



"I cannot say whether things will get better if we change; what I can say is they must change if they are to get better."

Georg Christoph Lichtenberg (1742-1799)

"Systems thinking is a discipline for seeing wholes. It is a framework for seeing interrelationships rather than things, for seeing 'patterns of change' rather than static 'snapshots'."

Peter M. Senge (1947-), MIT

Food for Thought ...

3S – Systematic, Systemic, and Systems Biology and Toxicology

Lena Smirnova¹, Nicole Kleinstreuer², Raffaella Corvi³, Andre Levchenko⁴, Suzanne C. Fitzpatrick⁵ and Thomas Hartung^{1,6}

¹Johns Hopkins University, Bloomberg School of Public Health, Center for Alternatives to Animal Testing (CAAT), Baltimore, MD, USA; ²NIH/NIEHS/DNTP/NICEATM, RTP, NC, USA; ³European Commission, Joint Research Centre (JRC), EU Reference Laboratory for Alternatives to Animal Testing (EURL ECVAM), Ispra, (VA), Italy; ⁴Yale Systems Biology Institute and Biomedical Engineering Department, Yale University, New Haven, CT, USA; ⁵Food and Drug Administration (FDA), Center for Food Safety and Applied Nutrition, College Park, MD, USA; ⁶CAAT-Europe, University of Konstanz, Konstanz, Germany

Summary

A biological system is more than the sum of its parts – it accomplishes many functions via synergy. Deconstructing the system down to the molecular mechanism level necessitates the complement of reconstructing functions on all levels, i.e., in our conceptualization of biology and its perturbations, our experimental models and computer modelling.

Toxicology contains the somewhat arbitrary subclass "systemic toxicities"; however, there is no relevant toxic insult or general disease that is not systemic. At least inflammation and repair are involved that require coordinated signaling mechanisms across the organism. However, the more body components involved, the greater the challenge to recapitulate such toxicities using non-animal models. Here, the shortcomings of current systemic testing and the development of alternative approaches are summarized.

We argue that we need a systematic approach to integrating existing knowledge as exemplified by systematic reviews and other evidence-based approaches. Such knowledge can guide us in modelling these systems using bioengineering and virtual computer models, i.e., via systems biology or systems toxicology approaches. Experimental multi-organ-on-chip and microphysiological systems (MPS) provide a more physiological view of the organism, facilitating more comprehensive coverage of systemic toxicities, i.e., the perturbation on organism level, without using substitute organisms (animals). The next challenge is to establish disease models, i.e., micropathophysiological systems (MPPS), to expand their utility to encompass biomedicine. Combining computational and experimental systems approaches and the challenges of validating them are discussed. The suggested 3S approach promises to leverage 21st century technology and systematic thinking to achieve a paradigm change in studying systemic effects.

Keywords: evidence-based toxicology, systems biology, repeated-dose toxicity, carcinogenicity, DART

1 Introduction

Systematic, systemic, and systems sound very much alike, but they represent three different approaches in the life sciences.

We will argue here that synergy between them is necessary to achieve meaningful understanding in biomedicine. Biology stands for the unperturbed "physiological" behavior of our model systems. Toxicology is certainly one of the more applied

Disclaimer: The views presented are those of the individual authors and do not necessarily reflect those of all authors or those of their institutions or official federal government policy. This article does not necessarily reflect the policy of the National Toxicology Program, National Institutes of Health, or the Food and Drug Administration.

Received April 5, 2018;
doi:10.14573/altex.1804051



This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 International license (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is appropriately cited.



sciences studying the perturbation of model systems (patho-physiology); ultimately, all experimental medical research links to this. Here, we focus primarily upon examples from toxicology, which is the authors' primary area of expertise, and the restriction to this area in the title appears prudent.

Systematic is a term most commonly used in the context of systematic reviews, i.e., evidence-based approaches that aim for a comprehensive, objective and transparent use of information. Born in the clinical and health care sciences, these approaches have gained significant traction in toxicology¹ but have not had major impacts on other pre-clinical and biological areas. We will argue that this represents an omission and an opportunity, as the respective tools for evidence evaluation (quality scoring, risk-of-bias analysis, etc.) and integration (meta-analysis, combination of information streams, etc.) are widely applicable across scientific disciplines. The resulting condensation of information and mapping of knowledge deficits as well as the cross-talk with quality assurance, Good Practices, and reporting standards, yield valuable lessons on how the systematic evaluation of available scientific knowledge can accelerate the organization of vast, and rapidly expanding, knowledge generation.

Systemic views are primarily organism-level views on problems (the big-picture view), the opposite of studying smaller and smaller elements of the machinery. However, it is also thinking in terms of functionalities. Cell culture is starting to embrace this with the advent of complex co-cultures with multiplexed endpoints (Kleinstreuer et al., 2014) and organotypic cultures reproducing organ architecture and functionalities (microphysiological systems, MPS) (Marx et al., 2016), now even moving to multi-organ models of a human-on-chip / body-on-chip (Skardal et al., 2016; Watson et al., 2017). The concomitant emerging availability of human stem cells that can be used to produce high-quality organoids further adds to this revolutionary change (Suter-Dick et al., 2015), as shown recently for the BrainSphere organoid model (Pamies et al., 2017a, 2018a). Functional thinking can also be applied to cellular biology when considering toxic impact, for example, repair, recovery and resilience (Smirnova et al., 2015). We are returning to seeing the forest, not just individual trees.

Systems biology and, more recently, toxicology (Hartung et al., 2012, 2017a) aim to study systems behavior: "*Systems biology begins to recognize the limitations of the reductionist approach to biology*" (Joyner and Pedersen, 2011). In its detailed definition (Ferrario et al., 2014), it is based on a comprehensive study of our knowledge on these systems, which is translated into computer models, allowing virtual experiments/simulations that can be compared to experimental results. Systems approaches require sufficient biological and physiological detail about the relevant molecular pathways, associated cellular behaviors, and complex tissue-level interactions, as well as computational models that adequately represent biological complexity while offsetting mathematical complexity. Bernhard Ø. Palsson wrote in his book *Systems Biology: Constraint-Based Reconstruction and Analysis*, "*The chemical interac-*

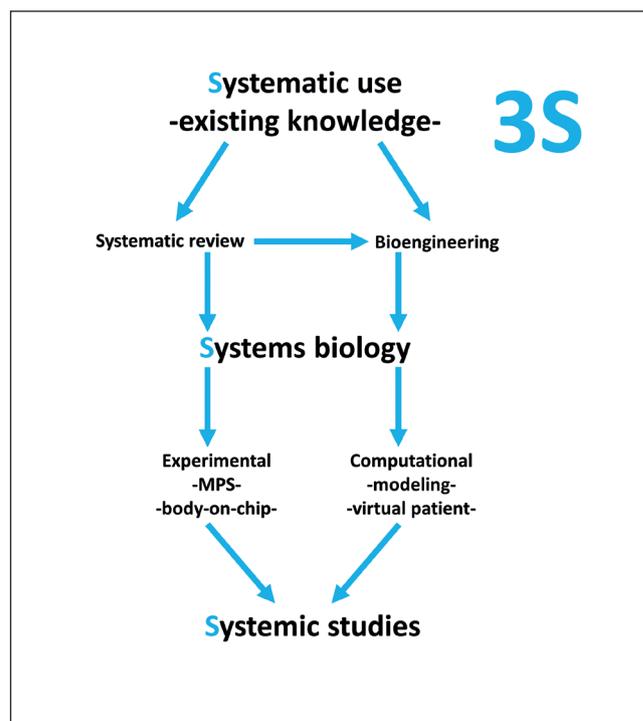


Fig. 1: The 3S approach to study systemic phenomena

tions between many of these molecules are known, giving rise to genome-scale reconstructed biochemical reaction networks underlying cellular functions". So, to some extent, the systems toxicology approach is systematic and systemic in view, but it brings in the additional aspects of knowledge organization using dynamic models of physiology.

Figure 1 shows how these different components come together. This paper suggests that the traditional 3Rs approach, which has served us well to replace a substantial part of acute and topical toxicities, might need approaches along the 3S for systemic toxicity testing replacement. It suggests that systematic organization of existing knowledge be combined with experimental and computational systems approaches to model the complexity of (patho-)physiology.

2 Systematic biology and toxicology

Perhaps a better term would be "systematic review" of biology and toxicology. Similar to the term evidence, the concept of being systematic sounds like it must be a given for any scientific endeavor. Unfortunately, it is not. Most of us are drowning in a flood of information. The seemingly straightforward request of evidence-based medicine (EBM) to assess all available evidence quickly reaches limits of practicality. A systematic evaluation of the literature often returns (tens of) thousands of articles. Only very important questions, which must at the same

¹ <http://www.ebtox.org>

time be very precise and very focused, warrant comprehensive efforts to analyze them. It is still worth the effort – as typically the result is strong evidence that is difficult to refute.

Earlier work in this area (Hoffmann and Hartung, 2006) led to the creation of the Evidence-based Toxicology Collaboration (EBTC) in 2011¹. Developments have been documented (Griesinger et al., 2008; Stephens et al., 2013) and have gained acceptance (Aiassa et al., 2015; Stephens et al., 2016; Ågerstrand and Beronius, 2016). The field is developing very rapidly (Morgan et al., 2016; Mandrioli and Silbergeld, 2016). The fundamentals and advantages of evidence-based approaches were previously detailed in a dedicated article (Hartung, 2009a) that appeared earlier in this series (Hartung, 2017a), and are not repeated here. Noteworthy, the call for a systematic review of animal testing methods is getting louder and louder (Basketter et al., 2012; Leist et al., 2014; Pound and Brakken, 2014); the work of SYRCLE, the SYstematic Review Centre for Laboratory Animal Experimentation² and CAMARADES³ (Collaborative Approach to Meta-Analysis and Review of Animal Data from Experimental Studies) is especially noteworthy. Here, we focus on two main points, one concerning opportunities for application in biology and other non-toxicology biomedical sciences, and the other framing the utility of systematic review in the context of systemic toxicities and systems toxicology.

Note that evidence-based approaches have a lot to offer also across diverse areas of biomedicine. Mapping what we know and what we do not, helps a field to focus research and resources, not only in areas where safety is at stake. In the clinical arena, EBM was the catalyst for many quality initiatives. Nobody wants to do research that is excluded from deriving authoritative conclusions by peers for quality (of reporting) reasons. A significant portion of irreproducible science could be avoided by using evidence-based approaches (Hartung, 2013; Freedman et al., 2015).

It should also be commented that in the context of systemic toxicities, we first of all need a systematic review of the traditional test methods, the information that they provide, and the decision contexts in which they are used. This was the unanimous recommendation of the roadmap exercise on how to overcome animal-based testing for systemic toxicology (Basketter et al., 2012; Leist et al., 2014). Systematic review was also suggested as a necessary element for a mechanistic validation (Hartung et al., 2013); this represents a key opportunity for the validation of both adverse outcome pathways (AOP) and mechanistic experimental models such as MPS. Lastly, systems toxicology should be based on a comprehensive analysis of biological systems characteristics, again calling for systematic literature analysis.

3 Systemic biology and toxicology

Systemic biology is not a common term – probably “physiology” covers it best, though much of physiology is studied in isolated organs. With the flourishing of molecular and cellu-

lar biology and biochemistry, systemic biology has been less prominent over the last few decades. However, the need to understand molecular and mechanistic findings in the context of an intact organism is obvious. This is one of the arguments for whole animal experimentation that are more difficult to refute. In fact, it is the use of genetically modified animals in academic research that is driving the steady increase of animal use statistics after three decades of decline (Daneshian et al., 2015).

Regulatory assessment of the complex endpoints of repeated-dose toxicity (RDT), carcinogenicity, and developmental and reproductive toxicity (DART), which are often grouped under systemic toxicities, still relies heavily on animal testing. Arguably, there is hardly any toxicity in nature that is not systemic as even topical effects such as skin sensitization include inflammatory components involving leukocyte infiltration and other acute effects, e.g., lethality, involve many parts of the organism. But, the aforementioned areas of toxicology represent the best examples of systemic toxicology, in which new approaches are needed but implementation is not straightforward.

The following section first addresses the limitations of current systemic toxicity testing and then reviews the alternative approaches that were developed in these areas of systemic toxicity in the last decades to waive testing or reduce the number of animals used.

The shortcomings of the current paradigm have been discussed earlier (Hartung, 2008a, 2013; Basketter et al., 2012; Paparella et al., 2013, 2017); some studies that cast doubt as to their performance are summarized in Table 1, using the more factual references, though the balance between opinion and evidence is difficult in the absence of systematic reviews (Hartung, 2017b). However, they stress the need for the strategic development of a new approach (Busquet and Hartung, 2017), especially for the systemic toxicities.

Alternative approaches range from the individual test methods (e.g., the cell transformation assay for carcinogenicity and the zebrafish and embryonic stem cell embryotoxicity tests for reproductive toxicity) to animal reduction approaches such as the International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH) strategy for carcinogenicity of pharmaceuticals and the extended-one-generation reproductive toxicity study. Currently, these areas are being revitalized owing to the broad recognition of the shortcomings of current *in vivo* testing requirements and the current regulatory environment (e.g., the European REACH and Cosmetic Regulation, the US amendment to the Toxic Substance Control Act (TSCA), i.e., the Lautenberg Chemical Safety for the 21st Century Act). More recent developments aimed at a more human-relevant chemical assessment, which rely on the integration of different sources of information, are also described.

The assessment of repeated-dose systemic toxicity, carcinogenicity and developmental and reproductive toxicity represent essential components of the safety assessment of all types of substances, being among the endpoints of highest concern. As

² <https://bit.ly/2q8yXXh>

³ <http://www.dcn.ed.ac.uk/camarades/#carousel-example-generic>


Tab. 1: Worrisome analyses as to the relevance of traditional systemic toxicity studies

Repeated-Dose Toxicity (RDT)	Developmental and reproductive toxicity (DART)	Cancer bioassay
Interspecies concordance of mice with rats (37 chemicals): 57%-89% (average 75%) in short-term and 65%-89% (average 80%) in long-term studies (Wang and Gray, 2015)	No relevant contribution to regulatory decision-making by second generation testing (Janer et al., 2007; Martin et al., 2009a)	While 53% of all chemicals test positive, age-adjusted cancer rates did not increase over the last century (Jemal et al., 2009)
Mouse-to-rat organ prediction (37 chemicals) in long-term studies with an average of 55%, in short-term studies with an average of 45%. For rat-to-mouse, the averages were 27% and 49%, respectively (Wang and Gray, 2015)	254 chemicals in ToxRefDB tested in both multi-generation and 2-year chronic studies, and 207 chemicals tested in both multigeneration and 90-day subchronic studies (Martin et al., 2009b); with an assessment factor of 10, the hazard of reproductive toxicity might be covered for 99.8% of substances	Exposure to mutagens does not correlate with oncomutations in people (Thilly, 2003)
Species concordance (310 chemicals) for non-neoplastic pathology between mouse and rat was 68% (Wang and Gray, 2015)	No experience for industrial chemicals: < 25 two-generation-studies and < 100 one-generation studies in EU and US in 30 years (Bremer et al., 2007)	Protocol has poorly defined endpoints and a high level of uncontrolled variation; could be optimized to include proper randomization, blinding, better necroscopy work, and adequate statistics (Freedman and Zeisel, 1988).
Inter-species differences mouse vs. rat (95 th percentile) of 8.3 for RDT (Bokkers and Slob, 2007)	Large number of individual skeletal variations (sometimes > 80%) even in control animals (Daston and Seed, 2007)	Most recent test guidelines (OECD, 2009) still do not make randomization and blinding mandatory, and statistics do not control for multiple testing, although about 60 endpoints are assessed (Basketter et al., 2012).
Low correlation between 28-day and 90-day NOAEL for 773 chemicals (Luechtefeld et al., 2016b, Fig.4)	Of those agents thought not to be teratogenic in man, only 28% are negative in all species tested (Brown and Fabro, 1983)	Not standardized for animal strains (" <i>young healthy adult animals of commonly used laboratory strains should be employed</i> ") (Basketter et al., 2012)
A limited set of only six targets consisting of liver, kidney, clinical chemistry, body weight, clinical symptoms and hematology within a study gives a probability of 86% to detect the LOEL (Batke et al., 2013)	Of 1223 definite, probable and possible animal teratogens, fewer than 2.3% were linked to human birth defects (Bailey et al., 2005)	Problems with standardization of strains that hamper the use of historical control groups (Haseman et al., 1997): the most commonly used strains showed large weight gain and changes in some tumor incidences that resulted in reduced survival over just one decade (attributed to intentional or inadvertent selection of breeding stocks with faster growth and easier reproduction)
	Not robust with about 25% equivocal studies (Bailey et al., 2005)	Analysis of 1,872 individual species/gender group tests in the US National Toxicology Program (NTP) showed that 243 of these tests resulted in "equivocal evidence" or were judged as "inadequate studies" (Seidle, 2006)
	74 industrial chemicals tested in New Chemicals Database: 34 showed effects on offspring, but only 2 chemicals were classified as developmental toxicants (Bremer and Hartung, 2004)	Questionable two-species paradigm as rats are more sensitive, and regulatory action is rarely taken on the basis of results in mice (Van Oosterhout et al., 1997; van Ravenzwaay, 2010)
	55% of positives in screening studies not in multi-generation studies (Bremer and Hartung, 2004)	Concordance of 57% comparing 121 replicate rodent carcinogenicity assays (Gottmann et al., 2001)
	Group size limits statistical power (Hotchkiss, 2008)	The apparent correlation between potency of carcinogens in mice and rats is largely an artifact (Bernstein et al., 1985).
	61% inter-species correlation (Hurt, 2003; Bailey et al., 2005)	Concordance of 57% between mouse and rat bioassays (Gray et al., 1995).



