reliability (PBEC, as described above).

Pathway-based HA values ($\text{BMC}_{\text{pw}50}$) were generally similar to gene-based HA values ($\text{BMC}_{\text{g}50}$) from the same assay and test compound with less than 60% difference in clearly activated assays ($N_{\text{DEG, cr}} > 20$ and $N_{\text{pw}} > 20$) except for RPTEC/TERT1 exposed to MBTC, where $\text{BMC}_{\text{pw}50}$ was a factor 4.9 lower than $\text{BMC}_{\text{g}50}$. Due to the high similarity, only the lower and thereby more conservative value was considered for PoD analysis.

3.2 Reporter assays and functional assays

In agreement with the gene expression pattern also in these assays, VPA, rotenone, and the tin compounds induced most of the types of perturbations assessed in reporter assays and functional assays, including activating reporters for oxidative stress and DNA damage, decreasing neurite outgrowth in LUHMES cells (only rotenone), and causing mitochondrial dysfunction (only rotenone and DBTC). Corrected and extrapolated thresholds of activation (plasma $C_{\text{u,max}}$) are depicted in Figure 3. Reporter assays targeting endocrine activity (PR-anti and TR β) were only activated by VPA (Fig. 3A). Response to VPA, which has been shown to induce liver steatosis (Abdel-Dayem et al., 2014; Escher et al., 2022a), was most sensitive in CALUX-based ESRE and TCF stress signaling assays and the PPAR α and PXR assays (plasma $C_{\text{u,max}} = 1.9 \log_{10} \mu\text{M}$), and at a just slightly higher concentration in the p53 (U2OS) assay. Activation of PPAR α and PXR are MIEs in the adverse outcome pathway (AOP) network for microvesicular liver steatosis (Escher et al., 2022a). The ESRE assay indicates endoplasmic reticulum stress, which is a key event (KE) in the same AOP network.

As expected, based on the MoA of rotenone, assays dedicated to mitochondrial toxicity and neuronal toxicity (neurite outgrowth in LUHMES, and to a much lesser extent LUHMES transcription) sensitively detected rotenone's toxicity. Similar thresholds of activation for rotenone were found in the p21 and ICAM1 assays. However, RPTEC/TERT1 reacted more sensitively to rotenone with concentration-responsive differential gene expression of hundreds of genes.

The group of organo-tin substances follows the potency pattern already described for *in vivo* repeated dose studies and transcriptomics assays, with TBTC > DBTC > MBTC. The most sensitive reporter assays or functional assays in terms of plasma C_{max} (unbound) for TBTC, DBTC, and MBTC were LXR ($-6.2 \log_{10} \mu\text{M}$), Nrf2 ($-4.3 \log_{10} \mu\text{M}$), and AKR1B10 ($-1 \log_{10} \mu\text{M}$), respectively. Besides, the same potency ranking is evident in assays activated by more than one of the tin substances such as SRXN1 GFP, the neurite outgrowth assay, and several assays that were activated by TBTC and DBTC but did not show a specific response to MBTC. However, a few reporter assays do not follow the same potency ranking. DBTC was most potent in the Nrf2 assays. Further, DBTC activated several assays (AKR1B10, p53 (HepG2), AP1, BIP, HSPA1B, ICAM1) that were not activated by TBTC. Finally, in the BTG2 assay, MBTC and TBTC showed activity, whereas DBTC was inactive.

DBTC and TBTC activated PPAR γ , confirming their potential to act as endocrine disruptors (DBT: Chamorro-Garcia et al., 2018, TBT: Chamorro-Garcia et al., 2013). As expected, neither DBTC nor TBTC activated the glucocorticoid agonist and antagonist assays. It has previously been found that DBT inhibits ligand binding to the glucocorticoid receptor and its transcriptional activity, thereby disturbing metabolic functions and modifying immune responses (Gumy et al., 2008).

The test compounds already described as inactive in transcriptomics assays, namely thiourea and imidazole compounds as well as butanone oxime, showed no activity in functional assays either

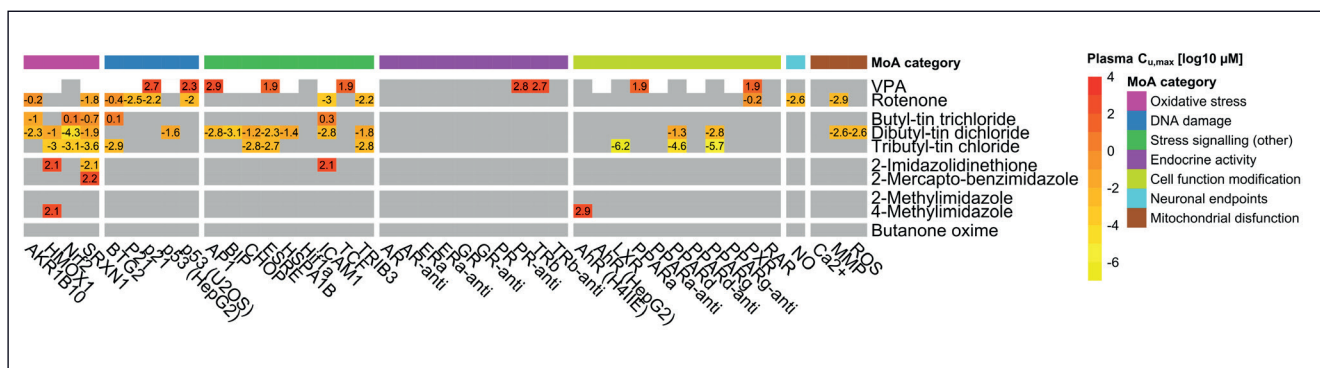


Fig. 3: Threshold of activation for reporter genes and functional readouts in assays without metabolic activation

Numbers indicate maximum unbound concentration in plasma, $C_{u,max}$ [$\log_{10} \mu M$]. Plasma $C_{u,max}$ were predicted from nominal *in vitro* concentrations by quantitative *in-vitro-to-in-vivo* extrapolation (qIVIVE) using the VIVD model. Read-outs are categorized according to the type of mode of action (MoA) they cover. Grey cells: inactivity; white cells: assay not applied.

