FutureTox III: Bridges for Translation

Daland R. Juberg,[a,1] Thomas B. Knudsen,[b] Miriam Sander,[c] Nancy B. Beck,[d] Elaine M. Faustman,[e] Donna L. Mendrick,[f] John R. Fowle III,[g] Thomas Hartung,[h] Raymond R. Tice,[i] Emmanuel Lemazurier,[j] Richard A. Becker,[d] Suzanne Compton Fitzpatrick,[k] George P. Daston,[l] Alison Harrill,[m] Ronald N. Hines,[b] Douglas A. Keller,[n] John C. Lipscomb,[o] David Watson,[p] Tina Bahadori,[q] and Kevin M. Crofton[b]

[a]Dow AgroSciences, Indianapolis, Indiana; [b]US Environmental Protection Agency, Research Triangle Park, North Carolina; [c]Page One Editorial Services, Boulder, Colorado; [d]American Chemistry Council, Washington, The District of Columbia; [e]University of Washington, Seattle, Washington; [f]US Food and Drug Administration, Silver Spring, Maryland; [g]Science to Inform, LLC, Pittsboro, North Carolina; [h]Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland; [i]National Toxicology Program/National Institute of Environmental Health Sciences, Durham, North Carolina; [j]INERIS-Chronic Risk Division, Verneeuil-en-Halatte, France; [k]US Food and Drug Administration, College Park, Maryland; [l]Procter & Gamble Company, Cincinnati, Ohio; [m]University of Arkansas for Medical Sciences, Little Rock, Arkansas; [n]Sanofi, Bridgewater, New Jersey; [o]US Environmental Protection Agency, Cincinnati, Ohio; [p]Lhasa Limited, Leeds, United Kingdom; and [q]US Environmental Protection Agency, Washington, The District of Columbia

[1]To whom correspondence should be addressed. E-mail: drjuberg@dow.com.

Disclaimer: The views expressed in this article are those of the authors and do not necessarily reflect the views or policies of the U.S. Environmental Protection Agency, the National Institute of Environmental Health Sciences, the National Institutes of Health, the U.S. Food and Drug Administration, or any other U.S. federal institute or agency. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

ABSTRACT

Future Tox III, a Society of Toxicology Contemporary Concepts in Toxicology workshop, was held in November 2015. Building upon Future Tox I and II, Future Tox III was focused on developing the high throughput risk assessment paradigm and taking the science of *in vitro* data and *in silico* models forward to explore the question—what progress is being made to address challenges in implementing the emerging big-data toolbox for risk assessment and regulatory decision-making. This article reports on the outcome of the workshop including 2 examples of where advancements in predictive toxicology approaches are being applied within Federal agencies, where opportunities remain within the exposome and AOP domains, and how collectively the toxicology community across multiple sectors can continue to bridge the translation from historical approaches to Tox21 implementation relative to risk assessment and regulatory decision-making.

Key words: predictive toxicology; *in vitro* and alternatives; regulatory/policy; risk assessment; testing alternatives.

In 2007 and 2009, the National Research Council published *Toxicity Testing in The Twenty-first Century: a Vision and a Strategy* (NRC, 2007) and *Science and Decisions: Advancing Risk Assessment* (NRC, 2009), respectively, 2 seminal reports on the state-of-the-science in the fields of toxicology and human risk assessment. These 2 documents were commissioned to advance progress in embracing and implementing novel emerging methods for toxicity testing and risk assessment, so that potential toxicants,

and their adverse effects on human health and the environment, would be identified in a more efficient and effective manner. At the time of their publication, the reports offered a vision and a strategy for advancing this process and stimulated a great deal of thought and discussion about the challenges of incorporating the methods of 21st Century Toxicology (TT21, derived from the title of the 2007 NRC report) into toxicological best practices (Andersen and Krewski, 2010; see also Boekelheide and Campion, 2010; Bus and Becker, 2009; Chapin and Stedman, 2009; Meek and Doull, 2009; Hartung, 2009a,b). Recognizing that new technologies were not yet mature and their potential was yet to be realized, the initial focus was on screening large inventories of previously untested chemicals, in order to identify those substances that were highest priority for follow-up testing. Thus, to date TT21 technologies have primarily been used for screening and prioritization, with few attempts to use these technologies for risk assessment and to inform risk management decisions.

TT21 technologies have largely been used by institutions and agencies that regulate and/or study toxicants, in order to identify, understand and potentially prevent adverse effects on human health and the environment. Several large initiatives are underway to implement TT21 science. Among the largest are ToxCast and ExpoCast at the U.S. Environmental Protection Agency (EPA), Integrated Approaches to Testing and Assessment (IATA) for skin sensitization at the Organization for Economic Co-operation and Development (OECD), and Tox21. The latter is a consortium between the National Toxicology Program at the National Institute of Environmental Health Sciences, the National Institute of Health Chemical Genomics Center, the National Center for Computational Toxicology of the U.S. EPA, and the U.S. Food and Drug Administration (FDA). Over the past decade, a large and steady stream of newly developed alternative testing approaches and toxicity data have been generated by automated high-throughput screening (HTS) and high-content screening (HCS) assays to prioritize and evaluate a vast number of potential toxicants across multiple platforms.

To help build the capacity to use TT21 approaches for risk assessment, toxicologists have also been working to develop computational models for organizing the content of this massive data stream. The models are essential, because they map the data onto relevant biological pathways, organize the data in meaningful ways, and correlate the data with linked events in defined pathways of toxicity. Some of the new models involve novel approaches and include the development of new complex human cell culture models and engineered model systems. Other examples include 3D organotypic cultures, microscale tissues, microphysiological systems, and other novel *in vitro* data and *in silico* models. In parallel, the Human Toxome project, sponsored by the NIH is developing the means for the identification of pathways of toxicity (Kleensang *et al.*, 2014) from multi-omics approaches (Bouhifd *et al.*, 2015). The field of toxicology as a whole is gradually building the capacity to measure exposure, assess pharmacokinetics and metabolism, integrate distinct data streams, and evaluate responses to perturbations in a systems biology framework, and toxicologists are becoming adept at using these powerful new tools, techniques and computational approaches. With vast data and powerful technologies in hand, the field of "predictive toxicology" now faces the challenge of defining best practices and consensus on exactly how to move forward. To reiterate, the vision expressed in the 2007 National Research Council report to transform and eventually replace traditional toxicology testing by bringing TT21 HTS

and HCS, and other *in vitro*, and computational (*in silico*) approaches into the realm of high throughput human risk assessment (HTRA) is slowly being realized. There is a strong feeling that HTRA can and will be used to support biologically based weight-of-evidence arguments in the context of regulatory decision-making, possibly moving towards probabilistic risk assessment. A significant benefit of this transition, when complete, will be the reduction or elimination of animals in testing of industrial and commercial chemicals for evaluating potential risk to human health.