Repeated-Dose Toxicity (RDT)	Developmental and reproductive toxicity (DART)	Cancer bioassay
	Given 2.5% true reproductive toxicants and 60% inter-species correlation, testing with two species will find 84% of the toxic but label 64% of the negatives falsely (Hartung, 2009b)	Less than 50% probability for known carcinogens that induce tumors in one species in a certain organ to also induce tumors in another species the same organ comparing rats, mice, and hamsters, as well as humans (Gold et al., 1991, 1998).
	Of 38 human teratogens, the following percentages tested positive in other species: mouse 85%, rat 80%, rabbit 60%, hamster 45%, monkey 30%, two or more species 80%, any one species 97% (Brown and Fabro, 1983)	Doses are hundreds to thousands of times higher than normal exposures and might be carcinogenic simply because they overwhelm detoxification pathways (Schmidt, 2002)
	Of 165 human non-teratogens, the following percentages tested negative in other species: mouse 35%, rat 50%, rabbit 70%, hamster 35%, monkey 80%, two or more species 50%, all species 28% (Brown and Fabro, 1983)	69% predictivity of human carcinogenicity for the two-species cancer bioassay (Pritchard et al., 2003)
	Reproductive toxicity within 10-fold of maternal repeated-dose toxicity for 99.8% of 461 chemicals (Martin et al., 2009b)	In 58% of cases considered by the EPA, the positive cancer bioassay was insufficient for assigning human carcinogenicity (Knight et al., 2006a,b)
		Cancer bioassays in nonhuman primates on 37 compounds were "... <i>inconclusive in many cases</i> " but carcinogenicity was shown unequivocally for four of them (Takayama et al., 2008)
		About 50% of all chemicals tested positive in the cancer bioassay test (Basketter et al., 2012), and 53% of 301 chemicals tested by the NTP were positive, with 40% of these positives classified as non-genotoxic (Ashby and Tennant, 1991)
		An early analysis of 20 putative human non-carcinogens found 19 false-positives, suggesting only 5% specificity (Ennever et al., 1987).
		Only one in ten positive compounds is truly carcinogenic (Rall, 2000)
		Not all human carcinogens are found: Diphenyl-hydantoin (phenytoin) (Anisimov et al., 2005); the combination of aspirin/phenacetin/ caffeine (Ennever and Lave, 2003); asbestos, nickel, benzidine-like compounds (Johnson, 2001); no cigarette smoke-induced lung cancer, no rodent leukemia induced by benzene, and no genetic point mutations induced by arsenic (Silbergeld, 2004).
		Estimate 70% sensitivity as well as specificity, assuming 10% real human carcinogens (Lave et al., 1988)
		Of 167 chemicals that caused neoplastic lesions in rat or mouse chronic/cancer studies, 35% caused neoplastic lesions in both rat and mouse (Martin et al., 2009a)
		Increasing the number of animals per group from 50 to 200 would result in statistically significant ($p < 0.01$) dose-responses for 92% of substances tested (Gaylor, 2005)



such, their assessment still relies mainly on animal tests. Progress toward replacing this paradigm is summarized in the following sections.

3.1 Repeated-dose systemic toxicity

Repeated chemical treatment, usually on a daily basis, from several days to life-long exposure, is key to the hazard assessment of substances as it covers toxicokinetic aspects, i.e., adsorption, distribution, metabolism and excretion (ADME), as well as toxicodynamics with the potential of all organ effects and interactions. The present testing schemes are based on rodent or non-rodent studies performed for 28 days (subacute toxicity), 90 days (subchronic toxicity), or 26-102 weeks (chronic toxicity). These tests typically form the basis for identifying hazards and their characterization, especially no-effect-levels (NOEL). This approach rests upon the key assumption that the animal models are representative of human ADME and effects. In fact, the enormous differences in ADME represented a key reason for drug attrition two decades ago (attrition has dropped from 40-60% to nowadays 10% (Kennedy, 1997; Kubinyi, 2003; Kola and Landis, 2004)), as the development of a portfolio of *in vitro* and *in silico* tools has drastically improved the situation as reviewed earlier in this series (Tsaïoun et al., 2016). The toolbox is neither perfect nor complete but, as discussed in the context of developing a roadmap for improvement (Basketter et al., 2012; Leist et al., 2014), there was general consensus among the experts involved that the missing elements are feasible and in reach. For example, epithelial barrier models (Gordon et al., 2015) as input into physiology-based (pharmaco-/toxicokinetic (PBPK) modelling were identified as a key opportunity for modelling RDT and were recently the subject of the Contemporary Concepts in Toxicology 2018 meeting “Building a Better Epithelium”.

The Adler et al. (2011) report already compiled the many partial solutions to RDT. The problem is how to integrate these elements into a testing strategy that provides predictivity of human toxicity that is equivalent or greater than that of an animal study. This is a difficult question to answer, as in most cases we do not actually know how predictive the animal studies are due to the absence of human data. A notable exception is in the area of topical toxicities such as skin sensitization, where the predictive performance of the animal studies against human data has been shown to be essentially equivalent to the reproducibility of the animal data (Kleinstreuer et al., 2018).

We can start by asking how reproducible they are and how well different laboratory animal species predict each other. An important analysis conducted by Wang and Gray (2015) gives us an idea: very little. They compared earlier RDT findings with the non-cancer pathologies observed in cancer bioassays in rats and mice of both genders run by the National Toxicology Program for 37 substances. They concluded: “Overall, there is considerable uncertainty in predicting the site of toxic lesions in different species exposed to the same chemical and from short-term to long-term tests of the same chemical.” Although this study was done for only 37 chemicals, it gives us a hint that there is no reason to assume that the predictivity of rodent data for humans will be any better. For a larger scale comparison, the key obsta-

cle is the lack of harmonized ontologies and reporting formats for RDT (Hardy et al., 2012a,b; Sanz et al., 2017). Very often it is unclear whether effects for certain organs or systemic effects were not reported because (a) they were assessed but not found and not reported as negative data; (b) there were already other organ toxicities at lower doses and, thus, the data on remaining organs was omitted or not assessed, or (c) only one organ was the focus of the study and the remaining and/or systemic effects were out of scope for the given study. Therefore, the standardized curation of databases with detailed organ effects is a resource-intensive problem, and there are currently none that facilitate widespread reproducibility assessments. Independent of the specific site of toxic manifestations, however, it is relatively easy to compare NOELs across studies. Using our machine-readable database from the REACH registration process (Luechtefeld et al., 2016a), such comparisons between 28- and 90-day studies showed strong discrepancies (Luechtefeld et al., 2016b). A systematic evaluation of RDT studies will enable further analysis of the current testing paradigm.

Given these problems, it will be very difficult to model such findings with a test strategy (Chen et al., 2014). Our t^4 workshop on *Adversity in vitro* (report in preparation) in the context of the Human Toxome Project (Bouhifd et al., 2015), took a different approach: Based on the observation that the majority of chemicals are quite promiscuous, i.e., start perturbing the same cellular pathways in a relatively narrow concentration range, it appears feasible to define *in vitro* benchmark doses at which adversity starts using a set of complementary cell-based assays (Judson et al., 2016). Quantitative *in vitro* to *in vivo* extrapolation based on exposure data (plus some safety factors) should allow definition of exposures necessary to reach such tissue concentrations. Without necessitating a prediction of which organs would be affected, a safe use range would be established. In fact, the current risk assessment paradigm also makes little use of which organ exhibits toxic effects first but relies upon the most sensitive endpoint to define a benchmark / no-effect dose. Obviously, this does not work for substances whose molecular initiating events (MIE) are not reflected in the cell test battery to derive benchmark doses or NOELs. This means that over time this should be complemented with specific assays for those substances whose effects may be missed with this approach. Read-across strategies could add safety measures to such an approach, i.e., besides defining the safe dose, read-across and other *in silico* tools could provide alerts for where to add additional safety factors. In cases where human exposure is not sufficiently below doses that can reach critical tissue concentrations, it will be necessary to follow a more investigative toxicological approach, i.e., a mechanistic evaluation addressing the human relevance of the findings.

Biological models for different organs, e.g., liver, kidney, lung or brain, have been established and new culture techniques, especially in form of 3D organoids and MPS, are expected to solve present *in vitro* testing issues concerning long-term culturing, absence of relevant immune cells (Hengstler et al., 2012) and availability of fully mature cell phenotypes. Stem cells, especially induced pluripotent stem cells (iPSC), are a major source of tissue and cell models not available otherwise. Therefore, re-

search on the generation of 2D cultures and 3D tissues from stem cells is of high importance. The formalization of our mechanistic knowledge via adverse outcome pathways (AOP) (Leist et al., 2017) further helps to assess whether these models are relevant. New prospects come from systems approaches, where human complexity is either modelled experimentally or virtually, as discussed below. The European Commission-funded Horizon 2020 consortium EU-ToxRisk was in fact set up to integrate advances in cell biology, omics technologies, systems biology and computational modelling to define the complex chains of events that link chemical exposure to toxic outcome in the areas of RDT, developmental and reproductive toxicity (Daneshian et al., 2016). The vision of EU-ToxRisk, which builds on the activities started by the SEURAT-1 EU framework project, is to progress towards an animal-free toxicological assessment based on human cell responses and a comprehensive mechanistic understanding of cause-consequence relationships of chemical adverse effects⁴.

3.2 Carcinogenicity

Substances are defined as carcinogenic if after inhalation, ingestion, dermal application or injection they induce (malignant) tumors, increase their incidence or malignancy, or shorten the

time to tumor occurrence. It is generally accepted that carcinogenesis is a multi-hit/multi-step process from the transition of normal cells into cancer cells via a sequence of stages and complex biological interactions, strongly influenced by factors such as genetics, age, diet, environment and hormonal balance (Adler et al., 2011). Although attributing observed cancer rates to individual specific causes remains a challenge, the fraction of all cancers currently attributed to exposure to carcinogenic pollutants is estimated to range from less than 1% to 10-15% to as high as 19% (Kessler, 2014; Colditz and Wei, 2012; Anand et al., 2008; President's Cancer Panel, 2010; GBD 2013 Risk Factors Collaborators, 2013).

For nearly half a century, the 2-year rodent cancer bioassay was widely considered the “gold standard” for determining the carcinogenic potential of a chemical and OECD Test Guidelines (TG) exist since 1981 (Madia et al., 2016). Its adequacy to predict cancer risk in humans, however, is the subject of considerable debate (Gottmann et al., 2001; Alden et al., 2011; Knight et al., 2006a,b; Paules et al., 2011) (Tab. 1). Recently, Paparella and colleagues (2017) conducted a systematic analysis of challenges and uncertainties associated with the cancer bioassay. Notably, extrapolating from rodents to humans and quantitative

Tab. 2: Milestones on the road towards a new approach to carcinogenicity testing

Date	Event	Who was involved
1995	Joint proposal for a new OECD TG for the <i>in vitro</i> Syrian Hamster Embryo (SHE) Cell Transformation Assay (CTA)	USA and France
1998	Workshop on “Cell transformation assays as predictors of carcinogenicity”	ECVAM
2006	Workshop on “How to reduce false positive results when undertaking <i>in vitro</i> genotoxicity testing and thus avoid unnecessary follow-up animal tests”	EURL ECVAM
2006-2011	EU-6 Framework Project CarcinoGENOMICS	DG RTD / EU Consortium
2011	ESAC Opinion on prevalidation of <i>in vitro</i> Syrian Hamster Embryo (SHE) Cell Transformation Assay	EURL ECVAM
2012	ESAC opinion on validation of <i>in vitro</i> Bhas42 Cell Transformation Assay	JaCVAM /EURL ECVAM
2009	Acceptance of transgenic models as alternative to bioassay in second species	ICH
2013	ICH Regulatory Notice Document announcing the evaluation of an alternative approach to the 2-year rat carcinogenicity test	ICH and Drug Regulatory Authorities
2015	Starting activity on IATA for non-genotoxic carcinogens	OECD
2015	Adoption of Guidance Document No. 214 on the <i>in vitro</i> Syrian Hamster Embryo (SHE) Cell Transformation Assay	OECD
2016	Adoption of Guidance Document No. 231 on the <i>in vitro</i> Bhas42 Cell Transformation Assay	OECD
2016	Inclusion of characteristics of carcinogens in systematic reviews for Monograph program	IARC
2017	Initiation of the project on predicting carcinogenicity of agrochemicals	EPAA

Abbreviations: DG RTD, EU Directorate General Research and Technological Development; ECVAM, European Centre for the Validation of Alternative Methods; EPAA, European Partnership for Alternative Approaches to Animal Testing; EURL, European Reference Laboratory; IARC, International Agency for Research on Cancer; ICH, International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use; JaCVAM, Japanese Center for the Validation of Alternative Methods; OECD, Organisation for Economic Co-operation and Development

⁴ <http://www.eu-toxrisk.eu>



risk estimation has limited accuracy (Knight et al., 2006b; Paparella et al., 2017; Paules et al., 2011). Moreover, the rodent bioassay, as originally designed, does not take into account windows of susceptibility over the life-time, and so may not have adequate sensitivity to detect agents, such as endocrine active chemicals, that alter susceptibility to tumors (Birnbaum et al., 2003). Furthermore, these studies are very time- and resource-consuming, taking up to three years to completion, and the high animal burden has raised ethical concerns.

From a regulatory perspective, the gradual recognition of non-genotoxic mechanisms of carcinogenesis (that do not involve direct alterations in DNA) has complicated the established relationship between genotoxicity and carcinogenicity and has challenged the conventional interpretation of rodent carcinogenicity results in terms of relevance to human cancer (Hengstler et al., 1999; Waters, 2016). Because of the default assumption in regulatory decision-making regarding the presumed linearity of the dose-response curve for genotoxic carcinogens, the classification of carcinogens as genotoxic or non-genotoxic became an essential but highly controversial component of cancer risk assessment.

The area of carcinogenicity has been very quiet for decades, but in recent years it has been revitalized due to broad recognition of the shortcomings of current regulatory *in vivo* testing requirements, and the awareness of information gaps in legislation that limit or ban the use of animals (e.g., European REACH Regulation (EC) No. 1907/2006 and Cosmetic Regulation (EC) No 1223/2009).

Table 2 shows some steps on the road to replacing the traditional paradigm, some of them are detailed in the following paragraphs.

Genotoxicity assays

Beginning in the late 1960s, highly predictive short-term genotoxicity assays were initially developed to screen for carcinogens. This led to a variety of well-established *in vitro* assays and, since the 1980s, to their respective OECD TGs that have been used successfully to predict genotoxicity, label chemical substances and inform cancer risk assessment. However, these tests are not at present considered to fully replace animal tests currently used to evaluate the safety of substances (Adler et al., 2011). In the last decade, several activities have been carried out worldwide with the aim of optimizing strategies for genotoxicity testing, both with respect to the basic *in vitro* testing battery and to *in vivo* follow-up tests. This was motivated by the scientific progress and significant experience of 40 years of regulatory toxicology testing in this area.