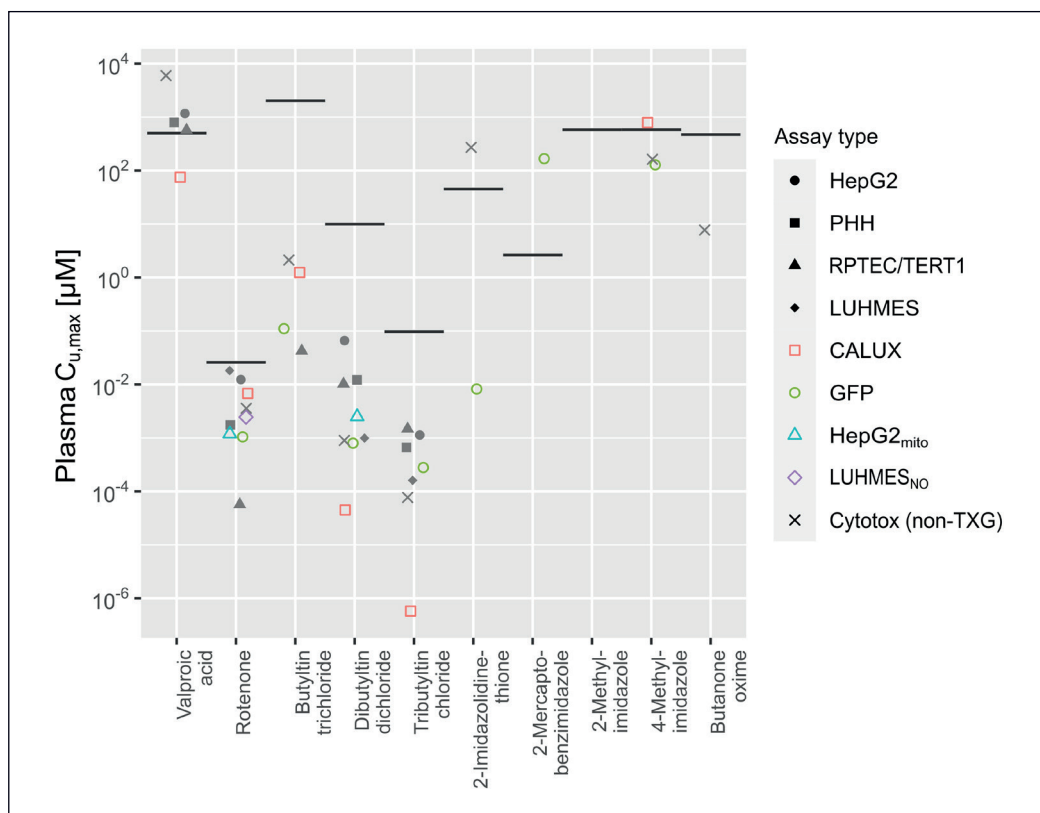


Fig. 4: Jitter plot of unbound plasma concentrations ($C_{u,max}$) corresponding to the lowest HA value per assay type and test compound

Filled markers represent min(BMC_{g50}, BMC_{pw50}) for transcriptional assays, unfilled markers indicate min(MECs) per reporter or functional assay type or for cytotoxicity thresholds of reporter and functional assays. HA values with low reliability (including PBEC and marginally activated transcriptomics assays) are not shown. Solid horizontal lines represent the result of forward dosimetry for *in vivo* LOELs.

and activated only few reporter assays. The few responding reporters included HMOX1 and SRXN1 (oxidative stress), ICAM1 (stress signaling), and AhR (H4IIE; cell function modification).

2-IT elicited a response of SRXN1 at a plasma $C_{u,max}$ of $-2.1 \log_{10} \mu M$ (Fig. 3), a concentration about 4 orders of magnitude lower compared to the activation of two other stress signaling reporters, namely HMOX1 (oxidative stress) and ICAM1 (stress signaling). 2-MBI, which is of similar toxicity *in vivo*, activated SRXN1 at a similar plasma concentration ($C_{u,max} = 2.2 \log_{10} \mu M$; Fig. 3).

Metabolic detoxification was found for the two active substances DBTC and TBTC as well as for VPA and rotenone, whereas all other test compounds did not show any activity in the assays addressing compound metabolism (data not shown).

3.3 Point of departure analyses

Forward dosimetry of oral *in vivo* doses allowed us to derive plasma $C_{u,max}$ from *in vivo* LOELs. Thereby we could compare them to the *in vitro* HA values equally expressed as plasma



Tab. 4: Comparison of maximum unbound plasma concentrations ($C_{u,max}$ (log₁₀ μ M)) corresponding to selected *in vitro* HA values and to LOEL values of high-quality *in vivo* studies (Tab. 1) obtained by *in-vitro-to-in-vivo* extrapolation and forward dosimetry, respectively

Bold numbers indicate HA values lower than the corresponding *in vivo* LOEL. The lowest fully reliable HA value per test compound is highlighted in green. Grey numbers indicate HA values based on 20 or fewer genes or pathways, dark grey fields mark experiments not performed, empty fields indicate inactivity. The difference between the most sensitive NAM and the *in vivo* PoD is used as a measure of protectiveness of the *in vitro* test battery with (more) negative values indicating (more) conservative *in vitro* HA values.