The present document describes key discussion points and outcomes of a Society of Toxicology (SOT) Contemporary Concepts in Toxicology (CCT) Workshop, entitled FutureTox III that was held in Crystal City, Virginia, November 19–20, 2015. The workshop built on the many lessons learned from the first 10 years of TT21 and the first 2 workshops in the FutureTox series [for summary of FutureTox II see (Knudsen *et al.*, 2015); for summary of FutureTox I see (Rowlands *et al.*, 2014)]. FutureTox III was attended in person and via webcast by more than 300 scientists from government research and regulatory agencies, research institutes, academia, and the chemical and pharmaceutical industries in Europe, Canada, and the US. At this workshop, participants reviewed and discussed the state of the science in toxicology and human risk and exposure assessment with a focus on moving TT21 science into the arena of regulatory decision-making. Whereas the 2-day conference included multiple presentations, what is presented below are 2 case studies of advancements using TT21 science, discussion of several presentations including exposome and AOPs, Day1/Day2 summary points and perspectives on the future landscape, including opportunities and challenges related to TT21 science. Summaries of the break-out group sessions are included in the Supplemental Materials.

SETTING THE SCENE

At the time when FutureTox III was held, the field of toxicology remained in a period of transition from the use of traditional toxicology approaches to more efficient and informative TT21 approaches. Despite excitement over the emerging TT21 technologies for screening and prioritization, the path towards using TT21 approaches in risk assessment and regulatory decision-making was unclear and the magnitude of the challenge remained great. The first 2 workshops in this series, FutureTox I and FutureTox II focused on the challenges and scientific opportunities inherent in the major paradigm shift already underway in the field of toxicology as well as on developing the methods, models, and capacity needed to realize the NRC's TT21 vision and strategy. At FutureTox III, the emphasis shifted to how best to harness the new paradigm, science and data for use in human risk assessment and regulatory decision-making.

FutureTox III occurred at a time of growing optimism and confidence that the vision laid out in the 2007 NRC report (NRC, 2007) was attainable. Importantly, key representatives from U.S. regulatory agencies, James Jones [Assistant Administrator for the EPA's Office of Chemical Safety and Pollution Prevention (OCSPP)], David Dix (Director of the Office of Science Coordination and Policy within OCSPP), and David Strauss (on behalf of Robert Califf, FDA Commissioner) expressed optimism and growing confidence that TT21 results can be used to help inform regulatory decision-making. In particular, Jones, Dix, and Strauss described and discussed 2 successful demonstration cases that illustrate the power of TT21 methods and approaches. They expressed a firm commitment to use these

approaches as the cornerstone of predictive toxicology in the immediate future to generate valuable *in vitro* data and *in silico* models. Expressing a similar sentiment, whereas alluding to recent substantial progress towards a TT21-based risk assessment paradigm, Kevin Crofton (Deputy Director of EPA's National Center for Computational Toxicology) suggested in closing remarks at the workshop, that "NowTox" might have been a more appropriate name for the workshop than FutureTox III.

"NowTox" Demonstration Case 1: The EDSP Pivot

The EPA has moved with direction and efficiency to apply TT21 approaches to inform regulatory decisions, as these approaches are proven to be scientifically sound and acceptable. The first example is the EPA's Endocrine Disruptor Screening Program (EDSP). Congress enacted the Food Quality Protection Act (FQPA) in 1996, which mandated that EPA first develop, validate, and then apply test systems to screen substances for estrogen receptor (and other endocrine receptor-mediated) effects in humans. After seeking expert advice, EPA developed a 2-stage test system comprised of 11 Tier 1 *in vitro* and *in vivo* screening assays for evaluating endocrine receptor-targeted bioactivity and 5 Tier 2 *in vivo* tests (2 mammalian, 1 avian, 1 amphibian, 1 fish) for evaluating adversity and the dose-dependence of these potentially toxic compounds. The proposed Tier 1 and Tier 2 testing methods were validated and subjected to peer review. At the same time as the EDSP Tier 1/Tier 2 tests were developed, EPA identified and ordered the screening of 67 chemicals, including active and high production volume inert ingredients used in pesticides (EDSP List 1) (U.S. EPA, 2009; Juberg *et al.*, 2014), whereas a second list of chemicals was prepared/modified in 2010/2014, including 107 additional chemicals to be screened using the EDSP Tier 1 battery of 11 assays. Dix and Jones reported that as of November 2015, Tier 1 tests were complete for all chemicals on EDSP List 1, analysis of these data was ongoing, and testing of a small number of chemicals on EDSP List 2 had begun. Based on this rate of progress and the fact that EPA is ultimately responsible for testing up to 10 000 chemicals [ie, chemicals identified by the Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA) and the Safe Water Drinking Act (SWDA)], it would take more than a century for EPA to complete its assigned task using the EDSP Tier 1 and Tier 2 assays. Given the scale of this problem, it was recognized that at the current rate of progress the anticipated completion of this task in a timely fashion was not possible. Concern over the projected time and cost of EDSP testing using Tier 1 tests, coupled with the recognition that ToxCast *in vitro* assays and other TT21 approaches could potentially be used to complete the same task within a more acceptable timeframe, resulted in EPA taking a bold step forward to apply these approaches to the EDSP (U.S. EPA, 2015). Over time it is expected that this approach will be applied to other regulatory programs and goals, as confidence in these approaches increase, and they are shown to be "fit for purpose".

The new, pivotal approach to EDSP testing, informally called the EDSP Pivot Program, employs a computational network model that integrates data from 18 *in vitro*, HTS ER assays. These assays quantify ER binding, dimerization, chromatin binding, transcriptional activation, and ER-dependent cell proliferation (Figure 1) (Judson *et al.*, 2015). In pilot studies, the model was validated using 43 reference chemicals based on the uterotrophic assays (30 active, 13 inactive). In the next phase of the program, the ER agonist/antagonist potential of 1812

commercial and environmental chemicals was quantified. Recent publications (Browne *et al.*, 2015; Kleinstreuer *et al.*, 2016) demonstrate that the ToxCast ER model performs as well or better than the EDSP Tier 1 ER binding, ERTA, and uterotrophic assays, with a reported accuracy of 84–93%, or better, if inconclusive scores are excluded from the dataset. In describing the success of this effort, Browne *et al.* (2015) write that the EDSP Pivot "represents the first step in a paradigm shift for chemical safety testing, a practical approach to rapidly screen thousands of environmental chemicals for potential endocrine bioactivity in humans and wildlife, and the first systematic application of ToxCast data in an EPA regulatory program."

In 2015, EPA published a commentary paper on the EDSP Pivot in the Federal Register, with the title "Use of High Throughput Assays and Computational Tools; Endocrine Disruptor Screening Program; Notice of Availability and Opportunity for Comment." At FutureTox III, Dix also discussed how EDSP Pivot's ER model fits into the ER agonist/antagonist-dependent adverse outcome pathway, a critical link to biological outcomes and real world human health risk. With regard to its screening mandate, both Jones and Dix pointed out that the EDSP Pivot will make it possible to complete what was previously a century-long task, in a fraction of the time, possibly only a few years (Figure 2). In terms of moving towards a new risk assessment paradigm, Dix also emphasized that the EDSP Pivot represents a large conceptual step forward for the toxicology/risk assessment communities, in that it challenges the idea that only *in vivo* data can inform our understanding of mechanism. Dix asserted that EPA is already using *in silico* approaches extensively, because it allows toxicologists and risk assessors to address the complexity inherent in biological systems without expensive and time-consuming animal studies. In summary, the EDSP Pivot demonstrates the power of TT21 science and the benefit of operationalizing TT21 science at EPA. Looking to the future, Jones and Dix laid out an ambitious plan to extend this approach to screen the EDSP universe of 10 000 plus chemicals for androgenic, steroidogenic and thyroid-like activities (Figure 2).

