OXFORD UNIVERSITY PRESS | Toxicological Sciences

# FutureTox IV Workshop Summary: Predictive Toxicology for Healthy Children

| | |
|---|---|
| Journal: | *Toxicological Sciences* |
| Manuscript ID | TOXSCI-20-0436.R2 |
| Manuscript Type: | Forum Article |
| Date Submitted by the Author: | 13-Jan-2021 |
| Complete List of Authors: | Knudsen, Thomas; US EPA, ORD; USEPA<br>Fitzpatrick, Suzanne; FDA, Office of the Commissioner<br>De Abrew, K. Nadira; The Procter & Gamble Company, Central Product Safety/Global Product Stewardship<br>Birnbaum, Linda; NIEHS, OD<br>Chappelle, Anne; Chappelle Toxicology Consulting, LLC<br>Daston, George; Procter & Gamble Company, Miami Valley Laboratories<br>Dolinoy, Dana; University of Michigan, Environmental Health Sciences<br>Elder, Alison; University of Rochester, Department of Environmental Medicine<br>Euling, Susan; U. S. Environmental Protection Agency, National Center for Environmental Assessment<br>Faustman, Elaine; University of Washington, Environmental and Occupational Health Sciences<br>Fedinick, Kristi; Natural Resources Defense Council<br>Franzosa, Jill; U.S. EPA, National Center for Computational Toxicology<br>Haggard, Derik; US Environmental Protection Agency, National Center for Computational Toxicology<br>Haws, Laurie; ToxStrategies, Inc.<br>Kleinstreuer, Nicole; National Toxicology Program,<br>Buck-Louis, Germaine; George Mason University<br>Mendrick, Donna; NCTR/FDA, Regulatory Activities<br>Rudel, Ruthann; Silent Spring Institute Inc., 29 Crafts Street<br>Saili, Katerine; US EPA, NCCT<br>Schug, Thaddeus; National Institute of Environmental Health Sciences, Population Health Branch<br>Tanguay, Robyn; Oregon State University, Environmental and Molecular Toxicology<br>Turley, Alexandra; FDA<br>Wetmore, Barbara; US Environmental Protection Agency Office of Research and Development, Office of Research and Development;<br>White, Kimberly; American Chemistry Council<br>Zurlinden, Todd; United States Environmental Protection Agency Research Triangle Park Campus, National Center for Computational Toxicology |
| Category - Please select one category that is most applicable to your | Developmental and Reproductive Toxicology |

| | |
|---|---|
| manuscript.: | |
| Category 2 - Optional: select a second category that is also applicable to your manuscript.: | In Vitro and Alternative Methods |
| Key Words: | high-throughput, new approach methodologies, developmental toxicity, Children's Health, human health, human risk assessment, mathemetical modelling, predictive toxicology, systems biology, regulatory sciences |
| | |

**FORUM**

# FutureTox IV Workshop Summary: *Predictive Toxicology for Healthy Children*

Thomas B. Knudsen[1*], Suzanne Compton Fitzpatrick[2*], K. Nadira De Abrew[3], Linda S. Birnbaum[4], Anne Chappelle[5], George P. Daston[3], Dana C. Dolinoy[6], Alison Elder[7], Susan Euling[8], Elaine M. Faustman[9], Kristi Pullen Fedinick[10], Jill A. Franzosa[1], Derik E. Haggard[1,11], Laurie Haws[12], Nicole Kleinstreuer[13], Germaine M. Buck Louis[14], Donna L. Mendrick[15], Ruthann Rudel[16], Katerine S. Saili[1], Thaddeus T. Schug[4], Robyn L. Tanguay[17], Alexandra E. Turley[2], Barbara A. Wetmore[1], Kimberly W. White[18], and Todd J. Zurlinden[1]

[1]US Environmental Protection Agency, ORD, Research Triangle Park, North Carolina
[2]US Food and Drug Administration, College Park, Maryland
[3]The Procter and Gamble Company, Cincinnati, Ohio
[4]National Institute of Environmental Health Science, NIH, Research Triangle Park, North Carolina
[5]Chappelle Toxicology Consulting, LLC, Chadds Ford Pennsylvania
[6]University of Michigan, Ann Arbor, Michigan
[7]University of Rochester, Rochester, New York
[8]US Environmental Protection Agency, Office of Children's Health Protection, Washington, DC
[9]Elaine M. Faustman, University of Washington, Seattle, Washington
[10]Natural Resources Defense Council, Washington, DC
[11]Oak Ridge Institute for Science and Education (ORISE)
[12]ToxStrategies, Inc., Austin, Texas
[13]National Toxicology Program, NIEHS, Durham, North Carolina
[14]George Mason University, Fairfax, Virginia
[15]US Food and Drug Administration, NCTR, Silver Spring, MD
[16]Silent Spring Institute, Newton, Massachusetts
[17]Oregon State University, Corvallis, Oregon
[18]American Chemistry Council, Washington, DC
(*) Chair and Co-chair

**DISCLAIMERS:** The views expressed in this article are those of the authors and do not necessarily reflect the views or policies of the US Environmental Protection Agency, the US 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

FutureTox IV, a Society of Toxicology (SOT) Contemporary Concepts in Toxicology (CCT) workshop, was held in November 2018. Building upon FutureTox I, II, and III, this conference focused on the latest science and technology for *in vitro* profiling and *in silico* modeling as it relates to predictive developmental and reproductive toxicity (DART). Publicly available high throughput screening data sets are now available for broad *in vitro* profiling of bioactivities across large inventories of chemicals. Coupling this vast amount of mechanistic data with a deeper understanding of molecular embryology and post-natal development lays the groundwork for using new approach methodologies (NAMs) to evaluate chemical toxicity, drug efficacy, and safety assessment for embryo-fetal development. NAM is a term recently adopted in reference to any technology, methodology, approach, or combination thereof that can be used to provide information on chemical hazard and risk assessment to avoid the use of intact animals (USEPA, 2018). There are challenges to implementing NAMs to evaluate chemicals for developmental toxicity compared with adult toxicity. This forum article reviews the 2018 workshop activities, highlighting challenges and opportunities for applying NAMs for adverse pregnancy outcomes (e.g., preterm labor, malformations, low birth weight) as well as disorders manifesting postnatally (e.g., neurodevelopmental impairment, breast cancer, cardiovascular disease, fertility). DART is an important concern for different regulatory statutes and test guidelines. Leveraging advancements in such approaches and the accompanying efficiencies to detecting potential hazards to human development are the unifying concepts toward implementing NAMs in DART testing. Although use of NAMs for higher level regulatory decision making is still on the horizon, the conference highlighted novel testing platforms and computational models that cover multiple levels of biological organization, with the unique temporal dynamics of embryonic development, and

novel approaches for estimating toxicokinetic parameters essential in supporting *in vitro* to *in vivo* extrapolation.