One of the major gaps identified was the need to ensure that *in vitro* tests do not generate a high number of false positive results, which trigger unnecessary *in vivo* follow-up studies, hence generating undesirable implications for animal welfare (Kirkland et al., 2005). The recommendations from a workshop organized by ECVAM (Kirkland et al., 2007) and from an EURL ECVAM strategy paper (EURL ECVAM, 2013a) on how to reduce genotoxicity testing in animals have contributed to several international initiatives aiming to improve the existing *in vitro* genotoxicity tests and strategy and to identify and evaluate new

test systems with improved specificity, while maintaining appropriate sensitivity. The outcome of this work led to the revision of OECD TGs for genotoxicity.

Meanwhile, the *in vitro* micronucleus test, which was the first test to be evaluated by ECVAM through retrospective validation (Corvi et al., 2008), is acquiring an increasingly prominent role in the genotoxicity strategy. It has in fact been proposed as the assay to be used in a two-test battery together with the Ames test (Kirkland et al., 2011; Corvi and Madia, 2017). Further *in vitro* methods are being developed and validated, especially aiming at a full replacement, as in the case of genotoxicity assays in 3D human reconstructed skin models, and for a better understanding of modes of action (MoA) using toxicogenomics-based tests and biomarker assays (Corvi and Madia, 2017).

Transgenic mouse models

Transgenic mouse model tests are possible alternatives to the classical two-year cancer bioassay owing to their enhanced sensitivity as predictors of carcinogenic risk to humans (Tenant et al., 1999). In fact, these models have a reduced tumor latency period (6-9 months) to chemically-induced tumors and may result in a significant reduction in the use of experimental animals (20-25 animals/sex/treatment group) (Marx, 2003). A study coordinated by ILSI/HESI (ILSI HESI ACT, 2001; MacDonald et al., 2004) led to the initial acceptance by pharmaceutical regulatory agencies of three primary models: p53^{+/-}, Tg.AC and rash2 model, to be used *in lieu* of a second species full carcinogenicity bioassay (ICH, 2009).

Cell transformation assays

In vitro cell transformation assays (CTA) for the detection of potential carcinogens have been in use for about four decades. Transformation involves several events in the cascade potentially leading to carcinogenesis and is defined as the induction of phenotypic alterations in cultured cells that are characteristic of tumorigenic cells (LeBoeuf et al., 1999; Combes et al., 1999). Despite the long experience in the use of CTAs, the intense and prolonged activities at the OECD from 1997 to 2016, and the performance of ECVAM and JaCVAM validation studies (EURL ECVAM, 2012, 2013b), the assays were adopted as OECD Guidance Documents (OECD, 2015, 2016), but they have so far not been adopted as OECD TGs in their own right. Among the obstacles to the development of an OECD TG for the Syrian Hamster Embryo (SHE) Cell Transformation Assay (SHE CTA) was the lack of a coordinated full validation. The combination of a detailed review paper (DRP) and a prospective limited validation study triggered the need for additional analyses by the OECD expert group (OECD, 2007; Corvi et al., 2012). This also demonstrates that a DRP cannot be considered equivalent to a retrospective validation. Moreover, with the lack of an Integrated Approach to Testing and Assessment (IATA) or alternative testing strategy available for carcinogenicity and since there was common agreement that the assay should not be used as a stand-alone, no strategy was available on how to apply it for regulatory decision-making. This dilemma, "What comes first: the chicken or the egg, the test method or the testing strategy (or IATA)?" raises the question whether in the future

the OECD should accept only new methods associated to a well-defined testing strategy or an IATA. A better characterization of the performance of the CTA to detect non-genotoxic carcinogens was considered important because the data collected in the DRP were biased towards genotoxic carcinogens, which reflects data available in the public domain. Another recurring concern was that the mechanistic basis of cell transformation and the link to tumorigenesis are not yet completely elucidated, which hampers interpretation of the findings from such an assay.

During the course of the OECD CTA activities, the regulatory framework in Europe changed considerably with the ban on animal testing for cosmetics (Hartung, 2008c) coming into force and the REACH evaluation of industrial chemicals commencing (Hartung, 2010b). This has put a huge burden on industry, which is limited in the use of *in vivo* tests to confirm results from *in vitro* tests, and on regulators, who have to assess carcinogenicity potential without *in vivo* data, leading to a more cautious uptake of *in vitro* tests to support assessment of such a critical endpoint. Many of these considerations apply also to the CTA based on Bhas 42 cells.

IATAs for non-genotoxic carcinogens

Non-genotoxic carcinogens contribute to an increased cancer risk by a variety of mechanisms that are not yet directly assessed by international regulatory approaches. In April 2014, the OECD WNT recognized that the CTA alone was insufficient to address non-genotoxic carcinogenicity and that a more comprehensive battery of tests addressing different non-genotoxic mechanisms of carcinogenicity would be needed in the future. This discussion led to the identification of the need for an IATA to properly address the issue of non-genotoxic carcinogenicity and where the CTA, together with other relevant assays, could fit. Under the auspices of the OECD, an expert working group was thus set up to examine the current international regulatory requirements and their limitations with respect to non-genotoxic carcinogenicity, and how an IATA could be developed to assist regulators in their assessment of non-genotoxic carcinogenicity (Jacobs et al., 2016). Moreover, the working group is tasked to review, describe and assess relevant *in vitro* assays with the aim of tentatively organizing them into levels of testing, following the adverse outcome pathway format such that possible structure(s) of the future IATA(s) can be created. Different *in vitro* methods are in fact already available as research tools to study a number of potential non-genotoxic mechanisms, such as oxidative stress or inhibition of gap junction intercellular communication (GJIC) (Basketter et al., 2012; Jacobs et al., 2016). Recent work has focused on mapping *in vitro* high-throughput screening (HTS) assays, e.g., from the ToxCast research program, to the hallmarks of cancer (Kleinstreuer et al., 2013a) and the characteristics of carcinogens (Chiu et al., 2018). However, these methods cannot currently be used to reliably predict carcinogenic potential; rather they are useful to better understand the mechanistic basis of effects elicited by a compound, as demonstrated by use in International Agency for Research on Cancer (IARC) monographs, within a weight of evidence strategy (i.e., IATA).

Toxicogenomics-based test methods for carcinogenicity

Toxicogenomics for the study of carcinogenicity has been applied

to several *in vitro* and short-term *in vivo* test systems (Vaccari et al., 2015; Schaap et al., 2015; Worth et al., 2014). For example, the EU-Framework Project carcinoGENOMICS, which aimed at developing *in vitro* toxicogenomics tests to detect potential genotoxicants and carcinogens in liver, lung and kidney cells, also assessed the preliminary reproducibility of the assay using different bioinformatics approaches (Doktorova et al., 2014; Herwig et al., 2016). Potential applications of toxicogenomics-based assays are clarification of mode of action (MoA), hazard classification, derivation of points of departure (PoD) and prioritization (Paules et al., 2011; Waters, 2016). Among these, the targeted use of transcriptomics tests for MoA determination seems to be the preferred application. However, there is still limited implementation of transcriptomics in regulatory decision-making, as discussed in a recent workshop featuring multi-sector and international perspectives on current and potential applications of genomics in cancer risk assessment organized by the Health and Environmental Sciences Institute (HESI), Health Canada and Mc Gill University in Montreal in May 2017. Even though companies make use of transcriptomics-based tests to guide internal decisions, the uncertainty on how these data would be interpreted by regulators is among the main roadblocks identified for submission of data. In addition, lack of validation and regulatory guidance were considered roadblocks (Corvi et al., 2016).

Systematic approaches to carcinogenicity assessment

Identification and incorporation of important, novel scientific findings providing insights into cancer mechanisms is an increasingly essential aspect of carcinogen hazard identification and risk assessment. In recent years, the IARC realized that its process for classifying human carcinogens was complicated by the absence of a broadly accepted, systematic method to evaluate mechanistic data to support conclusions regarding human hazard from exposure to carcinogens. First, no broadly accepted systematic approach was in place for identifying, organizing, and summarizing mechanistic data for the purpose of decision-making in cancer hazard identification. Second, the agents documented and listed as human carcinogens showed a number of characteristics that are shared among many carcinogenic agents. Hence, ten key properties that human carcinogens commonly exhibit and that can encompass many types of mechanistic endpoints were identified. These characteristics were used to conduct a systematic literature search focused on relevant endpoints that provides the basis for an objective approach to identifying and organizing results from pertinent mechanistic studies (Smith et al., 2016).

An example of a comprehensive systematic literature review was recently compiled by Rodgers et al. (2018). Here epidemiologic studies published since 2007, which were related to chemicals previously identified as mammary gland toxicants, were reviewed. The aim was to assess whether study designs captured relevant exposures and disease features, including windows of susceptibility, suggested by toxicological and biological evidence of genotoxicity, endocrine disruption, tumor promotion, or disruption of mammary gland development. Overall, the study added to evidence of links between environmental chemicals and breast cancer.



Beside systematic reviews, IATA can be considered approaches that integrate and weight all relevant existing evidence in a systematic manner to guide the targeted generation of new data, where required, and to inform regulatory decision-making regarding potential hazard and/or risk (e.g., IATA for non-genotoxic carcinogens as described above).

Alternative approaches to rodent long-term carcinogenicity studies for pharmaceuticals

Mainly due to deficiencies of animal carcinogenicity studies and based on some extensive data reviews, representatives of the pharmaceutical industry have leveraged decades of experience to make a proposal for refining the criteria for when carcinogenicity testing may or may not be warranted for pharmaceuticals. In August 2013, an ICH Regulatory Notice Document (RND) was posted by the Drug Regulatory Authorities (DRAs) announcing the evaluation of an alternative approach to the two-year rat carcinogenicity test (ICH Status Report, 2016). This approach is based on the hypothesis that the integration of knowledge of pharmacological targets and pathways together with toxicological and other data can provide sufficient information to anticipate the outcome of a two-year rat carcinogenicity study and its potential value in predicting the risk of human carcinogenicity of a given pharmaceutical. The rationale behind this proposal was supported by a retrospective evaluation of several datasets from industry and drug regulatory agencies, which suggests that up to 40-50% of rat cancer studies could be avoided (ICH Status Report, 2016; Sistare et al., 2011; van der Laan et al., 2016).

A prospective evaluation study to confirm the above hypothesis is ongoing. The industry sponsors are encouraged to submit a carcinogenicity assessment document (CAD) to address the carcinogenic potential of an investigational pharmaceutical and predict the outcome and value of the planned two-year rat carcinogenicity study prior to knowing its outcome (ICH Status Report, 2016). Predictions in the submitted CADs will then be checked against the actual outcome of the two-year rat studies as they are completed. The results of this study are expected for 2019. Currently, the EPAA (European Partnership for Alternative Approaches to Animal Testing) is carrying out a project to evaluate whether a similar approach is also applicable to the carcinogenicity assessment of agrochemicals.

3.3 Reproductive and developmental toxicity

Reproductive toxicity is defined as “*effects such as reduced fertility, effects on gonads and disturbance of spermatogenesis; this also covers developmental toxicity*” (Ferrario et al., 2014), while developmental toxicity is defined as effects of “*e.g., growth and developmental retardation, malformations, and functional deficits in the offspring*”. Often referred to in combination as DART (developmental and reproductive toxicity), the assessment of these endpoints aims to identify possible hazards to the reproductive cycle, with an emphasis on embryotoxicity. Only 2-5% of birth defects are associated with chemical and physical stress (Mattison et al., 2010), including mainly the abuse of alcohol and other drugs, with a far greater percentage attributable to known genetic factors. Overall, approximately 50% of birth defects have unknown causes (Christianson et al., 2006). The

available database is even more limited for the assessment of the prevalence of effects on mammalian fertility.

Similarly, DART was not in the foreground of updates to safety assessments for many years after the shock of the thalidomide disaster (Kim and Scialli, 2011) had died down. More recently, the European REACH legislation, which is extremely demanding in this field (Hartung and Rovida, 2009), has stirred discussion, notably because tests like the two-generation study are among the costliest and require up to 3,200 animals (traditional two-generation study) per substance. In the drug development area, the discussion has focused mainly around a possible replacement of the second species by human mechanistic assays and the value of using non-human primates for biologicals. Another driving force is the European ban on animal testing for cosmetics ingredients (Hartung, 2008b). A series of activities by ECVAM and CAAT, including several workshops, have tackled this challenge. The Integrated Project ReProTect (Hareng et al., 2005) was one of its offspring, pioneering several alternative approaches, followed by projects like Chem-Screen and most recently the flagship program EU-ToxRisk⁴ (Daneshian et al., 2016).

Developmental disruptions are especially difficult to assess (Knudsen et al., 2011), as the timing of processes creates windows of vulnerability, the process of development is especially sensitive to genetic errors and environmental disruptions, simple lesions can lead to complex phenotypes (and *vice versa*), and maternal effects can have an impact at all stages.

The treatment of one or more generations of rats or rabbits with a test chemical is the most common approach to identifying DART, detailed in seven OECD TGs. For specifically evaluating developmental toxicity, TGs were designed to detect malformations in the developing offspring, together with parameters such as growth alterations and prenatal mortality (Collins, 2006). For REACH, developmental toxicity tests are considered mainly as screening tests (Rovida et al., 2011). The shorter and less complex “screening” tests, which combine reproductive, developmental, and (optionally) repeated dose toxicity endpoints into a single study design, are variants.

The fundamental relevance of the current testing paradigm has only recently been addressed in a more comprehensive way (Carney et al., 2011; Basketter et al., 2012). There is considerable concern about inter-species differences (of about 60% concordance), reproducibility (in part due to a lack of standardization of protocols but also high background levels of developmental variants), and the value of the second generation in testing versus the costs, duration and animal use. An analysis of 254 chemicals (Martin et al., 2009b) suggests that 99.8% of chemicals show no-effect-levels for DART within a ten-fold range of maternal toxicity and thus might be simply covered by a safety assessment factor of 10.

An analysis by Bremer and Hartung (2004) of 74 industrial chemicals, which had been tested in developmental toxicity screening tests and reported in the EU New Chemicals Database, showed that 34 chemicals had demonstrated effects on the offspring, but only two chemicals were actually classified as developmentally toxic according to the standards applied by the national competent authorities (Bremer and Hartung, 2004).

**Tab. 3: Milestones on the road towards a new approach to DART**

Date	Event	Who was involved
2002	Validation of three embryotoxicity tests	ECVAM, ZEBET, RIVM
unclear	Zebrafish for DART	Many groups, currently validated by EBTC
2005-2010	ReProTect	ECVAM, University Tübingen (Coordinator Michael Schwarz), 35 partners
2009	Stemina DevTox assay commercially available	Stemina
2012	Acceptance of extended one-generation reproductive toxicity study	ECVAM, ECHA
2008-2017	Definition of TTC	BASF SE, CAAT
2017	Draft Guidance "Detection of toxicity to reproduction for human pharmaceuticals" including suggested reference chemicals for characterizing alternative DART assays	ICH

Abbreviations: BASF SE, German chemical company; CAAT, Center for Alternatives to Animal Testing at Johns Hopkins University; EBTC, Evidence-based Toxicology Collaboration; ECHA, European Chemicals Agency; ECVAM, European Centre for the Validation of Alternative Methods; ICH, International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use; RIVM, Netherlands National Institute for Public Health and the Environment; ZEBET – Center for Documentation and Evaluation of Alternative Methods to Animal Experiments at the German Federal Institute for Risk Assessment

This demonstrates the lack of confidence in the specificity of this “definitive” test. The same analysis showed that 55% of these chemical effects on the offspring could not be detected in multi-generation studies.

The status of alternative methods for DART has been summarized by Adler et al. (2011), endorsed by Hartung et al. (2011), and in the context of developing a roadmap to move forward by Basketter et al. (2012) and Leist et al. (2014). Some key developments are summarized in Table 3 and in the following text.

Extended one-generation reproductive toxicity study

Increasing doubt as to the usefulness of the second generation for testing of substances led to retrospective analyses by Janer et al. (2007), who concluded that this provided no relevant contribution to regulatory decision-making. The US EPA obtained similar data (Martin et al., 2009b) supporting the development of an extended one-generation study (OECD TG 443; OECD, 2011), originally proposed by the ILSI-HESI Agricultural Chemicals Safety Assessment (ACSA) initiative (Doe et al., 2006). The history of the new assay is summarized by Moore et al. (2009). This shows that elements of study protocols can indeed be useless and warrant critical assessment. The reduction brings the number of animals down from 3,200 to about 1,400 per substance tested. Ongoing discussions concern the new animal test’s modules for neurodevelopmental and developmental immunotoxicity, which may be triggered as a result of the extended one-generation study and which undo a lot of the reduction in terms of work and animal use.