Assay type	Cell type	HA value	Positive control		Butyltin			Thiourea		Imidazole		Oxime
			VPA	Rotenone	MBTC	DBTC	TBTC	2-IT	2-MBI	2-MI	4-MI	Butanone oxime
<i>in vitro</i> transcriptional	HepG2	BMC _g 50	3.1	-1.8	-0.2	-1.2	-2.9					
		BMC _{pw} 50	3.1	-1.9		-1.2	-2.8					
	PHH	BMC _g 50	2.9	-2.7	-0.5	-1.9	-3.2		0.5	2.1	2.2	
		BMC _{pw} 50	3.0	-2.8		-1.9	-3.4		1.2			
	RPTEC/TERT1	BMC _g 50	2.8	-4.2	-0.7	-2.0	-2.8			1.9		
		BMC _{pw} 50	2.8	-4.2	-1.4	-1.9	-2.8					
	PBEC ^a	BMC _g 50			-0.5	-3.3	-6.1	0.4	-0.1	0.4	0.3	0.3
		BMC _{pw} 50			-0.5	-3.2	-6.5	0.7	-0.2	0.4	-0.1	0.3
	LUHMES	BMC _g 50		-1.7		-3.0	-3.8					
		BMC _{pw} 50		-1.5		-3.0	-3.8					
<i>in vitro</i> reporter and functional	CALUX ^b	min MEC	1.9	-2.2	0.1	-4.3	-6.2				2.9	
	GFP	min MEC		-3.0	-1.0	-3.1	-3.6	-2.1	2.2		2.1	
	HepG2 _{mito} ^b	min MEC		-2.9		-2.6						
	LUHMES _{NO}	min MEC		-2.6								
	Cytotoxicity ^c	min MEC	3.8	-2.4	0.3	-3.0	-4.1	2.4			2.2	0.9
<i>in vivo</i> PoD			2.7	-1.6	3.3	1.0	-1.0	1.6	0.4	2.8	2.8	2.7
Difference between most sensitive NAM and <i>in vivo</i> PoD			-0.8	-2.7	-4.7	-5.3	-5.2	-3.7	1.8	NA	-0.7	-1.8

^a PBEC results are shown for completeness but are considered less reliable due to experimental irregularities. HA values based on 20 or fewer genes or pathways as well as PBEC results are not taken into account for comparisons with *in vivo* LOELs. ^b CALUX and HepG2 mitochondrial dysfunction assays without metabolic activation. ^c cytotoxicity thresholds of reporter and functional assays.

BMC, benchmark concentration; g, gene; pw, pathway; MEC, minimum effective concentration; NA, not applicable

$C_{u,max}$ (Tab. 4, summarized in Fig. 4) to assess the protectiveness of individual assays and the overall approach to testing and assessment. In the present study, we defined that protectiveness of a given assay or approach is achieved if it delivers a toxicity threshold lower than or equal to the *in vivo* reference values derived from preclinical animal studies.

One very encouraging result of this case study is that based on plasma $C_{u,max}$ the most conservative of all HA values derived using the present test battery is a factor of 6.7 to 222,000 lower than the *in vivo* LOEL for each active substance and three of five inactive substances (Fig. 4). In fact, several assays derive protective HA values for each of the active substances, reinforcing that the chosen test battery is appropriate for them.

It was no surprise that none of the assays or assay types presented as the single most sensitive assay overall. However, the CALUX test battery and RPTEC/TERT1 stand out with protective HA values for 5/5 and 4/5 active substances, respectively.

Further, RPTEC/TERT1 derived close-to-protective HA values for the remaining active substance. The SRXN1 assay performed similarly with protective HA values for 4/5 active substances. HepG2, PHH, and LUHMES transcriptional assays as well as Nrf2, ESRE, ICAM1, and TRIB3 each derived protective HA values for 3/5 active substances.

However, no activation or – in transcriptomics assays – only marginal activation at sub-cytotoxic levels leaves the inactive substances with few (2-IT, 4-MI) or no protective HA value (2-MBI, 2-MI) or only cytotoxicity deriving a protective HA value.

4 Discussion

The required minimal testing scope in NAM-based hazard assessment regarding complex endpoints like repeated dose systemic toxicity depends, for example, on the problem formulation, differing

between a prioritization or screening context and use as replacement of an *in vivo* study. In both cases, the required testing scope is unclear to date. In this study we focus on protection, while putting aside the goals of hazard identification and predicting the mechanism of action (Kavlock et al., 2018). This strategy assumes that *in vitro* derived BMCs, at which the onset of a biological perturbation is seen, occur generally at lower concentrations compared to the corresponding bioavailable concentrations from preclinical *in vivo* studies at which apical effects start to appear. Our study aims to better define the minimal scope of an *in vitro* test battery in terms of coverage of biological mechanisms so that protective *in vitro* HA levels can be derived with some confidence. In this context, it should be noted that NGRA does not aim to replace animal studies organ-by-organ or effect-by-effect. On the contrary, it requires new assessment concepts to derive protective and sufficiently robust thresholds of toxicity. Further, the most essential part of the study scope of repeated dose *in vivo* studies is encouragingly limited, with only a small set of target organs such as liver, kidney, and the respiratory tract frequently affected at the LOEL. Most other target organs show effects only at equal or higher dose level or start to be affected just one dose level lower (Batke et al., 2013).

The ivTB applied in the present study includes a set of largely non-targeted transcriptomics assays for comprehensive coverage of effects and responses in main target cell types, together with assays targeting a broad range of known MIEs and phenotypic effects. In a conservative screening approach, effect levels were integrated and extrapolated to internal exposure concentrations *in vivo*, which were compared to corresponding values based on *in vivo* study LOELs. Transcriptome data are usually measured to obtain first insights into the mechanism of action of a test compound. Therefore, assessing changes in the transcriptome is a well fitted tool for the standard situation of chemical safety assessment in which little or no evidence is available related to the MoA.