**KEY WORDS:** High-throughput screening**,** developmental and reproductive toxicity testing, *in vitro* profiling, *in silico* modeling, pediatric health, children's environmental health, new approach methodologies (NAMs)

## 1. INTRODUCTION

Progress and strategies have been addressed over the past two decades to increase the applicability, implementation, and acceptance of modern animal-free methods for efficiently and credibly evaluating chemical toxicity, drug efficacy, and safety assessment (ICCVAM, 2018; Luechtefeld et al., 2018; Mahony et al., 2020; Thomas et al., 2019). Stakeholders from governmental and regulatory agencies, research institutes, academia, and the chemical and pharmaceutical industry are working together under a 21st century paradigm to address these complex challenges (NAS, 2017); (NRC, 2007; 2009; 2012). Under the new paradigm, use of high-throughput and computational modeling approaches aims to transform the components in risk assessment while systematically addressing key challenges that hinder their acceptance for translation to public health and environmental regulatory decisions. New approach methodologies (NAMs) utilizing high-throughput *in vitro* profiling of chemical-bioactivity relationships, and *in silico* models can support efforts to minimize the use of vertebrate animal tests for prediction of hazard and exposure (EC, 2006; USA, 2016); (USEPA, 2019a). For example, the 2016 Frank R. Lautenberg Chemical Safety for the 21st Century Act (LSCA) that amended the Toxic Substances Control Act (TSCA) requires the United States Environmental Protection Agency (USEPA) to incorporate intrinsic and extrinsic factors that affect susceptibility, adequately assessing exposure among vulnerable groups, and accurately identifying highly exposed groups (Koman et al., 2019). Under LSCA the USEPA must encourage and facilitate *"… the use of scientifically valid test methods and strategies that reduce or replace the use of vertebrate animals while providing information of equivalent or better scientific quality and relevance that will support regulatory decisions …"* and also consider the impacts of chemicals and chemical mixtures to subpopulations who *"…may be at greater risk than the general population of adverse health effects*

*from exposure to a chemical substance or mixture, such as infants, children, pregnant women, workers, or the elderly*" (USA, 2016). In 2018, sixteen US federal agencies contributed to a strategic framework to enable development of, establish confidence in, and ensure use of new approaches to toxicity testing that improve human health relevance and reduce or eliminate the need for testing in animals (ICCVAM, 2018). A directive issued by USEPA Administrator Andrew Wheeler in September of 2019 called for reducing mammalian study requests 30% by 2025 and eliminating them by 2035 (USEPA, 2019a). USEPA convened a conference in December 2019 to discuss NAMs for achieving reduced animal testing in chemical safety research and issued a guidance in February 2020 for meeting its animal testing reduction goals [https://www.epa.gov/pesticide-registration/bridging-or-waiving-data-requirements].

The SOT FutureTox conference themes address the many challenges facing toxicology in the 21st-century. The series launched in 2012 with FutureTox: Building the Road for 21st-Century Toxicology and Risk Assessment Practices (Rowlands et al., 2014). A nexus was the expanding data collection from *in vitro* profiling of thousands of pharmaceutical compounds and environmental chemicals (Collins et al., 2008). Data-driven models and tiered-testing strategies would aim to translate points of departure, modes of action, and adverse outcome pathways (AOPs) into real-world human exposure scenarios for hazard prediction and translational challenges such as absorption-distribution-metabolism-elimination (ADME)/pharmacokinetics, life-stage considerations, and biological complexity (Rowlands et al., 2014). An AOP is conceptualized as a linear sequence of events progressing from a discrete molecular initiating event (MIE) to adverse outcome (AO) through a series of intermediate key events (KE) as a pragmatic tool for regulatory toxicology.

In 2014, FutureTox II: *In Vitro* Data and *In Silico* Models for Predictive Toxicology focused on priority concerns, such as predicting and modeling metabolism, cell growth and differentiation, effects on susceptible subpopulations and life-stages, and integrating data into risk assessment (Knudsen et al., 2015). In 2015, FutureTox III: Bridges for Translation focused on how best to harness the new paradigm, science, and data for use in human risk assessment and regulatory decision making (Juberg et al., 2017).

The Society of Toxicology's FutureTox IV conference was held November 14-16, 2018 in Arlington, Virginia. (http://www.toxicology.org/events/shm/cct/futuretoxiv.asp). It focused on predictive toxicology for children's health, emphasizing pregnancy and development. Basic, clinical, and regulatory scientists discussed the specific challenges and opportunities for NAMs in Developmental and Reproductive Toxicity (DART) testing. Replacing *in vivo* DART bioassays with new methods is complicated by the fact that development is a series of precisely orchestrated temporal and spatial events. Capturing each of those lifestage-specific events and the changes that occur over time is complex. This FutureTox IV conference report highlights the science and technology issues, needs, and recommendations to support decisions about chemical toxicity, drug efficacy, and safety assessment for adverse pregnancy (e.g., preterm labor, malformations, low birth weight) and childhood outcomes (e.g., neurodevelopmental impairment, pubertal outcomes).