Zebrafish embryotoxicity test

In the field of mammalian alternatives, the most complete reflection of embryonic development apparently can be achieved with zebrafish embryos (Selderslaghs et al., 2012; Sukardi et al., 2011; Weigt et al., 2010), for example using dynamic cell imag-

ing, or frog eggs (FETAX assay) (Hoke and Ankley, 2005), with the latter having been evaluated quite critically by ICCVAM.

Currently, the Evidence-based Toxicology Collaboration (EBTC) is evaluating available protocols and data in a systematic review. This retrospective analysis is also exploring whether such systematic reviews (Stephens et al., 2016; Hoffmann et al., 2017) can substitute for traditional validation approaches (Hartung, 2010a). The US National Toxicology Program (NTP) is currently leading the Systematic Evaluation of Application of Zebrafish in Toxicity Testing (SEAZIT) project to assess the impact of varying protocol elements, harmonize ontologies, and develop recommendations around best practices.

Embryotoxicity tests

By 2002, three well-established alternative embryotoxicity tests had already been validated, i.e., the mouse embryonic stem cell test, the whole rat embryo culture and the limb bud assay (Genschow et al., 2004; Piersma et al., 2004; Spielmann et al., 2004). This decade-long validation process represented a radical departure from other validation studies ongoing at that time. They covered only a small, though critical, part of the reproductive cycle and embryonic development. For this reason, none of the tests have received regulatory acceptance in the 15 years since. Although the embryonic stem cell test was validated, the exact regulatory use was still to be defined (Spielmann et al., 2006). The validation study was criticized because the validity statements had raised significant expectations, but such partial replacements could only be used in a testing strategy (Hartung et al., 2013; Rovida et al., 2015) as later attempted within ReProTect and other projects cited above. This prompted a restructuring of the validation process with earlier involvement of regulators and their needs (Bottini et al., 2008), leading among other outcomes to today’s PARERE network at EURL ECVAM. This is only one example, but in general a common problem of



tests that have undergone the classical validation process. This was also addressed and reflected on in the recently published ICCVAM strategic roadmap in conjunction with a clear statement “*The successful implementation of new approach methodologies (NAMs) will depend on research and development efforts developed cooperatively by industry partners and federal agencies. Currently, technologies too often emerge in search of a problem to solve. To increase the likelihood of NAMs being successfully developed and implemented, regulatory agencies and the regulated industries who will ultimately be using new technologies should engage early with test-method developers and stay engaged throughout the development of the technologies.*” (ICCVAM, 2018)

Other critical views as to the validation of alternative embryotoxicity tests concerned the low number of substances evaluated due to the costs of these assays, and the somewhat arbitrary distinction between weak and strong embryotoxicants, where a weak toxicant was defined as being reprotoxic in one species and a strong toxicant being reprotoxic in two or more.

Among the embryotoxicity tests, the murine embryonic stem cell test (EST) has attracted most interest, partly because it represents the only truly animal-free method of the three. Originally based on the counting of beating mouse embryonic stem cell-derived cardiomyocytes, this test has been adapted to other endpoints and to human cells (Leist et al., 2008). It is also used in pharmaceutical industry with revised prediction models. A dedicated workshop on the problems of the EST (Marx-Stoelting et al., 2009) pointed out that its prediction model is overly driven by the cytotoxicity of compounds. Importantly, a variant of the EST using either human embryonic stem cells or human induced pluripotent stem cells and metabolite measurements, which were identified by metabolomics, was introduced by Stemina Biomarker Discovery. This CRO offers contract testing in-house. The assay was evaluated with very promising results for more than 100 substances and is undergoing evaluation by the US EPA and the NTP. There is ongoing discussion with the FDA whether such assays might replace the second species in DART evaluations.

Endocrine disruptor screening assays

Endocrine disruption is one key element of DART but may also constitute a pathway of carcinogenesis. The important assay developments in the context of chemical endocrine disruptor screening go beyond the scope of this short overview. However, they could form critical building blocks in an integrated testing strategy for DART as suggested first by Bremer et al. (2007) and attempted in ReProTect, and for carcinogenicity (Schwarzman et al., 2015).

Computational methods and the threshold of toxicological concern (TTC)

Development of (Q)SAR models for reproductive toxicity is relatively meagre, due to both the complexity of the endpoint and the limited available public data (Hartung and Hoffmann, 2009). The more recent availability of larger toxicity datasets might change this (Hartung, 2016).

An alternative approach has been made by improving TTC for DART (van Ravenzwaay et al., 2017) by expanding earlier

attempts by BASF (Bernauer et al., 2008; van Ravenzwaay et al., 2011, 2012; Laufersweiler et al., 2012). The approach avoids testing by defining doses that are very unlikely to produce a hazard across a large number of chemicals based on the actual use scenario for a given substance of interest (Hartung, 2017c). This work resulted in remarkably high TTC (compared to other endpoints) of 100 µg/kg bw/day for rats and 95 µg/kg bw/day for rabbits for reproductive toxicity. If found acceptable, this could contribute to considerable test waiving.

4 Systems biology and toxicology

“*You think that because you understand ‘one’ that you must therefore understand ‘two’ because one and one make two. But you forget that you must also understand ‘and’.*” This quote by Donella H. Meadows in *Thinking in Systems: A Primer* hits the nail on the head: It is not about knowing the components but about their interrelationships. That is what systems approaches are about. The term has been used mainly for the computational approach of modelling these interrelationships. A key point made here is that there are two systems biology / toxicology approaches – one that is computational and one that is experimental – and they complement each other in addressing the complexity of the organism. Donella H. Meadows, quoted above, phrased it “*The behavior of a system cannot be known just by knowing the elements of which the system is made*”. We will ultimately propose to fuse these approaches, as we can sharpen our modeling tools with data generation in (quality-) controlled MPS. Mathematical modeling has a long history in physiology, but the new added value comes from the generation of big data via the respective measurement technologies (omics, high-content and sensor technologies), and the computational power to make sense of them.

4.1 Experimental systems biology and toxicology

We have recently comprehensively summarized the emergence of microphysiological systems (MPS) (Alépée et al., 2014; Hartung, 2014; Marx et al., 2016), which will not be repeated here. Here, the focus of this review will be on the understanding of how MPS can help to address systemic toxicities and aspects of their quality assurance. MPS bring a certain face-validity to the portfolio of tools as they introduce organ architecture, representative complexity and functionality to the *in vitro* approaches and increasingly even incorporate (patho-)physiological organ interactions. A critical element is the proper reflection of ADME, but microfluidics offers many opportunities to approach this goal (Slikker, 2014). The promise of MPS in biomedicine and drug development depends critically on their quality control. Especially, regulatory decisions based on them will require a high degree of confidence, which only strict quality control can create.

The quality assurance of MPS again requires an adaptation of the validation paradigm. Concepts of validation originally shaped around relatively simple cell systems for regulatory decision-taking as an alternative to animal testing. Three decades of experience have laid the foundation to broaden this concept

to MPS in the context of their use in the life sciences and especially in drug development (Abaci and Shuler, 2015; Ewart et al., 2017; Skardal et al., 2016, 2017).

The FDA MPS program

FDA recognizes that alternative test platforms like organs-on-chip can give regulators new tools that are more predictive. However, for these new alternative methods to be acceptable for regulatory use, confidence is needed that the questions can be answered by these new methods as with traditional testing. Fostering collaborations between government researchers and regulators and between regulators, industry, stakeholders and academia can ensure that the most promising technologies are identified, developed, validated and integrated into regulatory risk assessment. The FDA-DARPA-NIH Microphysiological Systems Program started in 2011 to support the development of human microsystems, or organ “chips”, to screen swiftly and efficiently for safe and effective drugs (before human testing). It represents a collaboration through coordination of independent programs:

- Defense Advanced Research Projects Agency (DARPA): Engineering platforms and biological proof-of-concept (DARPA-BAA-11-73: Microphysiological Systems)
- National Institutes of Health (NIH), National Center for Advancing Translational Sciences (NCATS): Underlying biology/pathology and mechanistic understanding (RFA-RM-12-001 and RFA RM-11-022)
- Food and Drug Administration (FDA): Advice on regulatory requirements, validation and qualification.

This was a unique partnership because it involved regulatory scientists at the very beginning and thus was able to address identified gaps in knowledge needed to regulate FDA products (Fig. 2).

As an outcome of the program, in April 2017, the FDA signed a Cooperative Research and Development Agreement (CRADA) with Emulate, Inc. to use organs-on-chips technology as a toxicology testing platform to understand how products affect human health and safety. It aims to advance and qualify their “Human Emulation System” to meet regulatory evaluation criteria for product testing^{5,6}. The FDA will evaluate the company’s “organs-on-chips” technology in laboratories at the agency’s Center for Food Safety and Applied Nutrition (CFSAN). Their miniature liver-on-chip will be evaluated as to its effectiveness to better understand the effects of medicines, disease-causing bacteria in foods, chemicals, and other potentially harmful materials on the human body. FDA will beta-test the Emulate system and look at concordance of chip data with *in vivo*, *in silico* and other *in vitro* (2-D) data on the same compounds; furthermore, FDA will begin to develop performance standards for organs-on-chips to create a resource for FDA regulators and researchers.

The work is part of the FDA Predictive Toxicology Roadmap announced December 6, 2017⁷. An FDA senior level toxicology working group was formed to foster enhanced communication among FDA product centers and researchers and leverage FDA resources to advance the integration of emerging predictive toxicology methods and new technologies into regulatory safety and risk assessments. This will include training of FDA regulators and researchers with continuing ongoing education in new pre-

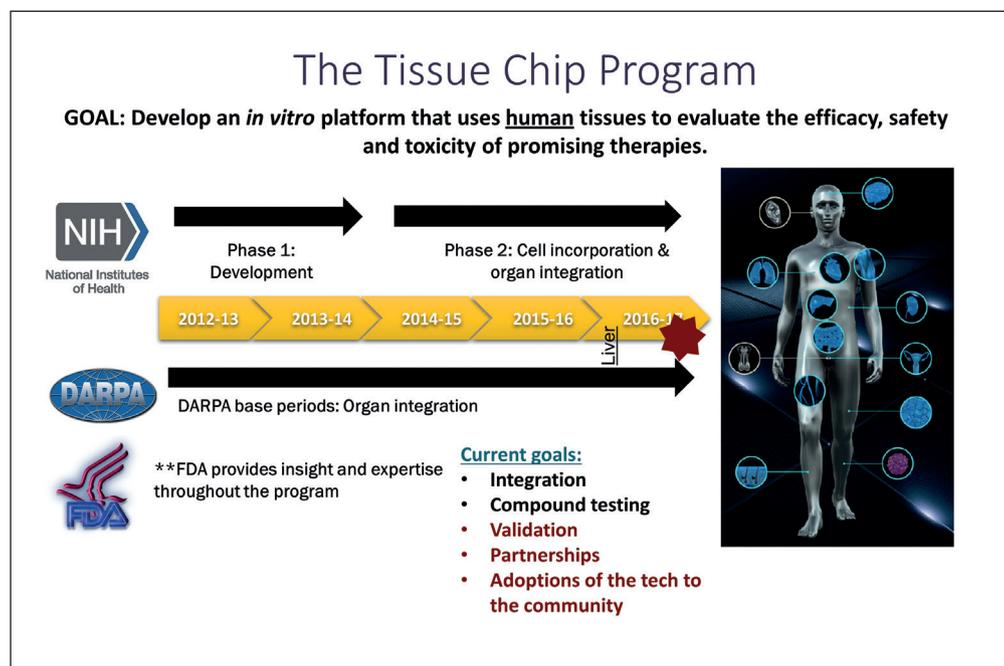


Fig. 2: The FDA-DARPA-NIH Microphysiological Systems Program

Abbreviations: NIH, National Institutes of Health USA; FDA, Food and Drug Administration USA; DARPA, Defense Advanced Research Projects Agency USA

⁵ <https://emulatebio.com/press/fda-collab-agreement-emulate/>

⁶ <https://bit.ly/2ovDVNC>

⁷ <https://bit.ly/2ixBTID>



dictive toxicology methods that are essential for FDA regulators. As part of this, FDA established an agency-wide education calendar of events and a Toxicology Seminar Series to introduce concepts of new toxicology methodologies and updates on toxicology-related topics. In order to promote continued communication, FDA reaffirmed its commitment to incorporate data from newly qualified toxicology methods into regulatory missions, is encouraging discussions with stakeholders as part of the regulatory submission process, and encourages sponsors to submit scientifically valid approaches for using a new method early in the regulatory process. FDA fosters collaborations with stakeholders across sectors and disciplines nationally and internationally. This is pivotal to identify the needs, maintain momentum, and establish a community to support delivery of new predictive toxicology methods. With this goal, FDA's research programs will identify data gaps and support intramural and extramural research to ensure that the most promising technologies are identified, developed, validated, and integrated into the product pipeline. Under the oversight of the Office of the Commissioner, the progress of these recommendations will be tracked, including an annual report to the Chief Scientist. This shall ensure transparency, foster opportunities to share ideas and knowledge, showcase technologies, and highlight collaborations on developing and testing new methods.

In conclusion, the FDA roadmap identifies the critical priority activities for energizing new or enhanced FDA engagement in transforming the development, qualification, and integration of new toxicology methodologies and technologies into regulatory application. Implementation of the roadmap and engagement with diverse stakeholders should enable FDA to fulfill its regulatory mission today while preparing for the challenges of tomorrow.

Validation of M(P)PS

Quality assurance and ultimately validation of the tools in the life sciences is a key contribution to overcome the stagnant drug development pipeline due to high attrition rates and the reproducibility crisis in biomedicine. MPS bring a certain face-validity to the portfolio of tools as they introduce organ architecture and functionality to the *in vitro* approaches and increasingly even incorporate (patho-)physiological organ system interactions. With more MPS developing, the major challenge for their use as translational drug development tools is to make micropathophysiological systems (MPPS). The promise of MPS in biomedicine and drug development depends critically on their quality control. Especially, regulatory decisions will require a confidence that only strict quality control can create.

Typically, the new test would be compared to a traditional method, usually an animal experiment, and the relative reproducibility of reference results would be used as the primary measure of success. Concurrent testing of new substances with the reference test represents another opportunity to gain comparative information without the information bias of the scientific literature (e.g., overrepresentation of toxic substances with specific mechanisms). In an ECVAM workshop (Hoffmann et al., 2008), it was suggested that instead of a specified reference (animal) test, a reference standard could be formed by expert

consensus by integrating all knowledge; a list of substances could be produced with results a hypothetical ideal test would provide. This can for example allow using also human data in combination with animal data or combination of results from various test systems.

These concepts of correlative validation are only partially applicable to MPS, which often have many purposes other than replacing an animal test, and for which in many cases a respective animal test does not even exist. For drug development, typically a pathophysiological state first needs to be introduced and then treatment effects are analyzed. This greatly complicates the validation process, as both the induction of pathophysiology and its correction need to be quality assured.

MPS are usually more relevant based on the mechanisms of physiology and pathophysiology they reflect. For this reason, mechanistic validation (Hartung et al., 2013) lends itself to the evaluation of MPS: This is first of all a comparison to mechanisms from the scientific literature, ideally by systematic review. Alternatively, high-content characterizations of a reference test and the new test can show that similar patterns of perturbation of physiology occur, in the easiest case that the same biomarkers of effect are observed. This experimental approach can be applied where the definition of mechanism is incomplete or the existing literature insufficient. Lastly, computational modelling of physiology and the prediction of test outcomes in comparison to real test data can show how well the test and the computational model align. The envisaged validation process for MPS has to start with the information need, which defines the purpose of the test.