In this case study, the number of DEGs provided an exploratory layer for investigating the onset of biological perturbation and concentration-response in the cell-based systems representing main target organs. In this analysis, we reasoned that a concentration-response in the number of DEGs at sub-cytotoxic concentrations may be indicative of a high validity of the transcriptional response with respect to the goal of deriving a threshold representing significant biological perturbation. This assumption was reinforced by the observation that in all tests that showed a clear concentration-responsive increase in the number of DEGs spanning several tested concentrations, such as VPA-treated HepG2, PHH, and RPTEC/TERT1, we also observed the number of concentration-responsive DEGs to be similarly high as the number of DEGs at the highest tested concentration ($N_{\text{DEG, cr}} \sim N_{\text{DEG, high}}$). In contrast, relatively low numbers of concentration-responsive DEGs ($N_{\text{DEG, cr}} \ll N_{\text{DEG, high}}$) always coincided either with a steep concentration-response in the number of DEGs, as e.g., for DBTC in PHH, or with a failure to detect a clear concentration-response in the number of DEGs, as e.g., for rotenone in PBEC. Both of these conditions are probably not ideal to produce precise and robust HA values, so transcriptional response with $N_{\text{DEG, cr}} \ll N_{\text{DEG, high}}$ should be used with increased caution.

The distribution of BMCs for transcriptional changes was analyzed with the intention to derive a threshold describing potency differences of test compounds in a robust and untargeted way. Several approaches have been proposed to derive toxicologically significant HA values based on transcriptional response concentrations, and this is still an active field of research (Farmahin et al., 2017; Gant et al., 2023; Harrill et al., 2021; Reardon et al., 2023). Generally, approaches build on estimates of (1) the central tendency or (2) the lower bound of the distribution of gene or pathway-based BMCs, or (3) on a predefined, absolute-rank BMC value. Recent examples include the mean BMC “of genes between 25th and 75th percentile” (Baltazar et al., 2020) of all BMC_{gs}, and the lowest transcriptional pathway benchmark doses (BMD) *in vivo*, which were found to correlate (Pearson) well with apical BMDs from the same time point (Thomas et al., 2013). Resulting HA values were found to be similar for several type (1) and (3) approaches, but those HA values building on an estimate of the central tendency of gene BMCs across all pathways were suggested to be particularly robust (Farmahin et al., 2017). Type (2) approaches may be appropriate for targeted approaches aiming to determine concentrations at which processes closely related to the MIE happen but may require additional adjustments when used to predict apical effects as is the case in this study. Further, in an untargeted approach, we cannot restrict ourselves to genes and responses corresponding to specific (known) MoA. Moreover, unlike in targeted approaches, for example, using hepatocytes to detect liver injury (Ramaiahgari et al., 2019), we cannot define the significance of the extent of a transcriptional response with reference to the extent of occasional perturbation elicited by compounds known to have an entirely unrelated MoA. Therefore, we settled for BMC_{g50} and BMC_{pw50}, two closely related approaches based on the central tendency of all genes and pathways. The approach of choosing BMC_{g50} and BMC_{pw50} as transcriptomics-derived HA values was considered less sensitive to outliers than approaches relying on the lowest BMCs (Farmahin et al., 2017). Another potential benefit of not relying on the lowest BMCs relates to the goal of having an approach that works for data-poor chemicals with no prior mechanistic information available (Webster et al., 2015). Supposing that for some chemicals very specific transcriptional effects or responses happen at concentrations much lower than the more general stress response, but for other chemicals this is not the case, we considered relying on estimates of central tendency among transcriptomics BMCs beneficial to even out such differences. In fact, we assume that BMC_{g50} and BMC_{pw50} represent rather generic xenobiotic responses that would be shown for most treatments by all tissues in a similar or foreseeably different way. The possibility that individual test substances could behave differently (e.g., not elicit such generic responses) must be accommodated for by applying appropriate assessment factors.

One concern with estimates of central tendency of transcriptional response such as BMC_{g50} and BMC_{pw50} may be that they constitute a deviation from the general rule of taking conservative assumptions that this hazard assessment approach adhered to otherwise. However, in this study the concern was not confirmed. Both approaches resulted in HA values lower or in a similar range as *in vivo* PoDs and most often comparable to those of



most reporter assays. If they are found not to be protective for some chemicals in further studies, then one can still introduce assessment factors.

VPA, rotenone, and the tin compounds elicited a profound transcriptional response. All other substances showed very limited response in most transcriptomics assays, which was interpreted in line with previous work (Baltazar et al., 2020) as minimal cellular effects and responses. Transcriptomics-based thresholds could be generated in PBEC for these rather inactive compounds. Interestingly, even these weakly founded HA values are protective. Moreover, we could confirm the finding of Farmahin et al. (2017) and others that the median transcriptomic BMCs of all pathways were lower or at least in the same range as apical systemic *in vivo* endpoints. Again, the value of the present study is not that it validated the broad coverage of chemical space. Instead, our results highlight that in principle surveillance of transcriptional changes in a small number of cellular models can be sufficient to represent a broader range of target organs and to derive protective systemic toxicity PoDs.

As discussed earlier, besides transcriptomics assays the EU-ToxRisk ivTB includes reporter assays and functional assays dedicated to sensitively detect specific molecular events and more complex endpoints, respectively. In this case study, thresholds of toxicity derived in reporter assays and functional assays were generally in a similar range as transcriptome-based HA values for active substances. This may indicate a broad activity of the test substances and some robustness of the derived PoDs. However, some *in vitro* PoDs were based on outliers orders of magnitude lower than the remaining HA values or *in vivo* PoDs (e.g., RPTEC/TERT1 for rotenone, LXR and PPAR γ for TBTC, and oxidative stress reporter SRXN1 for 2-IT). The case of VPA shows that these may not generally be considered overprotective.

Overall, we can discern two scenarios, one where several assays respond to a subcytotoxic treatment and another where very few do. In the first scenario, i.e., for the active substances, several assays derive protective HA values, thereby allowing to derive PoDs with sufficient confidence. Here, it appears feasible to even deviate from the most conservative approach based on the most sensitive HA value. The latter is in some cases orders of magnitude lower than all other HA values and clearly overprotective. Therefore, in this case deriving a PoD based on, e.g., the 10th percentile of HA values might significantly decrease overprotectiveness while still assuring protectiveness.