## 2. SETTING THE STAGE

Two examples set the stage for assessing children's health outcomes in the context of federal regulatory decisions over the last century: the Biologics Control Act of 1902, enacted after several children died after receiving adulterated diphtheria antitoxin; and the Food, Drug and Cosmetic Act of 1938 following many childhood deaths from elixir of sulfanilamide compounded with diethylene glycol. Firms were then required to prove to USFDA that any new drug was safe before it could be marketed. Following the episode of congenital anomalies with maternal use of thalidomide, the 1962 Kefauver-Harris Amendment required manufacturers to prove efficacy as well as safety. Pediatric health concerns have driven many historical milestones and legislations since then, including the Best Pharmaceuticals for Children Act (BPCA, enacted in 2002, amended in 2007) and the Pediatric Research Equity Act (PREA, enacted in 2003, amended in 2007) which provide the incentives (BPCA) and requirements (PREA) to study and label pediatric therapeutics. Although principles underlying pediatric drug development are consistent with general drug development, use of a product in children must demonstrate substantial evidence of effectiveness and safety while also considering their differences from adults both physiologically (they are still developing) and behaviorally (unique exposure pathways). Clinical trials cannot place children at a disadvantage nor extrapolate from adult trials just because it would be difficult to do a pediatric study. Although the number of approvals with pediatric labeling have steadily increased between 1998-2017, so have the attrition rates, where pediatric trials fail to establish safety or efficacy (Green et al., 2018; Momper et al., 2015).

There are several opportunities for NAMs in pediatric therapeutic development. Lifestage-specific *in vitro* approaches and *in silico* models - including pharmacokinetics (PK) and pharmacodynamics (PD) - can increase confidence in making predictions that are based on non-pediatric models.

Mechanistic modeling can also support regulatory decision-making by integrating available data from multiple analytical approaches. Consider thalidomide (and related compounds), for example, that may disrupt a transcriptional network working through zinc finger transcription factors such as *SALL4*. Heterozygous loss of function mutations in *SALL4* result in a human developmental condition that phenocopies thalidomide-induced birth defects, and thalidomide induces SALL4 protein degradation exclusively in humans, primates, and rabbits, but not in rodents or fish, providing a mechanistic link for the species-specific pathogenesis of thalidomide syndrome (Donovan et al., 2018). It remains to be demonstrated that a SALL4-based assay can eventually be shown to contribute to the predictive toxicology of other chemicals; however, that work (with thalidomide) shows how, in the absence of tractable animal models that closely resemble the human disease, predictive modeling can benefit from developmentally-relevant data that are generated using human embryonic stem cells (hESC) (Donovan et al., 2018).

Regarding NAMs to evaluate the safety of chemicals and medical products, the ICCVAM roadmap (ICCVAM, 2018) emphasized communication, collaboration, commitment. The so-called 3C's include not only understanding what needs to be done (e.g., organizations and stakeholders must communicate their decision contexts), but incentivizing the adoption of NAMs through building confidence in the *in vitro* data and identifying ways to improve their applicable domain through communication. An important consideration for validation is that NAMs must provide relevant information for the specific decision context in which they are being used. Furthermore, the validation process needs to take into consideration the impact of both false positive and false negative results on the decision-making process to avoid having the NAMs become more

conservative than the currently-used animal methods. An example is the assessment of potential estrogenicity (Juberg et al., 2014) for compounds tested in the ToxCast/Tox21 battery of HTS *in vitro* assays versus the rodent uterotrophic assay (Browne et al., 2015). The HTS data showed value in having large amounts of high-quality data for predicting the uterotrophic response. Computational models are needed to extrapolate *in vitro* data to *in vivo* toxicity in order to fully implement data from alternate test methods (Thomas et al., 2019). For DART, fit-for-purpose NAMs will require extrapolation to various lifestages including pregnancy as well as test methods that reflect morphogenesis, growth and differentiation for translatability into children's health protection strategies (Knudsen et al., 2017).

Dealing with the complexity of development is a significant challenge, due in part, to varying susceptibility windows and multifactorial etiologies (e.g., genetic, nutritional, infectious, environmental exposures, and social exposures. For example, 10-15% of US children may display neurodevelopmental disorders, indicating the potential for complex gene-environment-social interactions during pregnancy and postnatal development (Boyle et al., 2011). Up to 163 environmental chemicals that can adversely affect neurodevelopment in animal models were assessed in targeted evaluation of body fluids from pregnant women in the National Health and Nutrition Examination Survey (NHANES) (Woodruff et al., 2011). USEPA's 'TSCA Inventory' now lists more than 86,000 chemicals (USEPA, ). NAMs offer an opportunity to screen for key biological activity of these chemicals to identify chemicals requiring further research on exposure-dose-response relationships.

## 3. STATE OF THE SCIENCE AND FUTURE DIRECTION

Unique susceptibilities at developmental life stages include concepts such as critical windows of development, the germline as a unique target, and alternative and HTS tools for identifying and predicting developmental toxicants. Consider, for example, vulnerability of the developing brain and the permanent consequences in adulthood following chemical perturbation during pregnancy. Brain development is the most dynamic *in utero* although several processes affect human brain structure into childhood and adolescence, including myelination and arborization. Internal dosimetry and surrogate biomarkers can be used to identify risk profiles during embryonic patterning and morphogenesis (e.g., neurulation and neural tube differentiation), fetal brain growth (neuroprogenitor cell growth, differentiation, and apoptosis) and maturation (neural migration, synaptogenesis, myelination). Although total brain volume is about 95% of its adult size by the age of 5 years, various subcomponents of the brain do undergo developmental changes that continue well into adolescence and early adulthood (reviewed in (Giedd, 1999; Johnson et al., 2009)). Magnetic resonance imaging is well-suited for studies of childhood brain development because it supports repeated scans of the same individual. NAM data can support predictive models that link epidemiological data from internal dosimetry (e.g., NHANES) with disease biomarkers from diagnostic neuroimaging. For example, epidemiology may reveal statistical associations between exposure to specific chemical classes and altered neurodevelopmental outcomes, such as diminished IQ (high prenatal organophosphate pesticides) (Sagiv et al., 2019) or thyroid disruption (high polybrominated diphenyl ethers, PBDEs) (Dishaw et al., 2014). A battery of *in vitro* assays can then be used to profile large inventories of small molecules to identify effects on processes ranging from early patterning of the neural tube to neuroprogenitor differentiation, brain growth and functional maturation.