Although validation is often perceived as rigid and inflexible (which it has to be once a study is initiated), it is actually a highly flexible process, which needs to be adapted case by case and should be performed with the end use in mind (ICCVAM, 2018). Concepts of pre-validation, applicability domain, retrospective

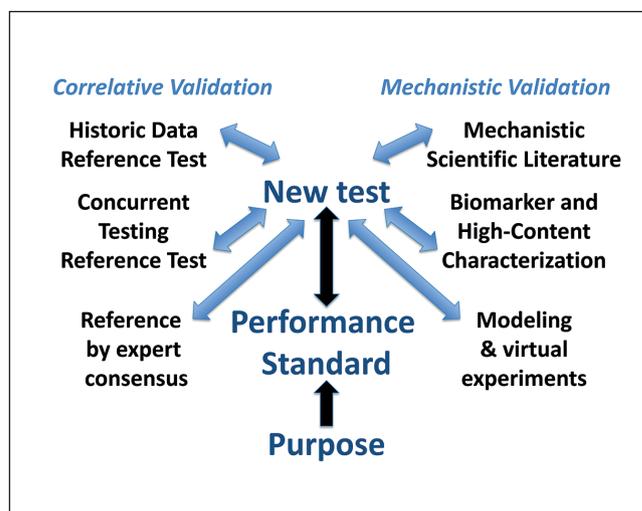


Fig. 3: The concept of performance standard-based validation
The different elements for anchoring a validation in a correlative or mechanistic manner will be combined by expert consensus to define a performance standard meeting a test purpose.

validation, catch-up-validation, minimal performance standards, prediction models, etc. are examples of the continuing adaptation to meet the needs of stakeholders (Hartung, 2007; Leist et al., 2012). Here, especially the concepts of “fit-for-purpose”, meeting defined “performance standards”, and “mechanistic validation” will have to be elaborated upon, specific to MPS (Fig. 3).

- “Fit-for-purpose”: The purpose of a test is its place and function in a testing strategy to meet an overall information need and decision context, e.g., the information need is developmental neurotoxicity with the focus on one of the key events of neural development such as myelination of axons. The question to be addressed can be the following: Do certain substances perturb myelination of neuronal axons? Then, a first test could assess toxicity to oligodendrocytes. A second test could quantify the level of myelin basic protein (MBP) in MPS. A third test might assess electrophysiology within the organoid as a functional outcome of perturbed myelination and as a consequence of the perturbation of neurodevelopment and neural differentiation. The testing strategy would need to combine these test results (evidence integration) with existing information.
- “Performance standards”: The concept of a performance standard for alternative methods was introduced in the Modular Approach in 2004 (Hartung et al., 2004) and incorporated into OECD validation guidance from 2005 (OECD, 2005). The basic idea is that if a successfully validated method is available, it should be defined what a similar method should demonstrate to be considered equivalent to the validated one. This has proven to be crucial for any modification of tests as well as to avoid extensive and expensive retesting for similar tests. For this reason, they were originally termed “minimum performance standards”. Over time, the concept has evolved, now also using the performance standards among others to show the proficiency of a laboratory to carry out a test. Most radically, the current work on developing a performance standard-based OECD TG for a skin sensitization defined approach (DA) aims at defining how any test or combination of tests should perform to be acceptable under the guidance without prescribing a specific method. By extension, a performance standard could be defined for an MPS: This means setting engineering goals (performance standards) and the quality assurance (validation) process would confirm that these standards are met. To some extent this is similar to the reasoning of an earlier ECVAM workshop on points of reference, where it was recommended to define a point of reference by expert consensus for a given validation, not by comparing to a dataset from a traditional animal test (Hoffmann et al., 2008). This was first applied in the retrospective validation of the micronucleus *in vitro* test validation (Corvi et al., 2008) and later in the more recent validation studies of micronucleus and comet assays in 3D skin models, and it will be applied in the future for the validation of thyroid endocrine disruptor tests.
- “Mechanistic validation” (Hartung et al., 2013) is another radical departure from current practice. Though validation has always included the aspect of mechanistic relevance when addressing test definition, this is usually only minimally covered. The traditional (animal) test and the new method are

typically taken as black boxes and the correlation of their results is the measure of validity. MPS bring (patho)physiology, i.e., mechanism, to the foreground. Thus, it makes sense to use a mechanistic basis for comparison. Mechanistic validation dictates first an agreement on the relevant mechanisms for a given information need, followed by evaluation based on coverage of the mechanism by the new method. This type of an approach increasingly takes place with the definition of adverse outcome pathways (AOP) and was the goal of the parallel Human Toxome Project (Bouhifd et al., 2015). One of the basic underpinnings of mechanistic validation is that a systematic review of the literature can be used to ascertain mechanism.

Even before attempting formal validation of MPS, their quality assurance will be of utmost importance. The Good Cell Culture Practice (GCCP) movement initiated by one of the authors in 1996 led to the first guidance of its kind (Coecke et al., 2005) under the auspices of ECVAM. The international community recognized a need to expand this to MPS and stem cell-based models ten years later, and under the lead of CAAT, with participation of FDA, NIH NCATS, NICEATM, ECVAM, UK Stem Cell Bank and others, GCCP 2.0 was initiated. In two dedicated workshops and several publications (Pamies et al., 2017b, 2018b; Pamies and Hartung, 2017; Eskes et al., 2017), the needs were defined, and a steering group plus scientific advisory group is currently formulating GCCP 2.0. The proof-of-principle of validation attempts by NIH NCATS in establishing Tissue Chip Testing Centers (TCTC) will cross-fertilize with these developments. The GCCP discussion was already the topic of workshops and conferences such as European Society of Toxicology *In Vitro* 2016, EuroTox 2017, Society of Toxicology 2018 and a joint conference with FDA and the IQ consortium in 2015. A 2017 workshop (Bal-Price et al., 2018) developed test readiness criteria for toxicology using the example of developmental neurotoxicity, which will be a further starting point for the performance standard development attempted here.

Recognizing the need for a stakeholder dialogue on the quality assurance of MPS, CAAT this year initiated a Public Private Partnership for Performance Standards for Microphysiological Systems (P4M), which aims to establish a stakeholder consensus process toward performance standards. P4M will discuss the core aspects, i.e., when is an MPS good enough and can we express this as a performance standard? Expressions of interest already received include various companies, academics, ECVAM, and US and Japanese agencies.

4.2 Computational systems biology and toxicology

J. B. S. Haldane (1892-1964), a biologist and mathematician, predicted “*If physics and biology one day meet, and one of the two is swallowed up, that one will be biology*”. Systems biology is biology swallowed by physics. Joyner and Pederson (2011) give an interesting reflection on this discipline. Systems toxicology (Kiani et al., 2016), its more applied sibling, was the topic of an earlier article in this series (Hartung et al., 2012), some dedicated conferences and symposia (Andersen et al., 2014; Sturla et al., 2014; Sauer et al., 2015; Hartung et al., 2017a) and a special issue of *Chemical Research in Toxicology* (Hartung et



al., 2017b). As experimental systems biology has been fueled by bioengineering and stem cell technologies, computational systems biology / toxicology has been driven by big data and machine learning technologies (Hartung, 2016; Luechtefeld and Hartung, 2017). The ultimate vision is using computational models of human metabolism, possibly as avatars or virtual patients, to try out pharmacological interventions or toxic insults; on the way, tissue and organ models are emerging (Hartung, 2017d).

Systems biology approaches biological function and its perturbation by various biochemically active compounds by complementing the traditional reductionist approaches. The emphasis of systems biology approaches is on the interactions between components rather than just the components themselves. This approach is therefore frequently focused on the dynamics of biological interactions and the emergent properties of biological cells, tissues and organisms stemming from the complexity of the underlying regulatory networks.

The systems biology analysis allows one to examine the disruption of network components by pharmacological and other interventions through the lens of their effects, not only on the designated target but on the network of molecular components, with frequently paradoxical, unexpected and counter-intuitive results. These results can be products of complex feedback interactions involving a specific target and the multiple phenotypes controlled by it, rather than just off-target biochemical effects. The network level effects can span multiple scales, from biochemical to cellular and tissue levels, which involve cell-cell communication through various signaling mechanisms, producing networks of networks. This complexity is captured through high-throughput experimentation and computational analysis and modelling, with a particular focus on the unanticipated, emergent properties. Below we provide some examples of the recent systems biology approaches to complex problems related to the mechanisms of drug action and possible toxic effects.

Several recent examples of the systems approach illustrate the philosophy and power of the approach. Particular attention so far has been paid to the complex mechanisms of action of cocktails of various pharmaceutical compounds. For instance, a recent systems analysis demonstrated that the order and timing of application of anti-cancer compounds can determine the efficacy of combinatorial treatments (Lee et al., 2012). This effect has been ascribed to re-wiring of the signaling network by the first compound, which might result in a more potent effect of the second compound if applied at the appropriate time. The same dynamic network view can be applied to combinatorial applications of radio- and chemotherapeutic treatments, as elucidated through mathematical modeling and experimental validation in another high-profile systems biology application (Chen et al., 2016). These types of network perspectives and associated modeling support will likely also inform the analysis of the effects of putative combinatorial treatments on other tissues and the associated toxicology outcomes.

Another example of a study benefiting from a systems approach is the paradoxical increase rather than decrease of the total kinase activity by ATP-competitive inhibitors of BRAF/CRAF kinases (Hatzivassiliou et al., 2010). As these kinases are a key target in various cancers, the paradoxical effect has

received much attention. However, it is virtually impossible to rationalize it without a modeling approach within a framework of systems biology. Two key insights had to be made to formulate hypotheses of how this might occur, including a feedback regulation inherent in the RAF/MAPK kinase signaling and potential allosteric action of the drugs on the enzyme (Kholodenko et al., 2015).

A number of recent efforts to build and apply computational systems models have focused specifically on mechanisms of developmental toxicity. The US EPA's Virtual Tissues research project uses cellular agent-based models to recapitulate developing embryonic systems and creates *in silico* testing platforms by parameterizing such models using the ToxCast/Tox21 HTS data to mimic chemical exposure and simulate effects on a tissue level. An AOP of embryonic vascular disruption was published based on a systematic literature review (Knudsen and Kleinstreuer, 2012) and was used to inform the construction of a computational model predicting disruption of blood vessel development (Kleinstreuer et al., 2013b). Putative vascular disruptor compounds and associated systems model predictions have been tested and confirmed in a number of functional assays such as transgenic zebrafish, human cell-based tubulogenesis assays, and whole embryo culture (Tal et al., 2016; McCollum et al., 2016; Ellis-Hutchings et al., 2017). Other work has focused on modelling key developmental toxicity mechanisms driving cleft palate formation (Hutson et al., 2017) and taking a systems toxicology approach to understanding disruption of male reproductive development and endocrine pathways (Leung et al., 2016; Kleinstreuer et al., 2016).

4.3 A fusion of experimental and computational systems biology / toxicology?

Even though MPS are complex, they are considerably simpler than human organisms and they are much more open for measurements and interventions. Thus, the opportunity to first model our experimental systems has enormous advantages; however, it represents an interdisciplinary challenge. Bioengineers and modelers have to be brought together. At the same time, funding bodies have to be convinced of the value of this interim step. As an example, in a recent organ-on-chip study (Kilic et al., 2016), mediator gradients were modeled computationally and then verified experimentally. By parameterization of the experimental systems, we can also start to scale our systems virtually as a quantitative *in vitro* to *in vivo* extrapolation (QIVIVE) (Tsaion et al., 2016; Hartung, 2018).

5 Conclusions

Overall, further discussion is needed as to the relevance of current carcinogenicity, RDT and DART assessments. Recognizing the societal need to ensure the safety of drugs, chemicals and consumer products, this might make it difficult to abandon current testing, but should lower the barrier for implementing alternative approaches that may improve the status quo. Data sharing and the harmonization of ontologies and data formats will be critical.



Repeated-dose toxicity, carcinogenicity and reproductive toxicity are three examples of systemic toxicology implementation. Since they are very complex endpoints, the uptake of alternative *in vitro* test methods is still very limited. Rather, some approaches are being investigated or are already in place for waiving testing and reducing the number of animals, such as the ICH strategy for pharmaceuticals and the extended one-generation assay.

Promise for all systemic toxicities comes from the starting development of integrated testing strategies driven by mechanistic relevance: By mapping the human reproductive cycle and its disturbance or the array of pathways of carcinogenesis with a number of assays, the hope is to design more human-relevant test strategies. These and other approaches form part of the emerging roadmap for replacement (Basketter et al., 2012; Leist et al., 2014; Corvi et al., 2017; ICCVAM, 2018) and will contribute to the momentum for implementing alternative approaches, which is also aided by the increasing recognition of the shortcomings of current testing methods.

Given the importance of these hazards and the backlog of testing for most substances of daily use, more efforts in the development of tests, design of testing strategies and their validation are needed. Quality assurance and ultimately validation of the tools in the life sciences is a necessity to unplug the drug development pipeline, which is blocked by high attrition rates, and the reproducibility crisis in biomedicine. The systematic condensation of our existing knowledge (including the mapping of gaps and shortcomings of existing evidence) can herald a more predictive systems approach to addressing systemic toxicity.

However, ultimately, we need a “new deal” for systemic toxicities. Albert Einstein once said, “*We can't solve the problems by using the same kind of thinking we used when we created them*”. The increasing awareness of the shortcomings of current tests with respect to reproducibility (Baker, 2016; Jilka, 2016; Voelkl et al., 2018), inter-species differences and thus lack of human relevance, ambiguity of results and steep costs, should make all who are in this field uneasy (Miller, 2014). It requires the “art of toxicology” to make good decisions on the basis of such compromised information sources. How can we sleep well when we know that our daily decisions are subject to these limitations? Rasheed Ogunlaru wrote, “*All the tools, techniques and technology in the world are nothing without the head, heart and hands to use them wisely, kindly and mindfully*”. This holds for the current art of toxicology and will likely be no different for any new approach. Especially as the new approaches come in the guise of objective “evidence-based” and high-tech approaches, they are still models created with a purpose. Frank Herbert (in God Emperor of Dune) warns “*Dangers lurk in all systems. Systems incorporate the unexamined beliefs of their creators. Adopt a system, accept its beliefs, and you help strengthen the resistance to change*”.

The 3Rs have served us to some extent to replace animal tests for acute and topical toxicities but have done so by modelling and reproducing the animal test results despite their shortcomings. A testing strategy modelling the outcomes of traditional tests will not serve us as well for systemic toxicities. The hazards are typically more severe and less directly attributable to exposure because they can manifest anywhere in the body and after

any time of exposure. The 3S approach suggested here is such a “new deal” for safety assessments. It goes far beyond the 3Rs as it does not aim to reproduce the results of a black box (animal) test, which may bear little resemblance to the human scenario. The combination of systematic evaluation of our knowledge and experimental as well as computational modelling of biological systems complexity promises a different approach to systemic toxicity prediction, even though it still has to prove its feasibility and utility.