The second scenario, with very few responding assay points is a limitation of the test battery chosen in this study. No protective HA value could be derived for two test substances (2-MBI and 2-MI). For another substance, butanone oxime, a protective PoD was based solely on cytotoxic effects. Apparently, more targeted assays are needed to reliably cover these compounds' toxic actions. Especially hematological effects, but to a smaller degree also thyroid effects, are frequently reported in subchronic *in vivo* rodent studies at the LOEL. Batke et al. (2013) found hematological effects and thyroid toxicity in about 30% and 1% of studies with comprehensive study scope (N = 88) as well as 16% and 5% in studies with limited study scope (N = 56; performed before the

publication of OECD test guidelines), respectively. Further, they reported that hematological effects occur frequently as the single most sensitive finding in *in vivo* studies, while thyroid effects are rarely observed in isolation (about 11% and none in the same sample of studies with comprehensive study scope for hematological effects and thyroid toxicity, respectively). This indicates that it is of higher importance to adjust future NAM-based approaches to quantitative HA to cover hematological effects than to cover thyroid effects.

Hematotoxicity can result from cytotoxicity towards mature blood cells or effects on hematopoietic stem/progenitor cells (Mahalingaiah et al., 2018). Further research will be needed to better determine the most important mechanisms leading to hematotoxicity in chemical safety and to design appropriate (e.g., high-throughput) assays.

In vitro-based derivation of PoDs may be particularly challenging for effects concerning thyroid hormone homeostasis, especially because the significance of thyroid hormone signaling varies largely across life stages and is still not fully understood (Noyes et al., 2019). Therefore, the applicability of *in vitro* assays remains limited to screening approaches, although several MIEs relating to thyroid hormone homeostasis have already been addressed in *in vitro* screening approaches including thyroperoxidase inhibition (Paul Friedman et al., 2016; Noyes et al., 2019), a presumed mechanism of action of both thiourea compounds used in this study (Maranghi et al., 2013; Norford et al., 1993).

Also, butanone oxime, which elicited only cytotoxicity in our assays, may have a very particular and therefore hard-to-detect MoA. Its hematotoxicity has been postulated to be mediated by GSH depletion through conjugation of butanone oxime or its metabolites to GSH followed by reactive oxygen species formation, oxidative stress, and subsequent hemolysis and methemoglobinemia (Yamada et al., 2022; Palmen and Evelo, 1998).

As detailed above, 2-MBI critically affects the thyroid, whereas 2-MI and butanone oxime both showed foremost hematological effects. However, both the thiourea and the imidazole group of compounds, which each elicited relatively homogenous adverse outcomes *in vivo*, include test substances for which protective HA values were derived in this case study (2-IT, 4-MI). 2-MBI, 2-MI, and butanone oxime may exhibit mechanisms of action that require more specifically dedicated assays than 2-IT and 4-MI to be covered with confidence, but even for the latter only a single oxidative stress GFP assay (SRXN1 and HMOX1 for 2-IT and 4-MI, respectively) provided a protective threshold of toxicity. Therefore, even for 2-IT and 4-MI the protectiveness may be less robust than for the active substances.

To sum up, we sought to define a panel of assays that can serve as a solid base for a MoA-agnostic approach to derive protective PoDs for systemic toxicity. As most *in vitro*-based HA values were protective in this study, the assay sensitivity appeared to be a less important criterion for inclusion in the test battery than the question, which assay could deliver protective HA values for the highest number of test substances. Although we can only give a very limited answer to this second question due to the small number of compounds tested, CALUX and RPTEC/TERT1 were particularly frequent and sensitive responders.

The idea of replacing animal-based toxicity testing with NAM-based approaches promises among others a more thorough assessment of the vast majority of (anthropogenic) chemicals that surround us and to overcome limitations due to inter-species differences. However, it also holds new challenges including the need to provide fit-for-purpose alternatives for the whole-organism perspective. In this study we highlighted that it may not only be impossible to fully reproduce the organism's complexity but also unnecessary. We showed that using a limited set of *in vitro* assays we could derive protective HA values for compounds with critical apical *in vivo* effects in target organs not represented by the cell types used in our *in vitro* assays. A main short-coming of our approach was that several test compounds provoked only a very limited response in our assays. Here, the question remains whether response in only a small fraction of assays applied can provide adequately robust PoDs for risk assessment. The broader question of how to assess chemicals with low *in vitro* activity will be one of several questions to be addressed more closely in the EU Horizon 2020 project RISK-HUNT3R. This will help to further define the applicability domain and the required scope of testing as well as any need for assessment factors.