Non-mammalian testing platforms can be used to evaluate effects unique to development. Testing chemical effects on the germline is an opportunity for the application of simple invertebrate systems. Several assays developed for assessing germline dysfunctions caused by direct or indirect (transgenerational) chemical exposures using the roundworm, *Caenorhabditis elegans*. The germline in this as in any organism, is established in early embryogenesis and then maintained throughout development and in the adult gonad (Schaner and Kelly, 2006). Germline impairment can be assessed by monitoring the production of green fluorescent protein (GFP) expressed in aneuploid embryos. Compounds with the potential to disrupt germline segregation in this manner were identified by screening the ToxCast Phase I chemical library. This evaluation demonstrated a balanced accuracy of 69% when compared to *in vivo* reproductive toxicants (Allard et al., 2013). This accuracy approaches that possible with alternative platforms based on vertebrate species, where performance levels have reached or exceeded 70% (as discussed in later sections). Another application of the *C. elegans* system was to test the hypothesis that changes in histone modifications may be required for transgenerational reproductive defects. While these studies could take several years and use several hundred animals if performed in mammals, they can be achieved at much lower cost and speed (2-3 weeks) in *C. elegans* which has a generation time of 4 days under standard culture conditions. Using a reporter for repetitive germline silencing that requires a proper regulation of histone modifications (Camacho et al., 2018; Lundby et al., 2016), BPA was found to dramatically reduce two repressive histone marks (H3K9me3 and H3K27me3) in the germline as well as cause germline apoptosis and embryo lethality at the F3 generation following ancestral (P0) bisphenol A exposure (100 µM). However, chemical or RNAi knockdown of the cognate demethylases reversed the effects of BPA on

these two marks and rescued all germline transgenerational effects at the F3. These results indicated that histone modifications can be an important vector of environmental information passed through the germline. NAMs based on *C. elegans* could thus serve as a rapid, medium-throughput screening platform for germ line effects of chemicals that may propagate across multiple generations.

High-throughput screening (HTS) non-mammalian testing platforms using zebrafish have already been implemented to evaluate chemicals for developmental toxicity (NRC, 2000). The morphogenesis and organogenesis of the vertebrate body plan in this species follows a basic principle of evolutionary conservation of developmental signaling pathways known to be critical in higher vertebrates. Transgenic, mutated, and transparent fish lines are available for HTS assays, as are technologies for gene knock-down (morpholino) and gene editing (CRISPR/Cas9) that can be used for functional analysis (Hwang et al., 2013). Transgenic reporter lines highlighting the vasculature, skeletal system, and neuromast cells were used to screen 293 ToxCast chemicals (McCollum et al., 2017). The reporter line highlighting bone (GFP) and cartilage (mCherry) identified 38 bone disruptor compounds of which at least seven caused cartilage malformations. Correlating these to the ToxCast assay portfolio returned a 'ToxPi' signature [http://toxpi.org] that included several human cytochrome (CYP) P450 enzyme inhibition assays (hCYP2J2, hCYP3A4, hCYP2C11). Since these CYPs participate in vitamin D metabolism (Jones, 2012), functional rescue was sought, and thus far, has confirmed for rotenone and simazine, with exogenous vitamin D supplementation (10 nM). RNAseq analysis of embryos exposed to cyproconazole inferred alterations in adipose development as well, and exposure to cyproconazole was subsequently shown to induce lipid accumulation in zebrafish larva. The capacity to run tiered screening in zebrafish reveals not only target-system response (e.g.,

skeletal defects) but also sensitive cell types (e.g., osteoblast, osteoclast, chondrocyte, adipocyte) and diverse biological processes for AOP elucidation, as was demonstrated for angiogenesis screens (McCollum et al., 2017), leading to the concept that *in vitro* profiling can barcode specific phenotypes and predict relevant *in vivo* developmental disruption.

In addition to their value as a screening platform for anatomical defects, zebrafish have a rich repertoire of behaviors that can be monitored quantitatively in an automatic manner to evaluate adverse and therapeutic effects of small molecules on behaviors including sleep patterns, learning, fear, aggression, addiction, social interaction, and so forth. (Geng and Peterson, 2019). To model socialization, a novel 'parallel social assessment arena' was devised wherein a test zebrafish subject placed in one chamber windowed to another can be monitored for how much time the subject spends near to, or away from, the stimulus (another zebrafish). Socialization score was tested for 1200 compounds and the top four reducers were all fluoroquinolone topoisomerase 2 inhibitors. Since the response depends on visual stimulus, further work is needed to resolve a primary effect (social avoidance) from one secondary to visual impairment; however, mutating topoisomerase recapitulated the phenotype in unexposed zebrafish and a topoisomerase-II inhibitor reduced social interaction in mice as well. Zebrafish embryos could therefore be used for assessing the impacts of chemical and pharmaceutical exposures on complex neurobehaviors that develop later in life.

Organotypic culture models and microphysiological systems have emerged for studying complex processes such as developmental neurotoxicity, testicular development, developmental effects and effects occuring through translational life stages. Although many diverse platforms are

providing rapid DART data, their power comes from combining and integrating data from different technologies and chemical classes (Ciallella and Zhu, 2019). Whereas the potential for 'big-data' yields mechanistic insight, AOPs provide one possible bridge for these data to phenotypic relevance to the human (Sturla et al., 2014). For DART, programming organotypic culture models engineered from human stem cells and microphysiological systems may capture enough tissue architecture to assess chemical effects on morphogenesis, growth and differentiation in an integrated human-relevant platform (Knudsen et al., 2017). Federal and European agencies have invested in the development and implementation of organoids, organ-on-chip and microphysiological system constructs to support human safety models (Truskey, 2018; Watson et al., 2017). Complex 3D (organotypic) systems and *in silico* models have potential applications for data generation and integration into AOP-based constructs.