References

- Abaci, H. E. and Shuler, M. L. (2015). Human-on-a-chip design strategies and principles for physiologically based pharmacokinetics/pharmacodynamics modeling. *Integr Biol* 7, 383-391. doi:10.1039/C4IB00292J
- Adler, S., Basketter, D., Creton, S. et al. (2011). Alternative (non-animal) methods for cosmetics testing: Current status and future prospects. *Arch Toxicol* 85, 367-485. doi:10.1007/s00204-011-0693-2
- Ågerstrand, M. and Beronius, A. (2016). Weight of evidence evaluation and systematic review in EU chemical risk assessment: Foundation is laid but guidance is needed. *Env Int* 92-93, 590-596. doi:10.1016/j.envint.2015.10.008
- Aiassa, E., Higgins, J. P. T., Frampton, G. K. et al. (2015). Applicability and feasibility of systematic review for performing evidence-based risk assessment in food and feed safety. *Crit Rev Food Sci Nutr* 55, 1026-1034. doi:10.1080/10408398.2013.769933
- Alden, C. L., Lynn, A., Bourdeau, A. et al. (2011). A critical review of the effectiveness of rodent pharmaceutical carcinogenesis testing in predicting human risk. *Vet Pathol* 48, 772-784. doi:10.1177/0300985811400445
- Alépée, N., Bahinski, T., Daneshian, M. et al. (2014). State-of-the-art of 3D cultures (organs-on-a-chip) in safety testing and pathophysiology – A t⁴ report. *ALTEX* 31, 441-477. doi:10.14573/altex1406111
- Anand, P., Kunnumakara, B. A., Sundaram, C. et al. (2008). Cancer is a preventable disease that requires major lifestyle changes. *Pharm Res* 25, 2097-2116. doi:10.1007/s11095-008-9661-9
- Andersen, M. E., Betts, K., Dragan, Y. et al. (2014). Developing microphysiological systems for use as regulatory tools – Challenges and opportunities. *ALTEX* 31, 364-367. doi:10.14573/altex.1405151
- Anisimov, V. N., Ukraintseva, S. V. and Yashin, A. I. (2005). Cancer in rodents: Does it tell us about cancer in humans? *Nat Rev Cancer* 5, 807-819. doi:10.1038/nrc1715
- Ashby, J. and Tennant, R. W. (1991). Definitive relationships among chemical structure, carcinogenicity and mutagenicity for 301 chemicals tested by the U.S. NTP. *Mutat Res* 257, 229-306. doi:10.1016/0165-1110(91)90003-E
- Bal-Price, A., Hogberg, H. T., Crofton, K. M. et al. (2018). Recommendation on test readiness criteria for new approach methods in toxicology: Exemplified for developmental neurotoxicity. *ALTEX*, Epub ahead of print. doi:10.14573/altex.1712081
- Bailey, J., Knight, A. and Balcombe, J. (2005). The future of



- teratology research is in vitro. *Biogenic Amines* 19, 97-145. doi:10.1163/1569391053722755
- Baker, M. (2016). Is there a reproducibility crisis? *Nature* 533, 452-454. doi:10.1038/533452a
- Basketter, D. A., Clewell, H., Kimber, I. et al. (2012). A roadmap for the development of alternative (non-animal) methods for systemic toxicity testing. *ALTEX* 29, 3-91. doi:10.14573/altex.2012.1.003
- Batke, M., Aldenberg, T., Escher, S. and Mangelsdorf, I. (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
- Bernaer, U., Heinemeyer, G., Heinrich-Hirsch, B. et al. (2008). Exposure-triggered reproductive toxicity testing under the REACH legislation: A proposal to define significant/relevant exposure. *Toxicol Lett* 176, 68-76. doi:10.1016/j.toxlet.2007.10.008
- Bernstein, L., Gold, L. S., Ames, B. N. et al. (1985). Some tautologous aspects of the comparison of carcinogenic potency in rats and mice. *Fundam Appl Toxicol* 5, 79-86. doi:10.1016/0272-0590(85)90051-X
- Birnbaum, L. S. and Fenton, S. (2003). Cancer and developmental exposure to endocrine disruptors. *Environ Health Perspect* 111, 389-394. doi:10.1289/ehp.5686
- Bokkers, B. G. H. and Slob, W. (2007). Deriving a data-based interspecies assessment factor using the NOAEL and the benchmark dose approach. *Crit Rev Toxicol* 37, 355-373. doi:10.1080/10408440701249224
- Bottini, A. A., Alepee, N., De Silva, O. et al. (2008). Optimization of the post-validation process. The report and recommendations of ECVAM workshop 67. *Altern Lab Anim* 36, 353-366.
- Bouhifd, M., Andersen, M. E., Baghdikian, C. et al. (2015). The human toxome project. *ALTEX* 32, 112-124. doi:10.14573/altex.1502091
- Bremer, S. and Hartung, T. (2004). The use of embryonic stem cells for regulatory developmental toxicity testing in vitro – The current status of test development. *Curr Pharm Des* 10, 2733-2747. doi:10.2174/1381612043383700
- Bremer, S., Pellizzer, C., Hoffmann, S. et al. (2007). The development of new concepts for assessing reproductive toxicity applicable to large scale toxicological programs. *Curr Pharm Des* 13, 3047-3058. doi:10.2174/138161207782110462
- Brown, N. A. and Fabro, S. (1983). The value of animal teratogenicity testing for predicting human risk. *Clin Obstet Gynecol* 26, 467-477. doi:10.1097/00003081-198306000-00028
- Busquet, F. and Hartung, T. (2017). The need for strategic development of safety sciences. *ALTEX* 34, 3-21. doi:10.14573/altex.1701031
- Carney, E. W., Ellis, A. L., Tyl, R. W. et al. (2011). Critical evaluation of current developmental toxicity testing strategies: A case of babies and their bathwater. *Birth Defects Res B: Develop Reprod Toxicol* 92, 395-403. doi:10.1002/bdrb.20318
- Chen, M., Bisgin, H., Tong, L. et al. (2014). Toward predictive models for drug-induced liver injury in humans: Are we there yet? *Biomark Med* 8, 201-213. doi:10.2217/bmm.13.146
- Chen, S.-H., Forrester, W. and Lahav, G. (2016). Schedule-dependent interaction between anticancer treatments. *Science* 351, 1204-1208. doi:10.1126/science.aac5610
- Chiu, W. A., Guyton, K. Z., Martin, M. T. et al. (2018). Use of high-throughput in vitro toxicity screening data in cancer hazard evaluations by IARC Monograph Working Groups. *ALTEX* 35, 51-64. doi:10.14573/altex.1703231
- Christianson, A., Howson, C. P. and Modell, B. (2006). *March of Dimes Global Report of Birth Defects: The hidden toll of dying and disabled children*. New York, USA: March of Dimes Birth Defects Foundation. <https://bit.ly/2GFyjXi>
- Coecke, S., Balls, M., Bowe, G. et al. (2005). Guidance on good cell culture practice. *Altern Lab Anim* 33, 261-287.
- Colditz, G. A. and Wei, E. K. (2012). Preventability of cancer: The relative contributions of biologic and social and physical environmental determinants of cancer mortality. *Annual Rev Publ Health* 33, 137-156. doi:10.1146/annurev-publhealth-031811-124627
- Collins, T. F. (2006). History and evolution of reproductive and developmental toxicology guidelines. *Curr Pharm Des* 12, 1449-1465. doi:10.2174/138161206776389813
- Combes, R., Balls, M., Curren, R. et al. (1999). Cell transformation assay as predictors of human carcinogenicity. *Altern Lab Anim* 27, 745-767.
- Corvi, R., Albertini, S., Hartung, T. et al. (2008). ECVAM retrospective validation of in vitro micronucleus test (MNT). *Mutagenesis* 23, 271-283. doi:10.1093/mutage/gen010
- Corvi, R., Aardema, M. J., Gribaldo, L. et al. (2012). ECVAM prevalidation study on in vitro cell transformation assays: General outline and conclusions of the study. *Mut Res* 744, 12-19. doi:10.1016/j.mrgentox.2011.11.009
- Corvi, R., Vilardell, M., Aubrecht, J. and Piersma, A. (2016). Validation of transcriptomics-based in vitro methods. In C. Eskes and M. Whelan (eds.), *Validation of Alternative Methods for Toxicity Testing, Advances in Experimental Medicine and Biology*. Vol. 856, 243-257. Switzerland: Springer. doi:10.1007/978-3-319-33826-2_10
- Corvi, R. and Madia, F. (2017). In vitro genotoxicity testing: Can the performance be enhanced? *Food Chem Toxicol* 106, 600-608. doi:10.1016/j.fct.2016.08.024
- Corvi, R., Madia, F., Guyton, K. Z. et al. (2017). Moving forward in carcinogenicity assessment: Report of an EURL ECVAM-ESTIV workshop. *Toxicol In Vitro* 45, 278-286. doi:10.1016/j.tiv.2017.09.010
- Daneshian, M., Busquet, F., Hartung, T. and Leist, M. (2015). Animal use for science in Europe. *ALTEX* 32, 261-274. doi:10.14573/altex.1509081
- Daneshian, M., Kamp, H., Hengstler, J. et al. (2016). Highlight report: Launch of a large integrated European in vitro toxicology project: EU-ToxRisk. *Arch Toxicol* 90, 1021-1024. doi:10.1007/s00204-016-1698-7
- Daston, G. P. and Seed, J. (2007). Skeletal malformations and variations in developmental toxicity studies: Interpretation issues for human risk assessment. *Birth Defects Res B Dev Reprod Toxicol* 80, 421-424. doi:10.1002/bdrb.20135
- Doe, J. E., Boobis, A. R., Blacker, A. et al. (2006). A tiered approach to systemic toxicity testing for agricultural chemical safety assessment. *Crit Rev Toxicol* 36, 37-68. doi:10.1080/10408440500534370

- Doktorova, T. Y., Yildirimman, R., Celeen, L. et al. (2014). Testing chemical carcinogenicity by using a transcriptomics HepaRG-based model? *EXCLI J* 13, 623-637.
- Ellis-Hutchings, R., Settivari, R., McCoy, A. et al. (2017). Embryonic vascular disruption adverse outcomes: Linking HTS signatures with functional consequences. *Reprod Toxicol* 70, 82-96. doi:10.1016/j.reprotox.2017.04.003
- Ennever, F. K., Noonan, T. J. and Rosenkranz, H. S. (1987). The predictivity of animal bioassays and short-term genotoxicity tests for carcinogenicity and non-carcinogenicity to humans. *Mutagenesis* 2, 73-78. doi:10.1093/mutage/2.2.73
- Ennever, F. K. and Lave, L. B. (2003). Implications of the lack of accuracy of the lifetime rodent bioassay for predicting human carcinogenicity. *Regul Toxicol Pharmacol* 38, 52-57. doi:10.1016/S0273-2300(03)00068-0
- Eskes, C., Boström, A.-C., Bowe, G. et al. (2017). Good cell culture practices & in vitro toxicology. *Toxicol In Vitro* 45, 272-277. doi:10.1016/j.tiv.2017.04.022
- EURL ECVAM (2012). EURL ECVAM Recommendation on three cell transformation assays using Syrian hamster embryo cells (SHE) and the BALB/c 3T3 mouse fibroblast cell line for in vitro carcinogenicity testing. <https://bit.ly/2qbJ50J>
- EURL ECVAM (2013a). EURL ECVAM strategy to avoid and reduce animal use in genotoxicity testing. *JRC Report EUR 26375*. <https://bit.ly/19V6dzF>
- EURL ECVAM (2013b). EURL ECVAM Recommendation on cell transformation assay based on the Bhas 42 cell line. <https://bit.ly/2qbJ50J>
- Ewart, L., Fabre, K., Chakilam, A. et al. (2017). Navigating tissue chips from development to dissemination: A pharmaceutical industry perspective. *Exp Biol Med* 242, 1579-1585. doi:10.1177/1535370217715441
- Ferrario, D., Brustio, R. and Hartung, T. (2014). Glossary of reference terms for alternative test methods and their validation. *ALTEX* 31, 319-335. doi:10.14573/altex.140331
- Freedman, D. A. and Zeisel, H. (1988). From mouse to man: The quantitative assessment of cancer risks. *Statist Sci* 3, 3-56. doi:10.1214/ss/1177012993
- Freedman, L. P., Cockburn, I. M. and Simcoe, T. S. (2015). The economics of reproducibility in preclinical research. *PLoS Biol* 13, e1002165. doi:10.1371/journal.pbio.1002165
- Gaylor, D. W. (2005). Are tumor incidence rates from chronic bioassays telling us what we need to know about carcinogens? *Regul Toxicol Pharmacol* 41, 128-133. doi:10.1016/j.yrtph.2004.11.001
- GBD 2013 Risk Factors Collaborators (2013). Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks in 188 countries, 1990-2013: A systematic analysis for the global burden of disease study 2013. *Lancet* 386, 2287-2323. doi:10.1016/S0140-6736(15)00128-2
- Genschow, E., Spielmann, H., Scholz, G. et al. (2004). Validation of the embryonic stem cell test in the international ECVAM validation study on three in vitro embryotoxicity tests. *Altern Lab Anim* 32, 209-244.
- Gold, L. S., Slone, T. H., Manley, N. B. et al. (1991). Target organs in chronic bioassays of 533 chemical carcinogens. *Environ Health Perspect* 93, 233-246. doi:10.1289/ehp.9193233
- Gold, L. S., Slone, T. H. and Ames, B. N. (1998). What do animal cancer tests tell us about human cancer risk? Overview of analyses of the carcinogenic potency database. *Drug Metab Rev* 30, 359-404. doi:10.3109/03602539808996318
- Gordon, S., Daneshian, M., Bouwstra, J. et al. (2015). Non-animal models of epithelial barriers (skin, intestine and lung) in research, industrial applications and regulatory toxicology. *ALTEX* 32, 327-378. doi:10.14573/altex.1510051
- Gottmann, E., Kramer, S., Pfahringer, B. and Helma, C. (2001). Data quality in predictive toxicology: Reproducibility of rodent carcinogenicity experiments. *Environ Health Perspect* 109, 509-514. doi:10.1289/ehp.01109509
- Gray, G. M., Li, P., Shlyakhter, I. and Wilson, R. (1995). An empirical examination of factors influencing prediction of carcinogenic hazard across species. *Regul Toxicol Pharmacol* 22, 283-291. doi:10.1006/rtph.1995.0011
- Griesinger, C., Hoffmann, S., Kinsner-Ovaskainen, A. et al. (2008). Proceedings of the First International Forum Towards Evidence-Based Toxicology. Conference Centre Spazio Villa Erba, Como, Italy. 15-18 October 2007. *Human Exp Toxicol* 28, *Spec Issue: Evidence-Based Toxicology (EBT) 2009*, 71-163.
- Hardy, B., Apic, G., Carthew, P. et al. (2012a). Food for thought ... A toxicology ontology roadmap. *ALTEX* 29, 129-137. doi:10.14573/altex.2012.2.129
- Hardy, B., Apic, G., Carthew, P. et al. (2012b). Toxicology ontology perspectives. *ALTEX* 29, 139-156. doi:10.14573/altex.2012.2.139
- Hareng, L., Pellizzer, C., Bremer, S. et al. (2005). The integrated project ReProTect: A novel approach in reproductive toxicity hazard assessment. *Reprod Toxicol* 20, 441-452. doi:10.1016/j.reprotox.2005.04.003
- Hartung, T., Bremer, S., Casati, S. et al. (2004). A modular approach to the ECVAM principles on test validity. *Altern Lab Anim* 32, 467-472.
- Hartung, T. (2007). Food for thought ... on validation. *ALTEX* 24, 67-72. doi:10.14573/altex.2007.2.67
- Hartung, T. (2008a). Food for thought ... on animal tests. *ALTEX* 25, 3-9. doi:10.14573/altex.2008.1.3
- Hartung, T. (2008b). Toward a new toxicology – Evolution or revolution? *Altern Lab Anim* 36, 635-639.
- Hartung, T. (2008c). Food for thought ... on alternative methods for cosmetics safety testing. *ALTEX* 25, 147-162. doi:10.14573/altex.2008.3.147
- Hartung, T. (2009a). Food for thought ... on evidence-based toxicology. *ALTEX* 26, 75-82. doi:10.14573/altex.2009.2.75
- Hartung, T. (2009b). Toxicology for the twenty-first century. *Nature* 460, 208-212. doi:10.1038/460208a
- Hartung, T. and Rovida, C. (2009). Chemical regulators have overreached. *Nature* 460, 1080-1081. doi:10.1038/4601080a
- Hartung, T. and Hoffmann, S. (2009). Food for thought ... on in silico methods in toxicology. *ALTEX* 26, 155-166. doi:10.14573/altex.2009.3.155
- Hartung, T. (2010a). Evidence based-toxicology – The toolbox of validation for the 21st century? *ALTEX* 27, 241-251. doi:10.14573/altex.2010.4.253