"NowTox" Demonstration Case 2: The Comprehensive *In Vitro* Pro-Arrhythmia Assay

In the 10 years between 1991 and 2001, 14 pharmaceutical agents, including cisapride and terfenadine, were introduced and then withdrawn from various commercial markets worldwide, because these drugs were linked to unacceptably high risk of a potentially lethal adverse cardiac event known as Torsade de Pointes (TdP) (Yap and Camm, 2003). Many of the drugs were withdrawn because of a perceived risk of TdP, even though the drugs were marketed and used to treat non-cardiac health concerns. In response to the high frequency of unintended and potential adverse cardiac events from these drugs, and to prevent additional similar events, the International Conference on Harmonization (ICH) issued 2 guidance documents in 2005, S7B and E14, which were immediately endorsed by the FDA. These guidance documents required specific preclinical and clinical tests to be conducted, to determine if candidate drugs block the hERG (human Ether-à-go-go Related Gene, or KCNH2 in the new nomenclature) potassium channel in the heart or trigger an event through a cardiac anomaly known as long QT syndrome. Although the ICH guidelines were effective in that they prevented the approval of new drugs with known effects on QT prolongation or TdP, the guidelines also had an associated cost/down-side. This is because the correlation between hERG
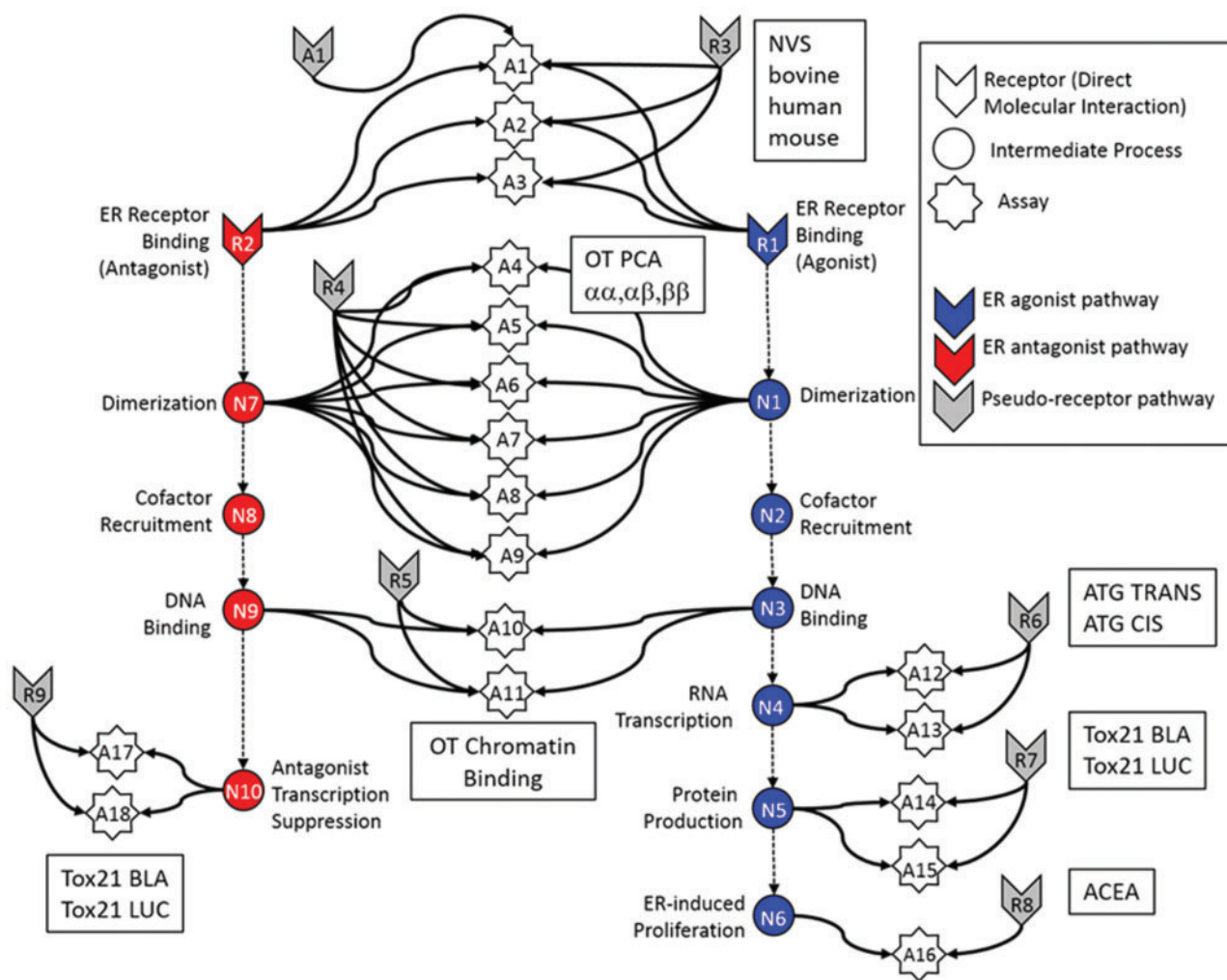
FIG. 1. High throughput assays integrated into an ER predictive pathway model. Graphical representation of the computational network used in the *in vitro* analysis of the ER pathway across assays and technology platforms for EDSP Pivot. Colored arrow nodes represent "receptors" with which a chemical can directly interact. Colored circles represent intermediate biological processes that are not directly observable. White stars represent the *in vitro* assays that measure activity at the biological nodes. Arrows represent transfer of information. Gray arrow nodes are the pseudo-receptors. Each *in vitro* assay (with the exception of A16) has an assay-specific pseudo-receptor, but only a single example is explicitly shown, for assay A1. Source: Judson et al. 2015 (reuse permission license number 3903120062524).
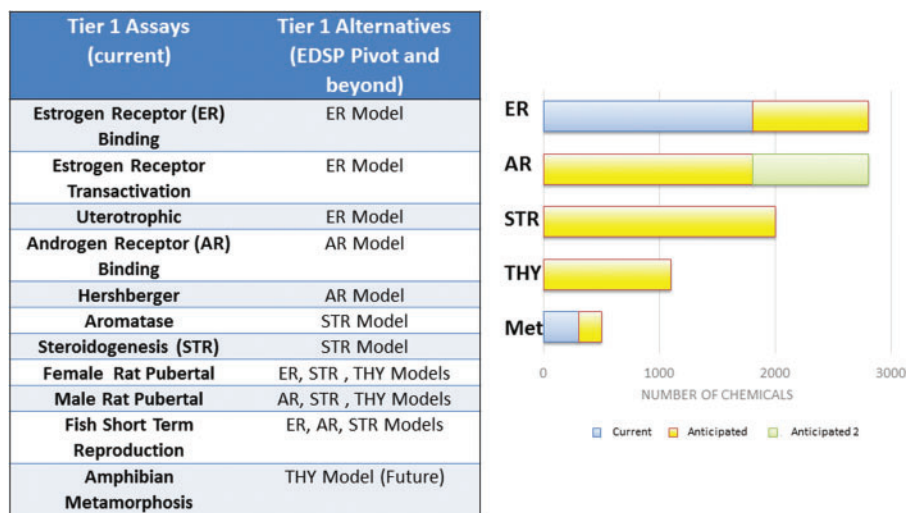