### *Neurodevelopment*

One example of a brain microphysiological systems for neurodevelopment is in the evaluation of Autism Spectrum Disorder (ASD). ASD prevalence (1 in 54, (Maenner et al., 2020)) is increasing in populations at a rate that is too fast to be explained by genetic drift, implicating complex interactions between genomic, environmental, and lifestyle factors beyond improved diagnostic capabilities (Maenner et al., 2020). For *in vitro* testing to more accurately and reliably recapitulate the complexity of neurodevelopment at a tissue/organ level, the need arises for 3D models that can be synthetically engineered into self-organizing heterogeneous tissues. Such models would be useful to "… *decode the toxicological blueprint of active substances that interact with the developing embryo*" (Sturla et al., 2014). Human iPSC (induced pluripotent stem cell) derived 3D 'BrainSpheres' yield a standardized,

reproducible system that has most cell types relevant to neurodevelopment. These structures develop spontaneous electrophysiological activity ($Ca^{2+}$ influx) and patterns of differential growth and organization that recapitulate neurodevelopment in a human cell-based system (Pamies et al., 2017). Studies have affirmed the potential for these BrainSpheres as an alternative model for testing the effects of an antidepressant (paroxiten) and other chemicals on the developing brain (Zhong et al., 2020). A case example using BrainSpheres to evaluate developmental neurotoxicity of chlorpyrifos in conjunction with heterozygous knock-out of CHD8 (Chromodomain Helicase DNA Binding Protein 8), a gene linked to ASD. They observed a synergistic oxidative stress response where *CHD8* expression was downregulated by exposure to chlorpyrifos or its active metabolite (chlorpyrifos-oxon) perhaps in response to an effect on the dopamine transporter. Oxidative stress identified with both CHD8-deficiency and chlorpyrifos toxicity implies a common response in at least some forms of ASD. An expansion of the technology to epidemiology would include, for example, identification of biomarkers of autism or constructs from patient-derived iPSCs for other neurological conditions.

Neural tube defects such as exencephaly, spina bifida, or retinal coloboma reflect cellular disruptions that are region-specific both with regards to positional information and timing. Both parameters are sensitive to local retinoic acid signaling (Lippmann et al., 2015) through temporal regulation of HOX (homeobox) gene expression (Iimura and Pourquie, 2007). A novel HTS platform has been developed based on self-organizing human neuroprogenitor stem cells locked into regional fates by extrinsic retinoic acid signaling (Lippmann et al., 2015). Clonal lines that harbor HOX gene expression patterns reflective of *in vivo* regional position in the neural tube were generated by *in vitro* exposure to retinoic acid for progressively longer periods. These cell lines are then arrayed in

96-well plates and therefore amenable to HTS to investigate regional susceptibility to xenobiotics. The neurogenic clones build on organizing principles of positional information set up in the neural tube by signaling in the primitive streak and notochord. Differentiating their genomic responses to pharmaceuticals or chemical exposures *in vitro* can allow one to model or screen the potential for regional-specific effects along the anterior-posterior or dorsal-ventral axes (Lippmann et al., 2015). In this system, spontaneous morphogenesis is minimized by controlled micropatterning, whereby the neuroprogenitor cells are seeded as 2D monolayers on micropatterned culture substrates and from there reproducibly self-organize into singularly polarized neuroepithelial rosettes, which are composites of relevant architectures 8-10 cell layers thick (Knight et al., 2018). These organoids polarize just like radial glia in the neuroepithelium and can display archetypical phenotypes of the forebrain, hindbrain, and spinal segments. For neurodevelopmental toxicity screening, the extent of *in vitro* rosette formation serves as a surrogate for neural tube closure defects. A proof of concept was performed using exposure to valproic acid (VPA), a drug known to induce neural tube closure defects. VPA exposure inhibited rosette formation at concentrations below the therapeutic range.

### *Testis and Ovary*

Environmental chemical impacts on reproductive development are a significant public health concern, and testicular toxicity is often discovered late in the drug-development process. A 3D co-culture model for of the rodent testis recapitulated *in vivo* functions of the somatic compartment (Sertoli cells, Leydig cells), germ line compartment (spermatogenesis) and their interplay (Wegner et al., 2013). In this model, the cellular consequences of drug or chemical exposure on testicular growth, differentiation, apoptosis, and stress response can be characterized via cell-type and stage-specific

biomarkers, including, for example, tight-junction formation in Sertoli cells that are key transients supporting spermatogenesis. Manipulation of the hormonal (e.g., Luteinizing Hormone (LH), testosterone) and paracrine (e.g., cytokines, retinoic acid) milieu is possible in these models as input or output variables in testing. Modeling the pathway-level dynamics *in vitro* with *in vivo* timelines can therein pinpoint windows of vulnerability during development across species (rat, mouse, human). Ground-truthing the platform utilized microarray profiles for compounds with well-characterized *in vivo* profiles to infer similar toxicity pathways (e.g., oxidative stress). To date, 70 compounds were evaluated, including many in the ToxCast/Tox21 database. Phthalate esters, for example, invoked alterations that recapitulated metabolic and hormonal effects observed *in vivo* (Harris et al., 2015). The testicular co-culture system thus captured molecular initiating events, cellular key events, and tissue dysfunction for fetal androgen disruption underlying AOPs for the rodent phthalate syndrome, the counterpart to human Testicular Dysgenesis Syndrome (TDS).

Applications of microphysiological systems to model different aspects of the female reproductive system have been built and tested to cover hormone-sensitive stages of follicular development regulated by pituitary hormones (e.g., LH, follicle stimulating hormone (FSH)) that direct follicular growth, oocyte maturation, ovulation, and luteinization, through fertilization and implantation. An 'ovary-on-a chip' device captures ovarian hormone-sensitive changes in a synthetic human menstrual cycle (28-day) for chemical testing on the follicles and oocytes, using doxorubicin as a reference compound. Further implantation as an expanded microphysiological system simulated the *in vivo* female reproductive tract and endocrine loops between organ modules for the ovary, fallopian tube, uterus, cervix and liver, with a sustained circulating flow between all tissues (Xiao et al., 2017). The