- Hartung, T. (2010b). Food for thought... on alternative methods for chemical safety testing. *ALTEX* 27, 3-14. doi:10.14573/altex.2010.1.3
- Hartung, T., Blaauw, G. J., Bosgra, S. et al. (2011). An expert consortium review of the EC-commissioned report "Alternative (non-animal) methods for cosmetics testing: current status and future prospects – 2010". *ALTEX* 28, 183-209. doi:10.14573/altex.2011.3.183
- Hartung, T., van Vliet, E., Jaworska, J. et al. (2012). Systems toxicology. *ALTEX* 29, 119-128. doi:10.14573/altex.2012.2.119
- Hartung, T. (2013). Look back in anger – What clinical studies tell us about preclinical work. *ALTEX* 30, 275-291. doi:10.14573/altex.2013.3.275
- Hartung, T., Luechtefeld, T., Maertens, A. and Kleensang, A. (2013). Integrated testing strategies for safety assessments. *ALTEX* 30, 3-18. doi:10.14573/altex.2013.1.003
- Hartung, T. (2014). 3D – A new dimension of in vitro research. *Adv Drug Deliv Rev* 69-70, vi. Preface Special Issue "Innovative tissue models for in vitro drug development".
- Hartung, T. (2016). Making big sense from big data in toxicology by read-across. *ALTEX* 33, 83-93. doi:10.14573/altex.1603091
- Hartung, T. (2017a). Food for thought ... the first ten years. *ALTEX* 34, 187-192. doi:10.14573/altex.1703311
- Hartung, T. (2017b). Opinion versus evidence for the need to move away from animal testing. *ALTEX* 34, 193-200. doi:10.14573/altex.1703291
- Hartung, T. (2017c). Thresholds of toxicological concern – Setting a threshold for testing where there is little concern. *ALTEX* 34, 331-351. doi:10.14573/altex.1707011
- Hartung, T. (2017d). Utility of the adverse outcome pathway concept in drug development. *Exp Opin Drug Metabol Toxicol* 13, 1-3. doi:10.1080/17425255.2017.1246535
- Hartung, T., FitzGerald, R., Jennings, P. et al. (2017a). Systems toxicology – Real world applications and opportunities. *Chem Res Toxicol* 30, 870-882. doi:10.1021/acs.chemrestox.7b00003
- Hartung, T., Kavlock, R. and Sturla, S. (2017b). Systems toxicology II: A special issue. *Chem Res Toxicol* 30, 869-869. doi:10.1021/acs.chemrestox.7b00038
- Hartung, T. (2018). Perspectives on in vitro to in vivo extrapolations. *J Appl In Vitro Toxicol*, in press. doi:10.1089/aivt.2016.0026
- Haseman, J. K., Boorman, G. A. and Huff, J. (1997). Value of historical control data and other issues related to the evaluation of long-term rodent carcinogenicity studies. *Toxicol Pathol* 25, 524-527. doi:10.1177/019262339702500518
- Hatzivassiliou, G., Song, K., Yen, I. et al. (2010). RAF inhibitors prime wild-type RAF to activate the MAPK pathway and enhance growth. *Nature* 464, 431-435. doi:10.1038/nature08833
- Hengstler, J. G., van der Burg, B., Steinberg, P. and Oesch, F. (1999). Interspecies differences in cancer susceptibility and toxicity. *Drug Metab Rev* 31, 917-970. doi:10.1081/DMR-100101946
- Hengstler, J. G., Marchan, R. and Leist, M. (2012). Highlight report: Towards the replacement of in vivo repeated dose systemic toxicity testing. *Arch Toxicol* 86, 13-15. doi:10.1007/s00204-011-0798-7
- Herwig, R., Gmuender, H., Corvi, R. et al. (2016). Inter-laboratory study of human in vitro toxicogenomics-based tests as alternative methods for evaluating chemical carcinogenicity: A bioinformatics perspective. *Arch Toxicol* 90, 2215-2229. doi:10.1007/s00204-015-1617-3
- Hoffmann, S. and Hartung, T. (2006). Towards an evidence-based toxicology. *Hum Exp Toxicol* 25, 497-513. doi:10.1191/0960327106het648oa
- Hoffmann, S., Edler, L., Gardner, I. et al. (2008). Points of reference in validation – The report and recommendations of ECVAM Workshop. *Altern Lab Anim* 36, 343-352.
- Hoffmann, S., de Vries, R. B. M., Stephens, M. L. et al. (2017). A primer on systematic reviews in toxicology. *Arch Toxicol* 91, 2551-2575. doi:10.1007/s00204-017-1980-3
- Hoke, R. A. and Ankley, G. T. (2005). Application of frog embryo teratogenesis assay-Xenopus to ecological risk assessment. *Environ Toxicol Chem* 24, 2677-2690. doi:10.1897/04-506R.1
- Hotchkiss, A. K., Rider, C. V., Blystone, C. R. et al. (2008). Fifteen years after "Wingspread" – Environmental endocrine disruptors and human and wildlife health: Where we are today and where we need to go. *Toxicol Sci* 105, 235-259. doi:10.1093/toxsci/kfn030
- Hurt, M. E., Cappon, G. D. and Browning, A. (2003). Proposal for a tiered approach to developmental toxicity testing for veterinary pharmaceutical products for food-producing animals. *Food Chem Toxicol* 41, 611-619. doi:10.1016/S0278-6915(02)00326-5
- Hutson, M. S., Leung, M., Baker, N. C. et al. (2017). Computational model of secondary palate fusion and disruption. *Chem Res Toxicol* 30, 965-979. doi:10.1021/acs.chemrestox.6b00350
- ICCVAM (2018). A Strategic Roadmap for Establishing New Approaches to Evaluate the Safety of Chemicals and Medical Products in the United States. January 2018. <https://ntp.niehs.nih.gov/go/natl-strategy> (accessed 22.03.2018).
- ICH (2009). S1B Guideline on Carcinogenicity Testing of Pharmaceuticals. EMA Document CPMP/ICH/299/95. <https://bit.ly/2H3PTw>
- ICH Status Report (2016). The ICHS1 Regulatory Testing Paradigm of Carcinogenicity in rats – Status Report. Safety Guidelines, 2 March, 1-5. <https://bit.ly/2IzczwZ>
- ILSI HESI ACT (2001). ILSI HESI Alternatives to carcinogenicity testing project. *Toxicol Pathol* 29, Suppl, 1-351.
- Jacobs, M. N., Colacci, A., Louekari, K. et al. (2016). International regulatory needs for development of an IATA for non-genotoxic carcinogenic chemical substances. *ALTEX* 33, 359-392. doi:10.14573/altex.1601201
- Janer, G., Hakkert, B. C., Slob, W. et al. (2007). A retrospective analysis of the two-generation study: What is the added value of the second generation? *Reprod Toxicol* 24, 97-102. doi:10.1016/j.reprotox.2007.04.068
- Jemal, A., Siegel, R., Ward, E. et al. (2009). Cancer statistics, 2009. *CA Cancer J Clin* 59, 225-249. doi:10.3322/caac.20006
- Jilka, R. L. (2016). The road to reproducibility in animal research. *J Bone Mineral Res* 31, 1317-1319. doi:10.1002/jbmr.2881
- Johnson, F. M. (2001). Response to Tennant et al.: Attempts to replace the NTP rodent bioassay with transgenic alternatives are unlikely



- to succeed. *Environ Mol Mutagen* 37, 89-92. doi:10.1002/1098-2280(2001)37:1<89::AID-EM1011>3.0.CO;2-4
- Joyner, M. J. and Pedersen, B. K. (2011). Ten questions about systems biology. *J Physiol* 589, 1017-1030. doi:10.1113/jphysiol.2010.201509
- Judson, R., Houck, K., Martin, M. et al. (2016). Analysis of the effects of cell stress and cytotoxicity on in vitro assay activity across a diverse chemical and assay space. *Toxicol Sci* 152, 323-339. doi:10.1093/toxsci/kfw092
- Kennedy, T. (1997). Managing the drug discovery/development interface. *Drug Discov Today* 2, 436-444. doi:10.1016/S1359-6446(97)01099-4
- Kessler, R. (2014). Air of danger. *Nature* S62, 509. doi:10.1038/509S62a
- Kholodenko, B. N. (2015). Drug resistance resulting from kinase dimerization is rationalized by thermodynamic factors describing allosteric inhibitor effects. *Cell Reports* 12, 1939-1949. doi:10.1016/j.celrep.2015.08.014
- Kiani, N. A., Shang, M.-M. and Tegner, J. (2016). Systems toxicology: Systematic approach to predict toxicity. *Curr Pharm Des* 22, 6911-6917. doi:10.2174/1381612822666161003115629
- Kilic, O., Pamies, D., Lavell, E. et al. (2016). Microphysiological brain model enables analysis of neuronal differentiation and chemotaxis. *Lab Chip* 16, 4152-4162. doi:10.1039/C6LC00946H
- Kim, J. H. and Scialli, A. R. (2011). Thalidomide: The tragedy of birth defects and the effective treatment of disease. *Toxicol Sci* 122, 1-6. doi:10.1093/toxsci/kfr088
- Kirkland, D., Aardema, M., Henderson, L. and Müller, L. (2005). Evaluation of the ability of a battery of three in vitro genotoxicity tests to discriminate rodent carcinogens and non-carcinogens I. Sensitivity, specificity and relative predictivity. *Mutat Res* 584, 1-256. doi:10.1016/j.mrgentox.2005.02.004
- Kirkland, D., Pfuhrer, S., Tweatsm, D. et al. (2007). How to reduce false positive results when undertaking in vitro genotoxicity testing and thus avoid unnecessary follow up animal tests: Report of an ECVAM Workshop. *Mutat Res* 628, 31-55. doi:10.1016/j.mrgentox.2006.11.008
- Kirkland, D., Reeve, L., Gatehouse, D. and Vanparys, P. (2011). A core in vitro genotoxicity battery comprising the Ames test plus the in vitro micronucleus test is sufficient to detect rodent carcinogens and in vivo genotoxins. *Mutat Res* 721, 27-73. doi:10.1016/j.mrgentox.2010.12.015
- Kleinstreuer, N., Dix, D., Houck, K. et al. (2013a). In vitro perturbations of targets in cancer hallmark processes predict rodent chemical carcinogenesis. *Toxicol Sci* 131, 40-55. doi:10.1093/toxsci/kfs285
- Kleinstreuer, N., Dix, D., Rountree, M. et al. (2013b). A computational model predicting disruption of blood vessel development. *PLoS Comput Biol* 9, e1002996. doi:10.1371/journal.pcbi.1002996
- Kleinstreuer, N., Houck, K., Yang, J. et al. (2014). Phenotypic screening of the ToxCast chemical library to classify toxic and therapeutic mechanisms. *Nat Biotechnol* 32, 583-591. doi:10.1038/nbt.2914
- Kleinstreuer, N. C., Ceger, P., Watt, E. D. et al. (2016). Development and validation of a computational model for androgen receptor activity. *Chem Res Toxicol* 30, 946-964. doi:10.1021/acs.chemrestox.6b00347
- Kleinstreuer, N., Hoffmann, S., Alepee, N. et al. (2018). Non-animal methods to predict skin sensitization (II): An assessment of defined approaches. *Crit Rev Toxicol* 23, 1-16. doi:10.1080/10408444.2018.1429386
- Knight, A., Bailey, J. and Balcombe, J. (2006a). Animal carcinogenicity studies: 2. Obstacles to extrapolation of data to humans. *Altern Lab Anim* 34, 29-38.
- Knight, A., Bailey, J. and Balcombe, J. (2006b). Animal carcinogenicity studies: 1. Poor human predictivity. *Altern Lab Anim* 34, 19-27.
- Knudsen, T. B., Kavlock, R. J., Daston, G. P. et al. (2011). Developmental toxicity testing for safety assessment: New approaches and technologies. *Birth Defects Res B Dev Reprod Toxicol* 92, 413-420. doi:10.1002/bdrb.20315
- Knudsen, T. B. and Kleinstreuer, N. C. (2012). Disruption of embryonic vascular development in predictive toxicology. *Birth Defects Res C Embryo Today* 93, 312-323. doi:10.1002/bdrc.20223
- Kola, I. and Landis, J. (2004). Can the pharmaceutical industry reduce attrition rates? *Nat Rev Drug Discov* 3, 711-715. doi:10.1038/nrd1470
- Kubinyi, H. (2003). Drug research: Myths, hype and reality. *Nat Rev Drug Discov* 2, 665-668. doi:10.1038/nrd1156
- Laufersweiler, M. C., Gadagbui, B., Baskerville-Abraham, I. M. et al. (2012). Correlation of chemical structure with reproductive and developmental toxicity as it relates to the use of the threshold of toxicological concern. *Regulat Toxicol Pharmacol* 62, 160-182. doi:10.1016/j.yrtph.2011.09.004
- Lave, L. B., Ennever, F. K., Rosenkranz, H. S. and Omenn, G. S. (1988). Information value of the rodent bioassay. *Nature* 336, 631-633. doi:10.1038/336631a0
- LeBoeuf, R. A., Kerckaert, K. A., Aardema, M. J. and Isfort, R. J. (1999). Use of Syrian hamster embryo and BALB/c 3T3 cell transformation for assessing the carcinogenic potential of chemicals. *IARC Sci Publ* 146, 409-425.
- Lee, M. J., Ye, A. S., Gardino, A. K. et al. (2012). Sequential application of anticancer drugs enhances cell death by rewiring apoptotic signaling networks. *Cell* 149, 780-794. doi:10.1016/j.cell.2012.03.031
- Leist, M., Bremer, S., Brundin, P. et al. (2008). The biological and ethical basis of the use of human embryonic stem cells for in vitro test systems or cell therapy. *ALTEX* 25, 163-190. doi:10.14573/altex.2008.3.163
- Leist, M., Hasiwa, M., Daneshian, M. and Hartung, T. (2012). Validation and quality control of replacement alternatives – Current status and future challenges. *Toxicol Res* 1, 8. doi:10.1039/c2tx20011b
- Leist, M., Hasiwa, N., Rovida, C. et al. (2014). Consensus report on the future of animal-free systemic toxicity testing. *ALTEX* 31, 341-356. doi:10.14573/altex.1406091
- Leist, M., Ghallab, A., Graepel, R. et al. (2017). Adverse outcome pathways: Opportunities, limitations and open questions. *Arch Toxicol* 91, 3477-3505. doi:10.1007/s00204-017-2045-3