FIG. 2. Tier 1 alternative models in the endocrine disruptor screening program (EDSP). Estrogen receptor (ER) model is currently available; androgen receptor (AR), steroidogenesis (STR), thyroid (THY) models are still in development stages. The progression of testing is demonstrated at the time of the conference: Current (published), Anticipated (data on hand but not yet published), and Anticipated 2 (data being collected).
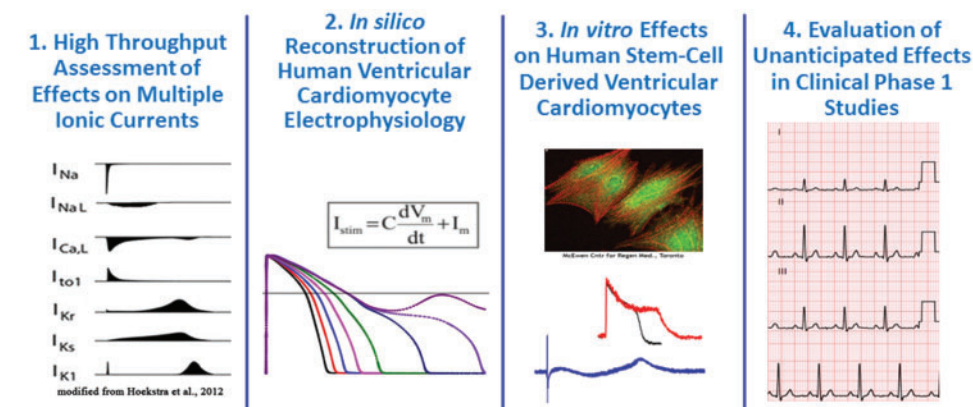
FIG. 3. Comprehensive *in vitro* proarrhythmia assay: 4 components. The goal is to develop a new *in vitro* paradigm for cardiac safety evaluation of new drugs that provides a more accurate and comprehensive mechanistic-based assessment of proarrhythmic potential.

potassium channel inhibition or QT interval prolongation and TdP is not absolute (Gintant et al., 2016; Johannesen et al., 2014). This situation was the stimulus that led to a forward-looking development and widespread endorsement of a program based on TT21 science, known as the Comprehensive *in vitro* Proarrythmia Assay (CiPA). The goal of CiPA is to make it possible to develop and market pharmaceuticals that benefit human health without increased risk of TdP, and without unnecessary prohibition of pharmaceuticals linked to low risk effects on hERG or the duration of QT (Fermini et al., 2016). Validation work is ongoing and the ICH S7B and E14 working group is being continually updated on progress (Colatsky et al., 2016).

At FutureTox III, CiPA was described by Dr Strauss, a Medical Officer at FDA's Center for Devices and Radiological Health (CDRH). In brief, CiPA is an international cross-sector collaboration between 8 partners (FDA, Health and Environmental Sciences Institute, Cardiac Safety Research Consortium, Health Canada, Japan National Institutes of Health, Safety Pharmacology Society, European Medicines Agency, and the Japan Pharmaceuticals and Devices Agency) whose goal is to develop and facilitate adoption of an evidence-based method to assess potential for drug-induced TdP using 4 interrelated program components: (1) mechanistically based HTS assays using homogenous, genetically engineered human cells that overexpress a human cardiac ion channel (one at a time); (2) *in silico* reconstruction of human cardiac action potential; (3) *in vitro* screening assays in human stem cell-derived cardiomyocytes; and (4) targeted small scale first-in-human Phase I clinical studies using body surface ECG (Figure 3; see also Sager et al., 2014). The components of CiPA are being developed by 4 expert working groups: the ion channel working group, the *in silico* working group, the cardiomyocyte working group and the translational/clinical working group. The FDA is strongly committed to successful implementation of CiPA.

In the same manner as the EDSP Pivot represented a conceptual shift for the EPA, CiPA represents a relatively radical departure from the "status quo" for the FDA. In continuing progress towards a new conceptual framework and a systems biology/systems toxicology-based risk assessment paradigm, *in silico* models will play a critical role in integrating disparate data streams. Remarkably, as described above, *in silico* models are being evaluated for inclusion, along with all other data, into 2 model programs, EDSP Pivot and CiPA, that are successful demonstration cases for high-throughput TT21-based hazard identification, risk assessment and regulatory decision-making.

Moving Towards Functional Exposure Science: Systems Exposure and the Exposome

In order to perform a human risk assessment for a single chemical, one must understand the relative toxicity of the chemical and the toxicokinetic/toxicodynamic parameters that determine how it interacts with human physiology/biology, and be able to estimate the concentration of the chemical to which humans are exposed. Until recently, the relative paucity of exposure data has been recognized as an unavoidable weakness that hinders progress in the pursuit of more informed, accurate, and refined human risk assessment (Egeghy et al., 2012). There is now a sense that this is beginning to change, that exposure science has been influenced by a conceptual "snowball" effect, in that the relatively new idea of *systems exposure* has emerged and is moving to the forefront of exposure science (Pleil, 2011). Along with systems exposure comes the idea of the *exposome*, which can be thought of as the cumulative exposure to industrial, commercial and consumer chemicals over time and its impact on human and environmental health. Dr John Wambaugh (EPA's National Center for Computational Toxicology (NCCT)) addressed the idea of systems exposure and exposure pathways/routes at FutureTox III, and went on to describe how exposure models are being built based on chemico-physico properties of chemicals as well as real or inferred/predicted information on chemical use and production volume. Wambaugh reported that nearly 50% of the exposure variance in a set of NHANES urine samples can be explained by 4 use-dependent "yes/no" variables (Industrial use; pesticide inert; pesticide active; consumer use) and the production volume of the chemical, if those variables are known (Dionisio et al., 2015; Goldsmith et al., 2014; Wambaugh et al., 2014). Unfortunately, these 5 variables remain unknown for a large number of chemicals in the exposome.