References

- Abdel-Dayem, M. A., Elmarakby, A. A., Abdel-Aziz, A. A. et al. (2014). Valproate-induced liver injury: Modulation by the omega-3 fatty acid DHA proposes a novel anticonvulsant regimen. *Drugs R D* 14, 85-94. doi:10.1007/s40268-014-0042-z
- Anonymous (1992). Toxicology and carcinogenesis studies of ethylene thiourea (CAS No. 96-45-7) in F344 rats and B6C3F1 mice (feed studies). *Natl Toxicol Program Tech Rep Ser* 388, 1-256. https://ntp.niehs.nih.gov/sites/default/files/ntp/htdocs/lt_rpts/tr388.pdf
- Anonymous (1999). NTP toxicity studies of methyl ethyl ketoxime administered in drinking water to F344/N rats and B6C3F1 mice (CAS No. 96-29-7). *Toxic Rep Ser* 51, 1-F9. https://ntp.niehs.nih.gov/sites/default/files/ntp/htdocs/st_rpts/tox051.pdf
- Anonymous (2012). N-Butylzinnverbindungen [MAK value documentation in German language, 2008]. In *The MAK-Collection for Occupational Health and Safety*. doi:10.1002/3527600418.mb68873verd0044
- Aschauer, L., Gruber, L. N., Pfaller, W. et al. (2013). Delineation of the key aspects in the regulation of epithelial monolayer formation. *Mol Cell Biol* 33, 2535-2550. doi:10.1128/mcb.01435-12
- Baltazar, M. T., Cable, S., Carmichael, P. L. et al. (2020). A next-generation risk assessment case study for coumarin in cosmetic products. *Toxicol Sci* 176, 236-252. doi:10.1093/toxsci/kfaa048
- Batke, M., Aldenberg, T., Escher, S. et al. (2013). Relevance of non-guideline studies for risk assessment: The coverage model based on most frequent targets in repeated dose toxicity studies. *Toxicol Lett* 218, 293-298. doi:10.1016/j.toxlet.2012.09.002
- Bitsch, A., Jacobi, S., Melber, C. et al. (2006). Repdose: A database on repeated dose toxicity studies of commercial chemicals – A multifunctional tool. *Regul Toxicol Pharmacol* 46, 202-210. doi:10.1016/j.yrtph.2006.05.013
- Callegaro, G., Schimming, J. P., Piñero González, J. et al. (2023). Identifying multiscale translational safety biomarkers using a network-based systems approach. *iScience* 26, 106094. doi:10.1016/j.isci.2023.106094
- Chamorro-Garcia, R., Sahu, M., Abbey, R. J. et al. (2013). Transgenerational inheritance of increased fat depot size, stem cell reprogramming, and hepatic steatosis elicited by prenatal exposure to the obesogen tributyltin in mice. *Environ Health Perspect* 121, 359-366. doi:10.1289/ehp.1205701
- Chamorro-Garcia, R., Shoucri, B. M., Willner, S. et al. (2018). Effects of perinatal exposure to dibutyltin chloride on fat and glucose metabolism in mice, and molecular mechanisms, in vitro. *Environ Health Perspect* 126, 057006. doi:10.1289/EHP3030
- Chan, P. C. (2004). NTP technical report on the toxicity studies of 2- and 4-methylimidazole (CAS No. 693-98-1 and 822-36-6) administered in feed to F344/N rats and B6C3F1 mice. *Toxic Rep Ser*, 1-G12. https://ntp.niehs.nih.gov/sites/default/files/ntp/htdocs/st_rpts/tox067.pdf
- Delp, J., Cediél-Ulloa, A., Suciu, I. et al. (2021). Neurotoxicity and underlying cellular changes of 21 mitochondrial respiratory chain inhibitors. *Arch Toxicol* 95, 591-615. doi:10.1007/s00204-020-02970-5
- Dent, M., Amaral, R. T., Da Silva, P. A. et al. (2018). Principles underpinning the use of new methodologies in the risk assessment of cosmetic ingredients. *Comput Toxicol* 7, 20-26. doi:10.1016/j.comtox.2018.06.001
- EFSA – European Food Safety Authority (2021). EFSA Strategy 2027: Science, safe food, sustainability. Publications Office of the European Union. doi:10.2805/886006
- EMA – European Medicines Agency (2020). EMA Regulatory Science to 2025. https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/ema-regulatory-science-2025-strategic-reflection_en.pdf
- Escher, S. E., Tluczkiewicz, I., Batke, M. et al. (2010). Evaluation of inhalation TTC values with the database repdose. *Regul Toxicol Pharmacol* 58, 259-274. doi:10.1016/j.yrtph.2010.06.009
- Escher, S. E., Mangelsdorf, I., Hoffmann-Doerr, S. et al. (2020). Time extrapolation in regulatory risk assessment: The impact of study differences on the extrapolation factors. *Regul Toxicol Pharmacol* 112, 104584. doi:10.1016/j.yrtph.2020.104584
- Escher, S. E., Aguayo-Orozco, A., Benfenati, E. et al. (2022a). Integrate mechanistic evidence from new approach methodologies (NAMs) into a read-across assessment to characterise trends in shared mode of action. *Toxicol In Vitro* 79, 105269. doi:10.1016/j.tiv.2021.105269
- Escher, S. E., Partosch, F., Konzok, S. et al. (2022b). Development of a roadmap for action on new approach methodologies in risk assessment. *EFSA Support Publ* 19, 7341E. doi:10.2903/sp.efsa.2022.en-7341
- Farmahin, R., Williams, A., Kuo, B. et al. (2017). Recommended approaches in the application of toxicogenomics to derive points of departure for chemical risk assessment. *Arch Toxicol* 91, 2045-2065. doi:10.1007/s00204-016-1886-5
- Fisher, C., Siméon, S., Jamei, M. et al. (2019). VIVD: Virtual in vitro distribution model for the mechanistic prediction of intracellular concentrations of chemicals in in vitro toxicity assays. *Toxicol In Vitro* 58, 42-50. doi:10.1016/j.tiv.2018.12.017