3D follicle environment was allowed 14 days to reach maturation. Fluid pumped through the device at given flow rates controlled FSH and LH delivery, emulating follicular growth (day 0 to 14) induced with FSH and tracked by estradiol production, followed by ovulation (day 15) tracked by follicular rupture and oocyte meiosis, and luteinization (day 15 to 28) tracked by progesterone secretion. Estradiol levels peaked at day 14, and progesterone at day 17, falling to baseline by day 28. Doxorubicin exposure flattened both peaks thereby recapitulating *in vivo* effects in both prepubertal and adult mice at concentrations that align reasonably well (Wang et al., 2019; Wang et al., 2018b). The system, termed EVATAR™, recapitulates pre- or peripubertal changes of follicle activation and development, hormone secretion and ovulation. Moreover, EVATAR™ is a powerful new *in vitro* tool that allows organ–organ integration of hormonal signaling as a phenocopy of menstrual cycle (follicular phase, ovulation, and luteal phase) and pregnancy-like endocrine loops for drug discovery and toxicity assessments (Xiao et al., 2017). An ultimate goal is to integrate EVATAR™ with the NIH Tissue Chip program's other organs-on-chips to make an integrated model of a human body on a chip, enabling scientists to test the varying effects of a drug across the entire body before testing it in people.

### *Mammary gland*

Progress has been made on a 3D microfluidics device recapitulating mammary gland development for screening toxicants (Markov et al., 2012). Developmental processes and toxicities of the mammary gland are lifelong health concerns, with changes in the mother during pregnancy, changes in the daughter during postnatal life, and of course susceptibility to breast cancer later in life. This poses the complexity of multiple windows of vulnerability to toxicants at various lifestages.

While conventional 2D culture models are amenable to HTS assays to identify drug and chemical hazards, 3D cultures provide a more realistic fluid balance and microphysiology to model self-organizing behaviors, changes in the extracellular matrix, and heterogeneous cellular interactions (e.g., toxicodynamics). Effects on organoid size, proliferation/apoptosis, polarization, and other relevant endpoints can be examined under continuous toxicant exposure to establish sensitive windows. The human mammary gland on-a-chip device can be directly linked to other organ-on-chips (e.g., liver, gut) for assessing toxicokinetic parameters, such as real-time conditioning of the fluid phase. For example, concentration-dependent effects on 'mammosphere' formation were observed with estrogen mimics in a liver-to-mammary gland circuit. Impacts such as oxygen and pH gradients on system responses are being tested in interaction with hormonal priming for sensitivity.

The ongoing 'Pregnancy Exposome and Breast Cancer Across Generations' project enrolled ~20,000 pregnancies in 1959 and followed women and their children, including the onset of breast cancer in the daughters (F1), grandchildren (F2), and future great-grandchildren (F3). This cohort study leverages initial data and biospecimens for the assessment of lifestyle, health and environmental exposures and generational health. For example, p,p'-dichlorodiphenyltrichloroethane (DDT), an endocrine disrupter with responsive breast targets from *in utero* to menopause, was associated with an increased odds ratio for premenopausal breast cancer among women first exposed during infancy through puberty (Cohn et al., 2019). Overall, breast cancer risk depended on timing of first exposure and diagnosis age, suggesting susceptibility windows during development and an induction period beginning in early life. Other factors being evaluated in the pregnancy exposome project include perfluorinated compounds, body mass index, smoking history, and hypertension. Of

interest is the placenta, because morphometry of placental volume, thickness, weight and density have been identified historically from the cohort as a potential risk factor for breast cancer among granddaughters (F2). As such, ancestral environmental exposures can define risks for breast cancer and provide a path forward to preventive strategies. We refer the reader to a recent special issue on 'The Pregnancy Exposome and Breast Cancer across Generations in The Child Health and Development Studies' [https://www.sciencedirect.com/journal/reproductive-toxicology/vol/92].

The availability of a human mammary gland on-a-chip device can provide a platform to test different risk factors for breast cancer identified from human epidemiology. Linking a placenta on-a-chip device (Lee et al., 2016) to the liver-mammary platform may provide a means to identify these complex interactions, leading to improvement in protecting mother-daughter health outcomes. Furthermore, use of mammary gland development and breast cancer as regulatory endpoints of concern in DART decisions can be enhanced by a new AOP framework structure for individual cancer risks (Morgan et al., 2016).

### *Predictive modeling*

Recent *in vitro* and *in silico* advances in approaches that translate in-well effective concentration estimates to administered equivalent doses using *in vitro*-to-*in vivo* extrapolation (IVIVE). Briefly, key ADME characteristics (e.g., hepatic clearance, nonmetabolic renal clearance, plasma binding) measured in *in vitro* assays of chemical binding to human plasma dialysate and parent chemical disappearance from hepatocyte culture are used as *in vitro* surrogates to estimate these parameters (Wetmore et al., 2015). Although useful in risk-based prioritization and increasingly

integrated in NAMs, these efforts utilize pooled adult human hepatocytes and Monte Carlo simulations to model population variability based primarily on physiologic differences. Incorporation of additional physiologic and metabolic information, including consideration of short-term exposures during critical developmental windows, is required to consider vulnerable populations and lifestages including pregnant women, infants, and children. Information of ontogenetic differences in xenobiotic metabolizing enzymes in a developing infant and child and resultant effects on toxicokinetics has been successfully incorporated by combining *in vitro* isozyme-level clearance data from recombinant systems with lifestage-specific libraries containing isozyme abundance and physiologic parameters in a modeling approach (Wetmore et al., 2014). Recently, a repository of empirical models for tissue volumes, blood flow rates, and other quantities that undergo substantial changes in a human mother and her fetus during the time between conception and birth was built to ultimately support decision-making with respect to optimal pharmacological dosing and risk assessment for pregnant women and their developing fetuses. These additional considerations and information are required to address maternal changes during pregnancy, placental expression of xenobiotic enzymes and transporters, individual genetic polymorphisms, and the pregnancy exposome (USEPA, 2015). Bringing all of this together into a predictive model for pregnancy that goes beyond predicting steady state concentrations of test chemicals is a challenge for continued research. However, as such data streams are collected and developed, parallel efforts are underway to update USEPA's open source high throughput toxicokinetics (httk) R package with the relevant data and modeling tools to continue to refine this research area (Pearce et al., 2017).