To move beyond this impasse, new approaches are being developed to infer or deduce information about those variables (Egeghy et al., 2015). Wambaugh also briefly described how exposure models are evaluated using a framework called Systematic Empirical Evaluation of Models (SEEM) (Wambaugh et al., 2013, 2014). Wambaugh emphasized that in many instances, even an estimated exposure that has relatively high uncertainty can be extremely valuable. For example, the relative risk of an adverse outcome is small if the highest possible exposure is 3 orders of magnitude below the lowest effect level for a single toxicant (ie, a large margin of exposure) (Wetmore et al., 2013). Looking to the future, including the challenge of

analyzing mixtures that contain many unknown toxicants, the newest frontier of exposure science is non-targeted exposure assessment. As an example, Wambaugh presented a mass spectrometer tracing of house dust with 947 peaks, and indicated that only 3% of the 947 peaks in a mass-spectrophotometry profile of house dust could be identified (Rager et al., 2016). The goal for the future is to identify and quantify a significant fraction of the chemicals in an environmental or biological sample such as house dust or human urine, a process analogous to sorting and identifying not one, but many different needles in a haystack. When this goal is achieved, exposure scientists will be one step closer to developing tools that will allow them to study the exposome.

Adverse Outcome Pathways, AOP Networks and the AOP Knowledge-Base

Traditional animal tests have been accepted as appropriate for toxicity testing for nearly 70 years because they are conducted in intact animals that integrate all the steps leading from exposure to adverse outcomes that may occur in humans. A major impediment to the acceptance of Tox21 testing results in risk assessment is defining the biological activity that each assay or combination of assays in a Tox21 approach appropriately measures and how best to combine suites of assays to appropriately mimic the entire range of events from exposure to adverse outcomes of concern for humans. The concept of Adverse Outcome Pathway (AOP) has been developed (Ankley et al., 2010) and agreed to internationally as a means to achieve this end, largely through the efforts of the Organization for Economic Cooperation and Development (OECD) http://www.oecd.org/chemicalsafety/ testing/adverse-outcome-pathways-molecular-screening-and-tox icogenomics.htm; last accessed September 29, 2016.

At FutureTox I, significant discussion was devoted to conceptualizing the steps that link perturbation of a biological system to an adverse [apical] outcome as a pathway. However, there were still improvement opportunities relative to terminology being used and how best to develop, share, test, and formalize pathways as tools for explaining/understanding interactions between biological systems and toxicants. In his presentation at FutureTox III, Dr Daniel Villeneuve [EPA's National Health and Environmental Effects Research Laboratory (NHEERL)] showed that this scenario has changed significantly in the last 4 years. Villeneuve defined an AOP as a conceptual framework that explains chemical perturbation of a biological system as a molecular initiating event (MIE) that progresses through a sequence of key events (KEs) to a specific adverse outcome (AO) of regulatory concern. The existence and acceptance of an AOP provides a rationale for monitoring the impacts of chemicals that activate the AOP, as well as their prevalence in the physical and biological environment, and their interactions with human and non-human biological systems. Villeneuve described several core principles of AOP development that have been codified over the past 2 years. Briefly, AOPs are not chemical-specific, and whereas they are modular, they are represented by simplified linear chains: MIE → KEs → AO. Multiple MIEs that converge on specific KEs may lead to the same AO, resulting in an inter-linked AOP network (Villeneuve et al., 2014). For a single pure ligand, a single AOP can be a functional unit of prediction; however, if more than one ligand is present, the AOP network becomes the functional unit of prediction.

An AOP is a conceptual framework that continues to evolve as more information is gained on the series of events. The AOP Knowledge-base (AOP-KB) is a formal internet-based repository for information about AOPs that went public in 2014 (http:// aopkb.org/; last accessed September 29, 2016). The AOP-KB is structured to be systematic, organized, transparent, defensible, credible, accessible, and searchable. By contributing to the AOP-KB, users can expect to facilitate integration and analysis of information about AOPs and to avoid duplication of effort among AOP developers/users. AOPs are gaining acceptance; as of June 20, 2016, the AOP wiki site (https://aopwiki.org/aops; last accessed September 29, 2016) listed about 130 AOPs. To date, one AOP (Covalent Protein binding leading to Skin Sensitisation) has been OECD endorsed by the Working Group of The National Coordinators for the Test Guidelines Programme (WNT) and the Task Force on Hazard Assessment (TFHA); another 6 AOPs are under WNT and TFHA review (declassification awaited in June 2016); 12 are under internal review by the OECD Extended Advisory Group on Molecular Screening and Toxicogenomics (EAGMST); and the remainder are still under development.

The majority of AOPs describe endpoints classified as reproductive/endocrine, central nervous system, developmental, carcinogenic, or hepatic. Some of the major challenges in developing and improving the AOP-KB are: encouraging consistent vocabulary across AOPs, avoiding duplicate entries, encouraging AOP development and submission of AOPs to the AOP-KB, bridging the research/regulatory interface, developing quantitative AOPs so that dose-response relationships can be incorporated, and encouraging predictive use/testing of AOPs. The quality and scope of the AOP-KB could play an important role in solidifying the emerging paradigm of systems toxicology, which is characterized by increased use of TT21 in vitro, and in silico tools.

Emerging Beyond the Tipping Point

Without TT21 science, testing the many thousands of chemicals used in commerce for potential adverse effects is an endless, intractable task. With TT21 science, the task becomes manageable, the willingness and impetus to use powerful TT21 science to solve other scientific problems increases, and the process of implementing the TT21 vision begins to accelerate. Dr. Maurice Whelan (European Commission Joint Research Centre, Ispra, Italy) described this transition in the field of toxicology graphically (Figure 4A), pointing out that the "Sisyphus Effect" is dominant on the uphill side of the transition from "old approach" to "new approach", the "Snowball Effect" is dominant on the downhill side of the transition, and the "Tipping Point" occurs at the top of the curve. The "Sisyphus Effect" refers to the punishment of Sisyphus, a figure from Greek mythology who was tasked with endlessly pushing a heavy stone to the top of the mountain over and over again, only for it to roll down the mountain under the force of gravity after each ascent. As described in Malcolm Gladwell's widely-read book *The Tipping Point: how little things make a big difference* (Gladwell, 2000), "the tipping point is the moment when an idea, trend, or social behavior crosses a threshold, tips, and spreads like wildfire." This book borrowed some concepts from epidemiology to describe and analyze in great detail how, why and when social trends turn into social epidemics. Thomas Schelling, joint recipient of the 2005 Nobel Prize in Economic Sciences for "having enhanced our understanding of conflict and cooperation through game-theory analysis," wrote a treatise in 1971 concerning social dynamics of discriminatory choices (Schelling, 1971). His simulation-analytics models showed that neighborhood segregation patterns resulted from the dynamics of collective movement, rather than individual motives. Schelling's "tipping point" theory of neighborhood segregation patterns may apply to sociology of the paradigm shift in
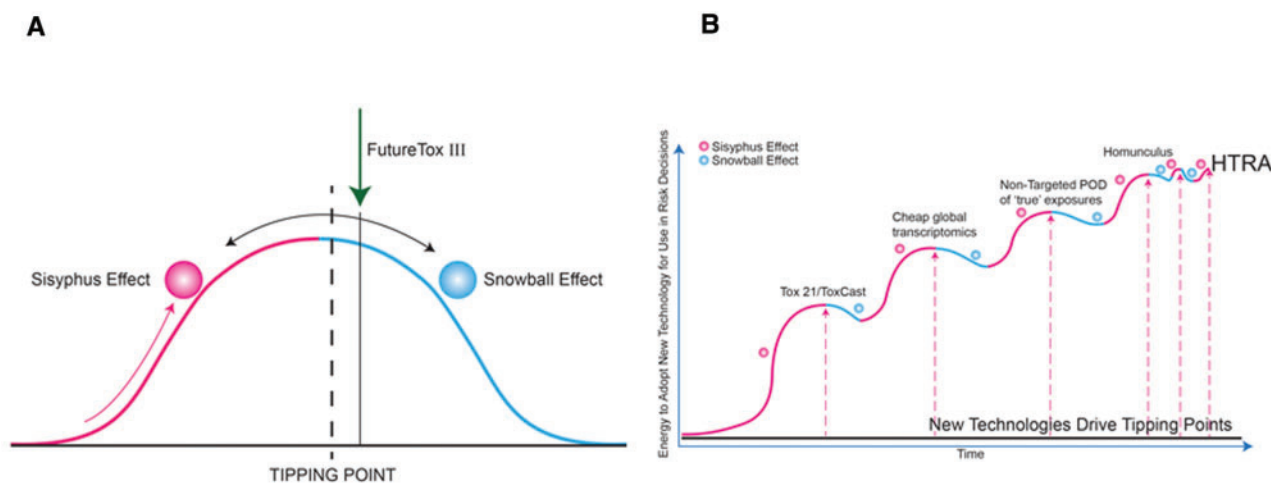
A