- Gant, T. W., Auerbach, S. S., Von Bergen, M. et al. (2023). Applying genomics in regulatory toxicology: A report of the ECETOC workshop on omics threshold on non-adversity. *Arch Toxicol* 97, 2291-2302. doi:10.1007/s00204-023-03522-3
- Gaulton, A., Hersey, A., Nowotka, M. et al. (2017). The ChEMBL database in 2017. *Nucleic Acids Res* 45, D945-D954. doi:10.1093/nar/gkw1074
- Gumy, C., Chandsawangbhuwana, C., Dzyakanchuk, A. A. et al. (2008). Dibutyltin disrupts glucocorticoid receptor function and impairs glucocorticoid-induced suppression of cytokine production. *PLoS One* 3, e3545. doi:10.1371/journal.pone.0003545
- Harrill, J. A., Everett, L. J., Haggard, D. E. et al. (2021). High-throughput transcriptomics platform for screening environmental chemicals. *Toxicol Sci* 181, 68-89. doi:10.1093/toxsci/kfab009
- Jamei, M., Turner, D., Yang, J. et al. (2009). Population-based mechanistic prediction of oral drug absorption. *AAPS J* 11, 225-237. doi:10.1208/s12248-009-9099-y
- Kamburov, A., Stelzl, U., Lehrach, H. et al. (2012). The Consensus-PathDB interaction database: 2013 update. *Nucleic Acids Res* 41, D793-D800. doi:10.1093/nar/gks1055
- Kamentsky, L., Jones, T. R., Fraser, A. et al. (2011). Improved structure, function and compatibility for cellprofiler: Modular high-throughput image analysis software. *Bioinformatics* 27, 1179-1180. doi:10.1093/bioinformatics/btr095
- Kavlock, R. J., Bahadori, T., Barton-Maclaren, T. S. et al. (2018). Accelerating the pace of chemical risk assessment. *Chem Res Toxicol* 31, 287-290. doi:10.1021/acs.chemrestox.7b00339
- Kawasaki, Y., Umemura, T., Saito, M. et al. (1998). Toxicity study of a rubber antioxidant, 2-mercaptobenzimidazole, by repeated oral administration to rats. *J Toxicol Sci* 23, 53-68. doi:10.2131/jts.23.53
- Klimisch, H. J., Andreae, M. and Tillmann, U. (1997). A systematic approach for evaluating the quality of experimental toxicological and ecotoxicological data. *Regul Toxicol Pharmacol* 25, 1-5. doi:10.1006/rtph.1996.1076
- Krug, A. K., Balmer, N. V., Matt, F. et al. (2013). Evaluation of a human neurite growth assay as specific screen for developmental neurotoxicants. *Arch Toxicol* 87, 2215-2231. doi:10.1007/s00204-013-1072-y
- Lobell, M. and Sivarajah, V. (2003). In silico prediction of aqueous solubility, human plasma protein binding and volume of distribution of compounds from calculated pKa and AlogP98 values. *Mol Divers* 7, 69-87. doi:10.1023/b:modi.00000006562.93049.36
- Love, M. I., Huber, W. and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15, 550. doi:10.1186/s13059-014-0550-8
- Mahalingaiah, P. K., Palenski, T. and Van Vleet, T. R. (2018). An in vitro model of hematotoxicity: Differentiation of bone marrow-derived stem/progenitor cells into hematopoietic lineages and evaluation of lineage-specific hematotoxicity. *Curr Protoc Toxicol* 76, e45. doi:10.1002/cptx.45
- Maranghi, F., De Angelis, S., Tassinari, R. et al. (2013). Reproductive toxicity and thyroid effects in Sprague Dawley rats exposed to low doses of ethylenethiourea. *Food Chem Toxicol* 59, 261-271. doi:10.1016/j.fct.2013.05.048
- Norford, D. C., Meuten, D. J., Cullen, J. M. et al. (1993). Pituitary and thyroid gland lesions induced by 2-mercaptobenzimidazole (2-MBI) inhalation in male fischer-344 rats. *Toxicol Pathol* 21, 456-464. doi:10.1177/019262339302100505
- Noyes, P. D., Friedman, K. P., Browne, P. et al. (2019). Evaluating chemicals for thyroid disruption: Opportunities and challenges with in vitro testing and adverse outcome pathway approaches. *Environ Health Perspect* 127, 95001. doi:10.1289/ehp5297
- NTP – National Toxicology Program (2018). NTP Research Report on National Toxicology Program Approach to Genomic Dose-Response Modeling. NTP RR 5. Research Triangle Park, NC. *National Toxicology Program*, 1-44. doi:10.22427/ntp-rr-5
- Palmen, N. G. M. and Evelo, C. T. A. (1998). Oxidative effects in human erythrocytes caused by some oximes and hydroxylamine. *Arch Toxicol* 72, 270-276. doi:10.1007/s002040050501
- Paul Friedman, K., Watt, E. D., Hornung, M. W. et al. (2016). Tiered high-throughput screening approach to identify thyroperoxidase inhibitors within the ToxCast phase I and II chemical libraries. *Toxicol Sci* 151, 160-180. doi:10.1093/toxsci/kfw034
- Paul Friedman, K., Gagne, M., Loo, L. H. et al. (2020). Utility of in vitro bioactivity as a lower bound estimate of in vivo adverse effect levels and in risk-based prioritization. *Toxicol Sci* 173, 202-225. doi:10.1093/toxsci/kfz201
- Phillips, J. R., Svoboda, D. L., Tandon, A. et al. (2019). BMDExpress 2: Enhanced transcriptomic dose-response analysis workflow. *Bioinformatics* 35, 1780-1782. doi:10.1093/bioinformatics/bty878
- Ramaiahgari, S. C., Auerbach, S. S., Saddler, T. O. et al. (2019). The power of resolution: Contextualized understanding of biological responses to liver injury chemicals using high-throughput transcriptomics and benchmark concentration modeling. *Toxicol Sci* 169, 553-566. doi:10.1093/toxsci/kfz065
- Reardon, A. J. F., Farmahin, R., Williams, A. et al. (2023). From vision toward best practices: Evaluating in vitro transcriptomic points of departure for application in risk assessment using a uniform workflow. *Front Toxicol* 5, 1194895. doi:10.3389/ftox.2023.1194895
- Rodgers, T., Leahy, D. and Rowland, M. (2005). Physiologically based pharmacokinetic modeling 1: Predicting the tissue distribution of moderate-to-strong bases. *J Pharm Sci* 94, 1259-1276. doi:10.1002/jps.20322
- Rodgers, T. and Rowland, M. (2006). Physiologically based pharmacokinetic modelling 2: Predicting the tissue distribution of acids, very weak bases, neutrals and zwitterions. *J Pharm Sci* 95, 1238-1257. doi:10.1002/jps.20502
- Sakuratani, Y., Zhang, H. Q., Nishikawa, S. et al. (2013). Hazard evaluation support system (HESS) for predicting repeated dose toxicity using toxicological categories. *SAR QSAR Environ Res* 24, 351-363. doi:10.1080/1062936x.2013.773375
- Schimming, J. P., ter Braak, B., Niemeijer, M. et al. (2019). System microscopy of stress response pathways in cholestasis research. In M. Vinken (ed.), *Experimental Cholestasis Research*. New York, NY, USA: Springer New York. doi:10.1007/978-1-4939-9420-5_13
- Scholz, D., Pörtl, D., Genewsky, A. et al. (2011). Rapid, complete and large-scale generation of post-mitotic neurons from the human luminescent cell line. *J Neurochem* 119, 957-971. doi:10.1111/j.1471-4159.2011.07255.x
- Stiegler, N. V., Krug, A. K., Matt, F. et al. (2011). Assessment of