Working to incorporate data from HTS *in vitro* bioactivity profiles into quantitative models that recapitulate critical phenomena during morphogenesis is an area of opportunity and challenge for *in silico* toxicodynamics. Self-organizing embryonic systems compute, in a biological sense, with complex genetic circuits akin to digital processers. Modeling these multicellular systems *in silico* can be achieved computationally with cell agent-based models (ABMs) where the functional unit (agent) is the 'cell' (Glen et al., 2019). Various morphogenetic systems have been designed with the CompuCell3d.org modeling environment (Dias et al., 2014; Hutson et al., 2017; Kleinstreuer et al., 2013; Leung et al., 2016). These ABMs build tissues cell by cell and interaction by interaction with enough computational intelligence to render emergent developmental phenotypes and adverse outcomes (e.g., malformations). Knock-down or knock-out of key signals generate *in silico* phenotypes that recapitulate the known biology. EPA's 'virtual embryo' platform contains a suite of such ABMs for different systems. One case study is a computer simulation of secondary palate morphogenesis (Hutson et al., 2017). Cleft palate has been linked to both genetic and environmental factors that perturb key events during palatal morphogenesis. Hierarchical clustering of ToxCast HTS bioactivity profiles and chemical structure revealed at least six AOPs for cleft palate (Baker et al., 2020). Specific bioactivity lesions imputed into the virtual palate resolved into probabilistic predictions of fusion defects underlying cleft palate based on epidermal growth factor and transforming growth factor (EGF-TGF) signaling switch. For example, synthetic dosing of the *in silico* model with retinoic acid, accomplished mathematically by imposing changes in the EGF-TGF switch dynamics from *in vitro* bioactivity data (Hutson et al., 2017), simulated disruption of morphogenetic fusion. A fusion competent OCM 'spheroid' model of the human palate (Belair et al., 2018) also revealed a disruption of fusion by exogenous retinoic acid (Belair et al., 2018). Therefore, tipping

points predicted from synthetic dose-response simulations can ultimately be compared to results from human cell-based organoids and back-translated to pregnancy using fetal httk exposure models (Wetmore et al., 2015). The suite of existing virtual embryo models can be expanded as a future goal to include systems underlying neurodevelopment, placental and mammary development.

## 4. CHALLENGES AND OPPORTUNITIES TO ADDRESS REGULATORY GAPS

### *Developmental neurotoxicity (DNT)*

NAMs could provide additional experimental data to inform regulatory decisions in guideline developmental neurotoxicity (DNT) studies (e.g., Organisation for Economic Co-operation and Development (OECD) TG 426, USEPA Office of Prevention, Pesticides & Toxic Substances (OPPTS) 870.6300). Alignment of academically and industry-driven assay development with regulatory needs in the field of DNT is a core mission of the International STakeholder NETwork (ISTNET) in DNT testing (Bal-Price et al., 2015). ISTNET has made progress toward a battery of mechanistically relevant, process-oriented assays for DNT (Fritsche, 2016). The battery covers different aspects of brain development from early embryonic through fetal and early neonatal stages (Figure 1).

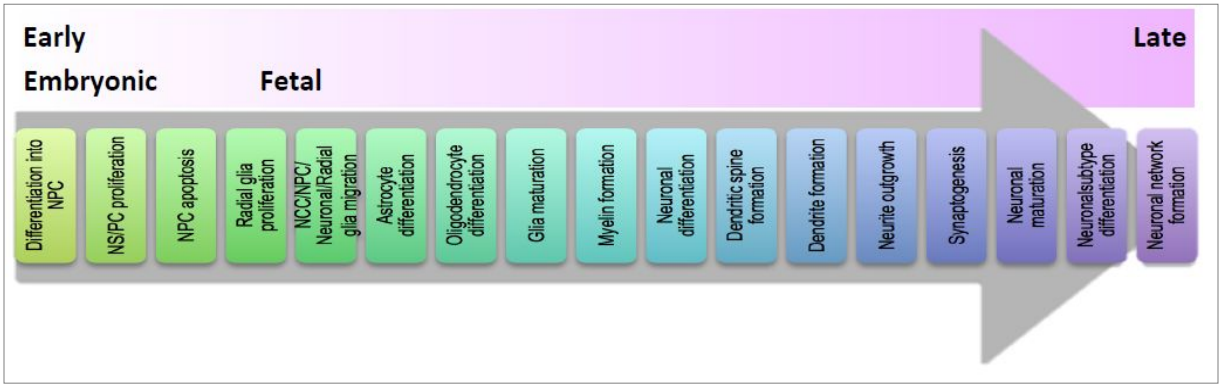

***Figure 1. Hallmark processes involved in brain development for developmental neurotoxicity (DNT).*** The *in vitro* assay battery is being tested with 117 compounds that include: 24 chemicals from an IATA case study; 48 chemicals with guideline study data (29 positive and 17 negative); 16 with some DNT evidence; and 19 negatives (Masjosthsumann et al., 2020). Abbreviations: IATA, Integrated Approaches to Testing and Assessment; NPC, neural progenitor cell; NS/PC (neural stem/progenitor cell); NCC, neural crest cell (Fritsche, 2016).

Consider, for example, an HTS neurosphere assay that using automated imaging to monitor chemical effects (Schmuck et al., 2017). EGF induces migration of radial glial cells, but not neurons. Neural crest cell migration, oligodendrocyte differentiation, neurite extension and branching, neuronal network formation by measuring electrical firing patterns in microelectrode arrays, and mini-brains are other features monitored in the battery. Performance does well with known positive and negatives, but translating results into a human risk (e.g., thyroid disruption) is one of the challenges that could be brought into the discussion. A positive result in the battery would trigger an alert for targeted testing, although additional information would be sought for complementary target cell types like astrocytes and toxicokinetics in future research.