B

FIG. 4. Moments of change. (A) The Sisyphus effect tipping over to the Snowball effect. (B) Keeping pace with changing scientific technologies that continue to overcome multiple tipping points toward high-throughput risk assessment (HTRA).

TT21 toxicology, whereby the "tipping point" will be in the dynamics of collective movement in the science rather than individual advancements. Maurice Whelan's account suggested that the field of toxicology was (at the time of FutureTox III) just beyond the tipping point in accepting/adopting/implementing TT21 science (Figure 4A), an idea that was mentioned in the workshop summations presented by Drs. Tina Bahadori (U.S. EPA) and Kevin Crofton (U.S. EPA). Stating that we are "beyond the tipping point," is equivalent to saying, "there is no going back to the old paradigm," an irreversible change has occurred and we are ready to accept the consequences and move forward. In other words, toxicologists are now committed to using *in vitro* and *in silico* human-based models to understand, predict, and manage human toxicological risk.

Certainly a fundamental underpinning of this impending transition to TT21 science, particularly if global acceptance is to occur, is the demonstration of assay/pathway/approach validation and that new approaches are sensitive, accurate and specific and can suitably and reliably supplant or perhaps replace traditional methods which may be in use today. This would, for example, include *in vitro* data and *in silico* models showing that novel test systems actually react to, and predict, responses from drug/chemical exposure in a manner that is qualitatively, and more importantly, quantitatively linked to adverse outcomes in more traditional toxicology bioassays and/or human exposures. Status of the underlying TT21 science with regards to assay sensitivity/specificity and efficiency/speed/cost was a central theme of the antecedent FutureTox-II conference (Knudsen *et al.*, 2015). Of course, many social, economic, legal, and political changes will also need to occur before TT21 approaches are fully accepted and implemented. However, the first step has been taken, because it is clear that TT21 testing approaches, used appropriately, can be more informative and efficient than traditional toxicology. Acceptance of TT21 science in non-scientific arenas is expected to follow in due course.

Stepping Ahead to Predictive Systems Toxicology

Given the complexity of biological processes and continuing advancements in the technologies for detecting toxicological responses (Langley *et al.*, 2015), multiple tipping points may abound. More recent large-scale empirical testing made feasible by census databases have suggested Schelling's tipping point

model is oversimplified (Easterly, 2009). Easterly acknowledges that tipping points may exist but real world scenarios must consider "dynamic instability of intermediate points". In the last slide presented at FutureTox III, Crofton represented the pathway to high throughput risk assessment (HTRA) as a stairway, where the first step was acceptance/adoption of TT21/ToxCast approaches (Figure 4B). Crofton encouraged participants at FutureTox III to move quickly to accept and begin to capitalize on future technological breakthroughs, especially at the regulatory level (Figure 4B). An appropriately circumspect and scientifically valid strategy is essential to avoid a "chain reaction due to herd behavior" (Easterly, 2009). Tipping points are generally not described as physically irreversible; however, enacting change in the reverse direction would be less facile, and require increasingly more energy.

Although many early changes following a toxicological perturbation are adaptive, in that they ultimately preserve cellular homeostasis, once a critical threshold is crossed, the integrity of the "old" system is weakened by the cumulative effect of many small biological changes: this is the point when the cell commits to an apical (ie, adverse) response, a.k.a. the intracellular tipping point, as it pertains to a biological system. Unfortunately, it is technologically difficult to track and quantify the myriad of biochemical interactions and thermodynamics of the intracellular response to a toxicant in real time: the process is complex and currently we lack the tools to gather a complete set of data (Shah *et al.*, 2016). However, *in vitro* data and *in silico* models may ultimately be able to capture this complexity with sufficient detail and sufficient fidelity for these models to be useful in predictive toxicology. If that is the first key goal of TT21 science, then the second goal is to apply the resulting knowledge to effectively regulate chemical compounds and to protect human health and the environment, through better understanding of these biological and cellular tipping points and to make the case in the larger social, economic, legal and political arenas to effect acceptance and change. FutureTox III shows that we are getting much closer to achieving these 2 goals.

Breakout Groups

FutureTox III offered breakout group discussions in 4 topical areas: Drug Development, EDSP, chemical legislation reform

under the Toxic Substances Control Act (TSCA), and Impact on Global Harmonization. The groups discussed current technologies, future needs for regulatory use and suggestions for a productive path forward. A few key points from the breakout groups are discussed below and further details can be found in Supplemental Materials.

Although much data are now available (eg, ToxCast/TT21), having the right kind of data for translation varies by topic. Important points were sorting the science before trying to apply it. For example, AOPs are an important framework for organizing weight-of-evidence information on chemicals. But as a decision-support system, it may take AOPs many years before they become acceptable in a regulatory context. The need to "prove" that the set of AOP's capture all the MIEs that will show a clinical risk for drug discovery or hazard for consumer product use has a long way to go. Thus, whereas the AOP framework is a positive start, much more information and data will be required to inform decision-making. However, when screening for endocrine activity at EPA, it is believed that the tools currently exist or are well under development. Further development is needed for assays for other endocrine endpoints and consideration of metabolism of chemicals is a necessary area of research and development. The chemical reform group focused on the need to be able to use upstream biological events to predict downstream biological events. There was not a strong consensus regarding the current state of the science, but all agreed that a better understanding of absorption, distribution, metabolism and excretion (ADME) is critical. The need to better understand ADME was recognized in all breakout groups.