- chemical-induced impairment of human neurite outgrowth by multiparametric live cell imaging in high-density cultures. *Toxicol Sci* 121, 73–87. doi:10.1093/toxsci/kfr034
- Thomas, R. S., Wesselkamper, S. C., Wang, N. C. Y. et al. (2013). Temporal concordance between apical and transcriptional points of departure for chemical risk assessment. *Toxicol Sci* 134, 180–194. doi:10.1093/toxsci/kft094
- Thomas, R. S., Bahadori, T., Buckley, T. J. et al. (2019). The next generation blueprint of computational toxicology at the U.S. Environmental Protection Agency. *Toxicol Sci* 169, 317–332. doi:10.1093/toxsci/kfz058
- Tisdell, M. (1985). Chronic toxicity study of rotenone in rats: Final report: Study No. 6115-100. Unpublished study prepared by Hazleton Laboratories America, Inc.
- Tolosa, L., Donato, M. T., Pérez-Cataldo, G. et al. (2012). Upgrading cytochrome P450 activity in HepG2 cells co-transfected with adenoviral vectors for drug hepatotoxicity assessment. *Toxicol In Vitro* 26, 1272–1277. doi:10.1016/j.tiv.2011.11.008
- US EPA (2021a). Estimation Programs Interface Suite™ for Microsoft® Windows V 4.1. United States Environmental Protection Agency, Washington, DC, USA.
- US EPA (2021b). New Approach Methods Work Plan (V2). U.S. Environmental Protection Agency, Washington, DC. EPA/600/X-21/209. <https://www.epa.gov/chemical-research/new-approach-methods-work-plan>
- US FDA (2021). Advancing New Alternative Methodologies at FDA. <https://www.fda.gov/media/144891/download>
- van der Burg, B., van der Linden, S., Man, H.-Y. et al. (2013). A panel of quantitative Calux® reporter gene assays for reliable high-throughput toxicity screening of chemicals and complex mixtures. In P. Steinberg (ed.), *High-Throughput Screening Methods in Toxicity Testing*. John Wiley & Sons, Inc. doi:10.1002/9781118538203.ch28
- van der Stel, W., Carta, G., Eakins, J. et al. (2020). Multiparametric assessment of mitochondrial respiratory inhibition in HepG2 and RPTEC/TERT1 cells using a panel of mitochondrial targeting agrochemicals. *Arch Toxicol* 94, 2707–2729. doi:10.1007/s00204-020-02792-5
- Webster, A. F., Chepelev, N., Gagne, R. et al. (2015). Impact of genomics platform and statistical filtering on transcriptional benchmark doses (BMD) and multiple approaches for selection of chemical point of departure (PoD). *PLoS One* 10, e0136764. doi:10.1371/journal.pone.0136764
- Wieser, M., Stadler, G., Jennings, P. et al. (2008). hTERT alone immortalizes epithelial cells of renal proximal tubules without changing their functional characteristics. *Am J Physiol Renal Physiol* 295, F1365–F1375. doi:10.1152/ajprenal.90405.2008
- Wink, S., Hiemstra, S., Huppelschoten, S. et al. (2014). Quantitative high content imaging of cellular adaptive stress response pathways in toxicity for chemical safety assessment. *Chem Res Toxicol* 27, 338–355. doi:10.1021/tx4004038
- Wink, S., Hiemstra, S., Herpers, B. et al. (2017). High-content imaging-based BAC-GFP toxicity pathway reporters to assess chemical adversity liabilities. *Arch Toxicol* 91, 1367–1383. doi:10.1007/s00204-016-1781-0
- Yamada, T., Kawamura, T., Tsujii, S. et al. (2022). Formation and evaluation of mechanism-based chemical categories for regulatory read-across assessment of repeated-dose toxicity: A case of hemolytic anemia. *Regul Toxicol Pharmacol* 136, 105275. doi:10.1016/j.yrtph.2022.105275
- Yeakley, J. M., Shepard, P. J., Goyena, D. E. et al. (2017). A trichostatin A expression signature identified by TempO-Seq targeted whole transcriptome profiling. *PLoS One* 12, e0178302. doi:10.1371/journal.pone.0178302
- Zhang, L.-F., Liu, L.-S., Chu, X.-M. et al. (2014). Combined effects of a high-fat diet and chronic valproic acid treatment on hepatic steatosis and hepatotoxicity in rats. *Acta Pharmacol Sin* 35, 363–372. doi:10.1038/aps.2013.135

Conflict of interest

B. M. van Vugt-Lussenburg was employed by BioDetection Systems, Amsterdam, The Netherlands. P. Walker was employed by Cypotex Discovery Ltd UK, Macclesfield, United Kingdom. C. Fisher and B. Islam were employed by Certara – Simcyp Division, Sheffield, United Kingdom. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Data availability

The experimental data presented in this study are available in supplementary files: “Zobl_Suppl.xlsx”¹ includes seven tables detailing the *in vivo* and *in vitro* data underlying this study; “Zobl_Suppl2.pdf”² includes supplementary materials with details regarding the characterization of test compounds, and “Zobl_Suppl3.7z”³ includes transcriptomics data for all test compounds.

Acknowledgements

This work has received funding from the European Union’s Horizon 2020 research and innovation program under grant agreements No. 681002 (EU-ToxRisk) and 964537 (RISK-HUNT3R) as well as from the Horizon Europe research and innovation program under grant agreement No. 101057014 (PARC). Further, it was supported by the German Federal Ministry of Education and Research (BMBF) (NeuroTool, InnosystoxMoving). The University of Konstanz would like to thank Viktoria Magel for cell culture support in the screening process. The symbolic pathway depicted as part of Figure 1 was created by Dmitriy Sachno with BioRender.com.

³ doi:10.14573/altex.2309081s3