- Leung, M. C., Phuong, J., Baker, N. C. et al. (2016). Systems toxicology of male reproductive development: Profiling 774 chemicals for molecular targets and adverse outcomes. *Environ Health Perspect* 124, 1050-1061. doi:10.1289/ehp.1510385
- Luechtefeld, T., Maertens, A., Russo, D. P. et al. (2016a). Global analysis of publicly available safety data for 9,801 substances registered under REACH from 2008-2014. *ALTEX* 33, 95-109. doi:10.14573/altex.1510052
- Luechtefeld, T., Maertens, A., Russo, D. P. et al. (2016b). Analysis of public oral toxicity data from REACH registrations 2008-2014. *ALTEX* 33, 111-122. doi:10.14573/altex.1510054
- Luechtefeld, T. and Hartung, T. (2017). Computational approaches to chemical hazard assessment. *ALTEX* 34, 459-478. doi:10.14573/altex.1710141
- MacDonald, J., French, J. E., Gerson, R. J. et al. (2004). The utility of genetically modified mouse assays for identifying human carcinogens: A basic understanding and path forward. *Toxicol Sci* 77, 188-194. doi:10.1093/toxsci/kfh037
- Madia, F., Worth, A. and Corvi, R. (2016). Analysis of carcinogenicity testing for regulatory purposes in the European Union (92 pp). *JRC Report EUR 27765*.
- Mandrioli, D. and Silbergeld, E. K. (2016). Evidence from toxicology: The most essential science for prevention. *Environ Health Perspect* 124, 6-11. doi:10.1289/ehp.1509880
- Martin, M. T., Judson, R. S., Reif, D. M. et al. (2009a). Profiling chemicals based on chronic toxicity results from the U.S. EPA ToxRef database. *Environ Health Perspect* 117, 392-399. doi:10.1289/ehp.0800074
- Martin, M. T., Mendez, E., Corum, D. G. et al. (2009b). Profiling the reproductive toxicity of chemicals from multigeneration studies in the toxicity reference database. *Toxicol Sci* 110, 181-190. doi:10.1093/toxsci/kfp080
- Marx, J. (2003). Building better mouse models for studying cancer. *Science* 299, 1972-1975. doi:10.1126/science.299.5615.1972
- Marx, U., Andersson, T. B., Bahinski, A. et al. (2016). Biology-inspired microphysiological system approaches to solve the prediction dilemma of substance testing using animals. *ALTEX* 33, 272-321. doi:10.14573/altex.1603161
- Marx-Stoelting, P., Adriaens, E., Ahr, H. J. et al. (2009). A review of the implementation of the embryonic stem cell test (EST). The report and recommendations of an ECVAM/ReProTect Workshop. *Altern Lab Anim* 37, 313-328.
- Mattison, D. R. (2010). Environmental exposures and development. *Curr Opin Pediatrics* 22, 208-218. doi:10.1097/MOP.0b013e32833779bf
- McCollum, C. W., de Vancells, J. C., Hans, C. et al. (2016). Identification of vascular disruptor compounds by analysis in zebrafish embryos and mouse embryonic endothelial cells. *Reprod Toxicol* 70, 60-69. doi:10.1016/j.reprotox.2016.11.005
- Miller, G. W. (2014). Improving reproducibility in toxicology. *Toxicol Sci* 139, 1-3. doi:10.1093/toxsci/kfu050
- Moore, N., Bremer, S., Carmichael, N. et al. (2009). A modular approach to the extended one-generation reproduction toxicity study: Outcome of an ECETOC task force and International ECETOC/ECVAM workshop. *Altern Lab Anim* 37, 219-225.
- Morgan, R. L., Thayer, K. A., Bero, L. et al. (2016). GRADE: Assessing the quality of evidence in environmental and occupational health. *Environ Int* 92-93, 611-616. doi:10.1016/j.envint.2016.01.004
- OECD (2005). Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment. Series on Testing and Assessment No. 34, NV/JM/MONO(2005)14. Paris: OECD Publishing. [http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?docLanguage=en&cote=env/jm/mono\(2005\)14](http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?docLanguage=en&cote=env/jm/mono(2005)14)
- OECD (2007). Detailed Review Paper on Cell Transformation Assays for Detection of Chemical Carcinogens. Series on Testing and Assessment No. 31. Paris: OECD Publishing. <http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=ENV/JM/MONO%282007%2918&docLanguage=En>
- OECD (2009). Carcinogenicity Studies. OECD Guidelines for Chemical Testing, TG No. 451. <http://www.oecd.org/dataoecd/30/46/41753121.pdf>
- OECD (2011). OECD Guideline on Extended One-Generation Reproductive Toxicity Study. Series on Testing and Assessment No. 443. Paris: OECD Publishing. doi:10.1787/9789264122550-en
- OECD (2015). Guidance Document on the in vitro Syrian Hamster Embryo (SHE) Cell Transformation Assay, Series on Testing and Assessment, No. 214. Paris: OECD Publishing. <http://www.oecd.org/env/ehs/testing/Guidance-Document-on-the-in-vitro-Syrian-Hamster-Embryo-Cell-Transformation-Assay.pdf>
- OECD (2016). Guidance Document on the in vitro Bhas 42 Cell Transformation Assay. Series on Testing and Assessment, No. 231. Paris: OECD Publishing. [https://www.oecd.org/env/ehs/testing/ENV_JM_MONO\(2016\)1.pdf](https://www.oecd.org/env/ehs/testing/ENV_JM_MONO(2016)1.pdf)
- Pamies, D. and Hartung, T. (2017). 21st century cell culture for 21st century toxicology. *Chem Res Toxicol* 30, 43-52. doi:10.1021/acs.chemrestox.6b00269
- Pamies, D., Barreras, P., Block, K. et al. (2017a). A human brain microphysiological system derived from iPSC to study central nervous system toxicity and disease. *ALTEX* 34, 362-376. doi:10.14573/altex.1609122
- Pamies, D., Bal-Price, A., Simeonov, A. et al. (2017b). Good cell culture practice for stem cells and stem-cell-derived models. *ALTEX* 34, 95-132. doi:10.14573/altex.1607121
- Pamies, D., Block, K., Lau, P. et al. (2018a). Rotenone exerts developmental neurotoxicity in a human brain spheroid model. *Toxicol Appl Pharmacol*, Epub ahead of print. doi:10.1016/j.taap.2018.02.003
- Pamies, D., Bal-Price, A., Chesne, C. et al. (2018b). Advanced good cell culture practice for human primary, stem cell-derived and organoid models as well as microphysiological systems. *ALTEX*, Epub ahead of print. doi:10.14573/altex.1710081.
- Paparella, M., Daneshian, M., Hornek-Gausterer, R. et al. (2013). Uncertainty of testing methods – What do we (want to) know? *ALTEX* 30, 131-144. doi:10.14573/altex.2013.2.131
- Paparella, M., Colacci, A. and Jacobs, M. N. (2016). Uncertainties of testing methods: What do we (want to) know about carcinogenicity? *ALTEX* 34, 235-252. doi:10.14573/altex.1608281
- Paules, R. S., Aubrecht, J., Corvi, R. et al. (2011). Moving forward in human cancer risk assessment. *Environ Health Persp* 119, 739-743. doi:10.1289/ehp.1002735

- Piersma, A. H., Genschow, E., Verhoef, A. et al. (2004). Validation of the postimplantation rat whole-embryo culture test in the international ECVAM validation study on three in vitro embryotoxicity tests. *Altern Lab Anim* 32, 275-307.
- President's Cancer Panel (2010). Reducing environmental cancer risk, NIH. https://deainfo.nci.nih.gov/advisory/pcp/annual_reports/pcp08-09rpt/pcp_report_08-09_508.pdf
- Pound, P. and Bracken, M. B. (2014). Is animal research sufficiently evidence based to be a cornerstone of biomedical research? *BMJ* 348, g3387. doi:10.1136/bmj.g3387
- Pritchard, J. B., French, J. E., Davis, B. J. and Haseman, J. K. (2003). The role of transgenic mouse models in carcinogen identification. *Environ Health Perspect* 111, 444-454. doi:10.1289/ehp.5778
- Rall, D. P. (2000). Laboratory animal tests and human cancer. *Drug Metab Rev* 32, 119-128. doi:10.1081/DMR-100100565
- Rodgers, K. M., Udesky, J. O., Rudel, R. A. and Brody, J. G. (2018). Environmental chemicals and breast cancer: An updated review of epidemiological literature informed by biological mechanisms. *Environ Res* 160, 152-182. doi:10.1016/j.envres.2017.08.045
- Rovida, C., Longo, F. and Rabbit, R. R. (2011). How are reproductive toxicity and developmental toxicity addressed in REACH dossiers? *ALTEX* 28, 273-294. doi:10.14573/altex.2011.4.273
- Rovida, C., Alépée, N., Api, A. M. et al. (2015). Integrated testing strategies (ITS) for safety assessment. *ALTEX* 32, 171-181. doi:10.14573/altex.1411011
- Sanz, F., Pognan, F., Steger-Hartmann, T. et al. (2017). Legacy data sharing to improve drug safety assessment: The eTOX project. *Nat Rev Drug Discov* 16, 811-812. doi:10.1038/nrd.2017.177
- Sauer, J. M., Hartung, T., Leist, M. et al. (2015). Systems toxicology: The future of risk assessment. *Int J Toxicol* 34, 346-348. doi:10.1177/1091581815576551
- Schaap, M. M., Wackers, P. F., Zwart, E. P. et al. (2015). A novel toxicogenomics-based approach to categorize (non-)genotoxic carcinogens. *Arch Toxicol* 89, 2413-2427. doi:10.1007/s00204-014-1368-6
- Schmidt, C. W. (2002). Assessing assays. *Environ Health Perspect* 110, A248-251. doi:10.1289/ehp.110-a248
- Schwarzman, M. R., Ackerman, J. R., Dairkee, S. H. et al. (2015). Screening for chemical contributions to breast cancer risk: A case study for chemical safety evaluation. *Environ Health Perspect* 123, 1255-1264. doi:10.1289/ehp.1408337
- Seidle, T. (2006). Chemicals and cancer: What the regulators won't tell you about carcinogenicity testing. PETA Europe Ltd. <https://www.peta.de/mediadb/EUreport300.pdf>
- Selderslaghs, I. W. T., Blust, R. and Witters, H. E. (2012). Feasibility study of the zebrafish assay as an alternative method to screen for developmental toxicity and embryotoxicity using a training set of 27 compounds. *Reprod Toxicol* 33, 142-154. doi:10.1016/j.reprotox.2011.08.003
- Silbergeld, E. K. (2004). Commentary: The role of toxicology in prevention and precaution. *Int J Occup Med Environ Health* 17, 91-102.
- Sistare, F. D., Morton, D., Alden, C. et al. (2011). An analysis of pharmaceutical experience with decades of rat carcinogenicity testing: Support for a proposal to modify current regulatory guidelines. *Toxicol Pathol* 9, 716-744. doi:10.1177/0192623311406935
- Skardal, A., Shupe, T. and Atala, A. (2016). Organoid-on-a-chip and body-on-a-chip systems for drug screening and disease modeling. *Drug Discov Today* 21, 1399-1411. doi:10.1016/j.drudis.2016.07.003
- Skardal, A., Murphy, S., Devarasetty, M. et al. (2017). Multi-tissue interactions in an integrated three-tissue organ-on-a-chip platform. *Sci Rep* 7, 8837. doi:10.1038/s41598-017-08879-x
- Slikker, W. (2014). Of human-on-a-chip and humans: Considerations for creating and using microphysiological systems. *Exp Biol Med* 239, 1078-1079. doi:10.1177/1535370214537754
- Smirnova, L., Harris, G., Leist, M. and Hartung, T. (2015). Cellular resilience. *ALTEX* 32, 247-260. doi:10.14573/altex.1509271
- Smith, M. T., Guyton, K. Z., Gibbons, C. F. et al. (2016). Key characteristics of carcinogens as a basis for organizing data on mechanisms of carcinogenesis. *Environ Health Perspect* 124, 713-721. doi:10.1289/ehp.1509912
- Spielmann, H., Genschow, E., Brown, N. A. et al. (2004). Validation of the rat limb bud micromass test in the international ECVAM validation study on three in vitro embryotoxicity tests. *Altern Lab Anim* 32, 245-274.
- Spielmann, H., Seiler, A., Bremer, S. et al. (2006). The practical application of three validated in vitro embryotoxicity tests. The report and recommendations of an ECVAM/ZEBET workshop (ECVAM workshop 57). *Altern Lab Anim* 34, 527-538.
- Stephens, M. L., Andersen, M., Becker, R. A. et al. (2013). Evidence-based toxicology for the 21st century: Opportunities and challenges. *ALTEX* 30, 74-103. doi:10.14573/altex.2013.1.074
- Stephens, M. L., Betts, K., Beck, N. B. et al. (2016). The emergence of systematic review in toxicology. *Toxicol Sci* 152, 10-16. doi:10.1093/toxsci/kfw059
- Sturla, S. J., Boobis, A. R., FitzGerald, R. E. et al. (2014). Systems toxicology: From basic research to risk assessment. *Chem Res Toxicol* 27, 314-329. doi:10.1021/tx400410s
- Sukardi, H., Chang, H. T., Chan, E. C. et al. (2011). Zebrafish for drug toxicity screening: Bridging the in vitro cell-based models and in vivo mammalian models. *Expert Opin Drug Metab Toxicol* 7, 579-589. doi:10.1517/17425255.2011.562197
- Suter-Dick, L., Alves, P. M., Blauboer, B. J. et al. (2015). Stem cell-derived systems in toxicology assessment. *Stem Cells Develop* 24, 1284-1296. doi:10.1089/scd.2014.0540
- Takayama, S., Thorgeirsson, U. P. and Adamson, R. H. (2008). Chemical carcinogenesis studies in nonhuman primates. *Proc Jpn Acad Ser B Phys Biol Sci* 84, 176-188. doi:10.2183/pjab.84.176
- Tal, T., Kilty, C., Smith, A. et al. (2016). Screening for angiogenic inhibitors in zebrafish to evaluate a predictive model for developmental vascular toxicity. *Reprod Toxicol* 70, 70-81. doi:10.1016/j.reprotox.2016.12.004
- Tennant, R. W., Stasiewicz, S., Mennear, J. et al. (1999). Genetically altered mouse models for identifying carcinogens. *IARC Sci Publ* 146, 123-150.
- Thilly, W. G. (2003). Have environmental mutagens caused on-



- comutations in people? *Nat Genet* 34, 255-259. doi:10.1038/ng1205
- Tsaïoun, K., Blaauboer, B. J. and Hartung, T. (2016). Evidence-based absorption, distribution, metabolism, excretion and toxicity (ADMET) and the role of alternative methods. *ALTEX* 33, 343-358. doi:10.14573/altex.1610101
- Vaccari, M., Mascolo, M.G., Rotondo, F. et al. (2015). Identification of pathway-based toxicity in the BALB/c 3T3 cell model. *Toxicol In Vitro* 29, 1240-1253. doi:10.1016/j.tiv.2014.10.002
- van der Laan, J. W., Kasper, P., Silva Lima, B. et al. (2016). Critical analysis of carcinogenicity study outcomes. Relationship with pharmacological properties. *Crit Rev Toxicol* 46, 587-614. doi:10.3109/10408444.2016.1163664
- Van Oosterhout, J. P., Van der Laan, J. W., De Waal, E. J. et al. (1997). The utility of two rodent species in carcinogenic risk assessment of pharmaceuticals in Europe. *Regul Toxicol Pharmacol* 25, 6-17. doi:10.1006/rtph.1996.1077
- van Ravenzwaay, B. (2010). Initiatives to decrease redundancy in animal testing of pesticides. *ALTEX* 27, 159-161.
- van Ravenzwaay, B., Dammann, M., Buesen, R. et al. (2011). The threshold of toxicological concern for prenatal developmental toxicity. *Regul Toxicol Pharmacol* 59, 81-90. doi:10.1016/j.yrtph.2010.09.009
- van Ravenzwaay, B., Dammann, M., Buesen, R. et al. (2012). The threshold of toxicological concern for prenatal developmental toxicity in rabbits and a comparison to TTC values in rats. *Regul Toxicol Pharmacol* 64, 1-8. doi:10.1016/j.yrtph.2012.06.004
- van Ravenzwaay, B., Jiang, X., Luechtefeld, T. and Hartung, T. (2017). The threshold of toxicological concern for prenatal developmental toxicity in rats and rabbits. *Regul Toxicol Pharmacol* 88, 157-172. doi:10.1016/j.yrtph.2017.06.008
- Voelkl, B., Vogt, L., Sena, E. S. and Würbel, H. (2018). Reproducibility of preclinical animal research improves with heterogeneity of study samples. *PLoS Biology* 16, e2003693. doi:10.1371/journal.pbio.2003693
- Wang, B. and Gray, G. (2015). Concordance of Noncarcinogenic Endpoints in Rodent Chemical Bioassays. *Risk Analysis* 35, 1154-1166. doi:10.1111/risa.12314
- Waters, M. D. (2016). Introduction to predictive carcinogenicity. In *Issues in Toxicology No. 28 Toxicogenomics in Predictive Carcinogenicity* (Waters MD and Thomas RS, eds). Cambridge, UK: Royal Society of Chemistry. doi:10.1039/9781782624059-00001
- Watson, D. E., Hunziker, R. and Wikswo, J. P. (2017). Fitting tissue chips and microphysiological systems into the grand scheme of medicine, biology, pharmacology, and toxicology. *Exp Biol Med* 242, 1559-1572. doi:10.1177/1535370217732765
- Weigt, S., Huebler, N., Braunbeck, T. et al. (2010). Zebrafish teratogenicity test with metabolic activation (mDarT): Effects of phase I activation of acetaminophen on zebrafish *Danio rerio* embryos. *Toxicology* 275, 36-49. doi:10.1016/j.tox.2010.05.012
- Worth, A., Barroso, J. F., Bremer, S. et al. (2014). Alternative Methods for Regulatory Toxicology – A State-of-the-art Review, 470pp. *JRC Report 26797*. <https://bit.ly/2q91DiM>

Conflict of interest

The authors declare the following competing financial interest(s): T.H. is the founder of Organome LLC, Baltimore, and consults AstraZeneca, Cambridge, UK, in the field of organo-typic cultures / MPS. A. L. is a co-founder of Sidera Medicine. The opinions expressed in this article are not informed by this affiliation.

Acknowledgements

This work was supported by the EU-ToxRisk project (An Integrated European “Flagship” Program Driving Mechanism-Based Toxicity Testing and Risk Assessment for the 21st Century) funded by the European Commission under the Horizon 2020 program (Grant Agreement No. 681002). The work on human BrainSpheres mentioned was supported by NIH NCATS (grant U18TR000547 “A 3D Model of Human Brain Development for Studying Gene/Environment Interactions”, PI Hartung) and Alternatives Research & Development Foundation (“A 3D *in vitro* ‘mini-brain’ model to study Parkinson’s disease”, PI Hartung). Andre Levchenko is the PI of a U54 NCI Cancer Systems Biology grant CA209992.

Correspondence to

Thomas Hartung, MD PhD
Center for Alternatives to Animal Testing
Johns Hopkins Bloomberg School of Public Health
615 North Wolfe Street
W7032, Baltimore, MD 21205, USA
e-mail: thartun1@jhu.edu