Breakout group discussions also recognized the importance of developing *in vitro* data and computational models that are fit-for purpose. There is a need to strategically implement specific *in vitro* assays to replace the animal uses that are of most concern. Progress will likely be seen more quickly when the questions are targeted and specific. In addition, to accelerate the acceptance of new assays, it may be reasonable to focus on an area where the risk of a false positive or false negative is of minor consequence. A key ingredient for global harmonization is "accept change". This requires education (next generation and current practitioners), engaging multiple stakeholders, allowing time for learning curves, and overcoming uncertainties that underlie societal resistance to change. Once there is acceptance of a few assays, it is hoped that further innovations will be welcomed by not only researchers but also regulators. Educating and involving regulators early will help them to gain familiarity with new approaches.

Challenges and Gaps

Although FutureTox III was characterized by strong optimism about the future of toxicology, several challenges and knowledge gaps were recognized. Near the end of his keynote presentation, Jim Jones (EPA) stated that "the reality of our lives is a reality of mixtures," suggesting that the time has come to address the risk of real-world simultaneous exposure to multiple chemicals using computational tools. This, along with the desire to understand more about chronic, low-dose and multiple sequential exposures remain unmet challenges (and opportunities for advancement in knowledge) for the field. Knowledge of the variance in toxicant susceptibility across population subgroups also remains an unmet challenge, and this goes hand-in-hand with the need for a less human-centric interpretation of the responsibility of toxicologists. Human toxicology can be seen as one piece in the larger puzzle of ecotoxicology; in this context, the adverse impact of toxicants on sentinel species should remind us of this fact. In the spirit of embracing the large picture in all its complexity, toxicologists must engage in discussion with stakeholders in academia, industry, regulatory, and non-governmental organizations as well as all parts of civic society as they shape the future of toxicology. To this point, Dr Elaine Faustman described both challenges as well as opportunities for the academic community. These included university training programs to enhance translational science which offer expanded opportunities for bench scientists to mentor students and fellows in emerging approaches/technologies, such as microphysiological and engineered microsystems for *in vitro* models and computational toxicology for *in silico* models. For example, the EPA has been holding in-person workshops and webinars to encourage use of ToxCast; many stakeholder groups including non-governmental organizations (NGOs) and students have participated and these efforts have been well-received and well-attended. Finally, referring to the future of toxicology, Crofton offered a quote from Yogi Berra, "It's pretty far, but it doesn't seem like it." Another Yogi Berra quote might be equally appropriate at this time: "The future ain't what it used to be." In other words: what we called the "future of toxicology" yesterday, is already here and now.

SUPPLEMENTARY DATA

Supplementary data are available online at http://toxsci. oxfordjournals.org/.

TERA Center, University of Cincinnati, ToxServices, American Academy of Clinical Toxicology, American College of Toxicology, CAAT, Johns Hopkins Center for Alternatives to Animal Testing (CAAT), In Sphero, Lhasa Limited, Ramboll Environ, Safety Pharmacology Society, Society of Toxicologic Pathology, The Teratology Society.

REFERENCES

Andersen, M. E., and Krewski, D. (2010). The vision of toxicity testing in the 21st century: Moving from discussion to action. *Toxicol. Sci.* **117**, 17–24.

Ankley, G. T., Bennett, R. S., Erickson, R. J., Hoff, D. J., Hornung, M. W., Johnson, R. D., Mount, D. R., Nichols, J. W., Russom, C. L., Schmieder, P. K., *et al.* (2010). Adverse outcome pathways: A conceptual framework to support ecotoxicology research and risk assessment. *Environ. Toxicol. Chem.* **29**, 730–741.

Boekelheide, K., and Campion, S. N. (2010). Toxicity testing in the 21st century: Using the new toxicity testing paradigm to create a taxonomy of adverse effects. *Toxicol. Sci.* **114**, 20–24.

Bouhifd, M., Beger, R., Flynn, T., Guo, L., Harris, G., Hogberg, H. T., Kaddurah-Daouk, R., Kamp, H., Kleensang, A., Maertens, A., *et al.* (2015). Quality assurance of metabolomics. *ALTEX* **32**, 319–326.

Browne, P., Judson, R. S., Casey, W. M., Kleinstreuer, N. C., and Thomas, R. S. (2015). Screening chemicals for estrogen receptor bioactivity using a computational model. *Environ. Sci. Technol.* **49**, 8804–8814.

Bus, J. S., and Becker, R. A. (2009). Toxicity testing in the 21st century: A view from the chemical industry. *Toxicol. Sci.* **112**, 297–302.

Chapin, R. E., and Stedman, D. B. (2009). Endless possibilities: Stem cells and the vision for toxicology testing in the 21st century. *Toxicol. Sci.* **112**, 17–22.

Colatsky, T., Fermini, B., Gintant, G., Pierson, J. B., Sager, P., Sekino, Y., Strauss, D. G., and Stockbridge, N. (2016). The comprehensive in vitro proarrhythmia assay (CiPA) initiative—Update on progress. *J. Pharmacol. Toxicol. Methods*. doi: 10.1016/j.vascn.2016.06.002.

Dionisio, K. L., Frame, A. M., Goldsmith, M. R., Wambaugh, J. F., Lidell, A., Cathey, T., Smith, D., Vail, J., Emstoff, A. S., Fantke, P., *et al.* (2015). Exploring consumer exposure pathways and patterns of use for chemicals in the environment. *Toxicol. Rep.* **2**, 228–237.

Easterly, W. (2009). Empirics of strategic interdependence: The case of the racial tipping point. *J. Macroecon.* **9**, 25.

Egeghy, P. P., Judson, R., Gangwal, S., Mosher, S., Smith, D., Vail, J., and Cohen hubal, E. A. (2012). The exposure data landscape for manufactured chemicals. *Sci. Total Environ.* **414**, 159–166.

Egeghy, P. P., Sheldon, L. S., Isaacs, K. K., Ozkaynak, H., Goldsmith, M. R., Wambaugh, J. F., Judson, R. S., and Buckley, T. J. (2015). Computational exposure science: An emerging discipline to support 21st-century risk assessment. *Environ. Health Perspect.* **124**, 697–702.

Fermini, B., Hancox, J. C., Abi-Gerges, N., Bridgland-Taylor, M., Chaudhary, K. W., Colatsky, T., Correll, K., Crumb, W., Damiano, B., Erdemli, G., *et al.* (2016). A new perspective in the field of cardiac safety testing through the comprehensive in vitro proarrhythmia assay Paradigm. *J. Biomol. Screen.* **21**, 1–11.

Gintant, G., Sager, P. T., and Stockbridge, N. (2016). Evolution of strategies to improve preclinical cardiac safety testing. *Nat. Rev. Drug Discov.* **15**, 457–471.

Gladwell, M. (2000). *The Tipping Point: How Little Things Can Make a Big Difference*. 1st ed. Little Brown, Boston.

Goldsmith, M. R., Grulke, C. M., Brooks, R. D., Transue, T. R., Tan, Y. M., Frame, A., Egeghy, P. P., Edwards, R., Chang, D. T., Tornero-Velez, R., *et al.* (2014). Development of a consumer product ingredient database for chemical exposure screening and prioritization. *Food Chem. Toxicol.* **65**, 269–279.

Hartung, T. (2009a). Toxicology for the twenty-first century. *Nature* **460**, 208–212.

Hartung, T. (2009b). Toxicology for the 21st century: Mapping the road ahead. *Toxicol. Sci.* **109**, 18–23.

Johannesen, L., Vicente, J., Mason, J. W., Sanabria, C., Waite-Labott, K., Hong, M., Guo, P., Lin, J., Sørensen, J. S., Galeotti, L., *et al.* (2014). Differentiating drug-induced multichannel block on the electrocardiogram: Randomized study of dofetilide, quinidine, ranolazine, and verapamil. *Clin. Pharmacol. Ther.* **96**, 549–558.

Juberg, D. R., Borghoff, S. J., Becker, R. A., Casey, W., Hartung, T., Holsapple, M. P., Marty, M. S., Mihaich, E. M., Van Der Kraak, G., Wade, M. G., *et al.* (2014). Lessons learned, challenges, and opportunities: The U.S. endocrine disruptor screening program. *Altex* **31**, 63–78.

Judson, R. S., Magpantay, F. M., Chickarmane, V., Haskell, C., Tania, N., Taylor, J., Xia, M., Huang, R., Rotroff, D. M., Filer, D. L., *et al.* (2015). Integrated model of chemical perturbations of a biological pathway using 18 *in vitro* high-throughput screening assays for the estrogen receptor. *Toxicol. Sci.* **148**, 137–154.

Kleensang, A., Maertens, A., Rosenberg, M., Fitzpatrick, S., Lamb, J., Auerbach, S., Brennan, R., Crofton, K. M., Gordon, B., Fornace, A. J. Jr., *et al.* (2014). Pathways of toxicity. *Altex* **31**, 53–61.

Kleinstreuer, N. C., Ceger, P. C., Allen, D. G., Strickland, J., Chang, X., Hamm, J. T., and Casey, W. M. (2016). A curated database of rodent uterotrophic bioactivity. *Environ. Health Perspect.* **124**, 556–562.

Knudsen, T. B., Keller, D. A., Sander, M., Carney, E. W., Doerrer, N. G., Eaton, D. L., Fitzpatrick, S. C., Hastings, K. L., Mendrick, D. L., Tice, R. R., *et al.* (2015). FutureTox II: *In vitro* data and in silico models for predictive toxicology. *Toxicol. Sci.* **143**, 256–267.

Langley, G., Austin, C. P., Balapure, A. K., Birnbaum, L. S., Bucher, J. R., Fentem, J., Fitzpatrick, S. C., Fowle, J. R., 3rd, Kavlock, R. J., Kitano, H., *et al.* (2015). Lessons from toxicology: Developing a 21st-century paradigm for medical research. *Environ. Health Perspect.* **123**, A268–A272.

Meek, B., Doull, J. (2009). Pragmatic challenges for the vision of toxicity testing in the 21st century in a regulatory context: Another Ames test? ...or a new edition of "the Red Book"? *Toxicol. Sci.* **108**, 19–21.

National Research Council. (2007). *Toxicity Testing in the 21st Century: A Vision and a Strategy*. National Academies Press, Washington D.C.

National Research Council (2009). *Science and Decisions: Advancing Risk Assessment*. National Academies Press, Washington (DC).

Pleil, J. D., Sheldon, L. S. (2011). Adapting concepts from systems biology to develop systems exposure event networks for exposure science research. *Biomarkers* **16**, 99–105.

Rowlands, J. C., Sander, M., Bus, J. S., and FutureTox Organizing, C. (2014). FutureTox: Building the road for 21st century toxicology and risk assessment practices. *Toxicol. Sci.* **137**, 269–277.

Rager, J. E., Strynar, M. J., Liang, S., McMahen, R. L., Richard, A. M., Grulke, C. M., Wambaugh, J. F., Isaacs, K. K., Judson, R., Williams, A. J., *et al.* (2016). Linking high resolution mass spectrometry data with exposure and toxicity forecasts to advance high-throughput environmental monitoring. *Environ. Int.* **88**, 269–280.

Sager, P. T., Gintant, G., Turner, J. R., Pettit, S., and Stockbridge, N. (2014). Rechanneling the cardiac proarrhythmia safety paradigm: A meeting report from the Cardiac Safety Research Consortium. *Am. Heart J.* **167**, 292–300.

Schelling, T. C. (1971). Dynamic models of segregation. *J. Math. Soc.* **1**, 143–186.

Shah, I., Setzer, R. W., Jack, J., Houck, K. A., Knudsen, T. B., Martin, M. T., Reif, D. M., Richard, A. M., Dix, D. J., and Kavlock, R. J. (2016). Elucidating dynamic modulation of cellular state function during chemical perturbation. *Environ. Health Perspect.* **124**, 910–919.

U.S. EPA (2009). Endocrine disruptor screening program; Tier 1 screening order issuance announcement. *Federal Register* **74**, 54422–54428.

U.S. EPA (2015). Use of high throughput assays and computational tools; Endocrine disruptor screening program; Notice of availability and opportunity for comment. *Federal Register* **80**, 35350–35355.

Villeneuve, D. L., Crump, D., Garcia-Reyero, N., Hecker, M., Hutchinson, T. H., LaLone, C. A., Landesmann, B., Lettieri, T., Munn, S., Nepelska, M., *et al*. (2014). Adverse outcome pathway (AOP) development I: Strategies and principles. *Toxicol. Sci.* **142**, 312–320.

Wambaugh, J. F., Setzer, R. W., Reif, D. M., Gangwal, S., Mitchell-Blackwood, J., Arnot, J. A., Joliet, O., Frame, A., Rabinowitz, J., Knudsen, T. B., *et al*. (2013). High-throughput models for exposure-based chemical prioritization in the ExpoCast project. *Environ. Sci. Technol.* **47**, 8479–8488.

Wambaugh, J. F., Wang, A., Dionisio, K. L., Frame, A., Egeghy, P., Judson, R., and Setzer, R. W. (2014). High throughput heuristics for prioritizing human exposure to environmental chemicals. *Environ. Sci. Technol.* **48**, 12760–12767.

Wetmore, B. A., Wambaugh, J. F., Ferguson, S. S., Li, L., Clewell, H. J., 3rd, Judson, R. S., Freeman, K., Bao, W., Sochaski, M. A., Chu, T. M., *et al*. (2013). Relative impact of incorporating pharmacokinetics on predicting *in vivo* hazard and mode of action from high-throughput in vitro toxicity assays. *Toxicol. Sci.* **132**, 327–346.

Yap, Y. G., and Camm, A. J. (2003). Drug induced QT prolongation and torsades de pointes. *Heart* **89**, 1363–1372.