### *Endocrine disruption*

The European Commission (EC) has sponsored work on new testing and screening methods to identify endocrine disrupting chemicals. This includes efforts on four themes (thyroid disruption, DNT, metabolism, and female reproduction) with emphasis on regulatory usability of the tools/methods. In the European Union (EU), under the Biocidal Products and Plant Protection Products Regulations, a substance is identified as an endocrine disruptor when meeting three criteria (paraphrased) (ECHA/EFSA, 2018): (i) adverse effect in an intact organism or its progeny (e.g., change in morphology, physiology, growth, development, reproduction or life span) that results in impaired function or increased susceptibility to other influences; (ii) has an endocrine mode of action (e.g.,

alters the function(s) of the endocrine system); and (iii) the adverse effect is a consequence of the endocrine mode of action.

AOPs and Integrated Approaches to Testing and Assessment (IATAs) point to potential paths forward to address the regulatory testing with reduced animal use. AOPs provide an analytical construct that describes a sequential chain of causally linked events at different levels of biological organization that lead to an adverse health or ecotoxicological effect (Villeneuve et al., 2014). IATAs are approaches for chemical hazard characterization that rely on an integrated analysis of existing information coupled with the generation of new information using testing strategies (Tollefsen et al., 2014). With regards to DART, AOPs would ideally utilize scientific evidence to bolster confidence in an apical endpoint based on *in vitro* data rather than actual observation from new animal studies. Because AOPs are biologically driven and chemically agnostic, additional data and information are needed from IATAs on *in vivo* extrapolation (IVIVE) and metabolic competency (ADME). Compiling data and information into a user-friendly AOP framework will help navigate differences in regulatory guidelines and policies between countries.

Future research needs include endocrine disruption for systems beyond EATS (estrogen, androgen, thyroid, and steroidogenesis), such as the retinoid signaling pathway (EC, 2018) that is critically important for embryonic development and function of the male and female reproductive systems (Ghyselinck and Duester, 2019). OECD's Endocrine Disruptors Testing and Assessment (EDTA) Advisory Group commissioned a Detailed Review Project (DRP project 4.97b) in its Test Guidelines Programme workplan to focus on the identification of relevant *in vitro* assays and *in vivo* endpoints

for male and female reproductive systems, and on CNS and skeletal development that can identify candidate assays and endpoints to be included in OECD test guidelines or testing strategies to evaluate chemical interactions with the retinoid signaling pathway (Grignard et al., 2020). The recent development of the key characteristics of endocrine disruptors – a uniform, pathway-free basis for searching, organizing, and evaluating mechanistic evidence to support the classification of endocrine disruptors – provides a compound-specific complimentary approach for use in regulatory decisions (La Merrill et al., 2020).

In the US, Congress mandated that the EPA screen pesticides, chemicals, and environmental contaminants for their potential to be endocrine disruptors leading to the EDSPs [https://www.epa.gov/endocrine-disruption/endocrine-disruptor-screening-program-edsp-overview]; however, at a cost of $1M per chemical for the Tier 1 battery on *in vitro* and *in vivo* EDSP tests, and throughput limited to about 50 chemicals per year, it would take an estimated $1B and more than 20 years to complete screening of the approximately 1,150 US registered pesticide active ingredients. This motivates the need for computational approaches to inform the EDSP. Knowledge of the complexities in identifying endocrine disruption and the limited biological coverage provided by any one *in vitro* assay underscores the need to develop a testing framework that combines and integrates data from multiple assays into a battery for screening, prioritization and further evaluation. Two international collaborative projects addressed this problem: CERAPP (Collaborative Estrogen Receptor Activity Prediction Project) (Mansouri et al., 2016) and CoMPARA (Collaborative Modeling Project for Androgen Receptor Activity) (Mansouri et al., 2020). These projects involved over 100 scientists worldwide from academia, industry, and government using available *in vitro* data

from ToxCast/Tox21, QSAR-ready training and prioritization sets, and experimental data curated from the literature. A crowd-sourcing approach was undertaken. Every participant submitted their model for the organizers to evaluate, score, and include in the consensus. The data and models were combined into a consensus approach that was used prospectively for rapid screening of actives and inactives from QSAR-ready structures (32,464 for CERAPP and 55,450 for CoMPARA).

### *Pregnancy exposome*

*In utero* exposure to environmental chemicals can adversely impact pregnancy outcomes and childhood health, but only minimal biomonitoring data exist for most chemicals used in commerce. Efforts are ongoing to address targeted and non-targeted analysis of chemical exposures associated with the human pregnancy exposome (Wang et al., 2018a). The technology uses high-resolution mass spectrometry to profile a 'suspect screen' of chemicals that can be matched qualitatively to reference metabolomic databases and inform targeted assays for confirmation. For example, an analysis of maternal serum samples collected at delivery from 75 pregnant women detected, on average, 56 (range 36-82) suspect environmental chemical metabolites in a study focused mainly on pesticides, phenols, polyfluoro alkyls, phthalates, and unknown classes (Wang et al., 2018a; Wang et al., 2016). Two high use compounds with evidence of human health effects were confirmed in HTS yeast and *C. elegans* assays. Suspect screening in human pregnancy biomonitoring provides a viable method to characterize a broad spectrum of environmental chemicals.

### *Perinatal regulatory needs*

Bioimaging modalities such as micro-CT (computer tomography) have shown promise in teratological study designs (Vasquez et al., 2012). With micro-CT, the fetal skeleton as well as soft tissues can be evaluated in the same subjects across the entire litter (rather than parsing pups out for skeletal or visceral exams). In addition, the evaluation can yield novel quantitative information on tissue mass and volume. Another bioimaging modality, PET (positive emission tomography) scans, enables *in vivo* tracing of physiological changes. For example, adverse effects of exposure to anesthetics such as ketamine or isoflurane administered during late pregnancy can be evaluated by PET scans of the offspring during early postnatal development. FEPPA - a PET labeled ligand that binds to the outer mitochondrial translocator membrane protein (TSPO), is a non-invasive marker for early brain injury that can detect and quantitate microglial activation, foreshadowing neuroinflammation (Ghadery et al., 2019). Other physiological tracers such as neurotransmitters or apoptosis can be evaluated to assess the progression of neurological damage over time in the same animal subjects. These *in vivo* anchors provide endpoints in modeling data from *in vitro* bioactivity profiles. The anesthetic propofol, a gamma aminobutyric acid (GABA) receptor agonist with 1.5 million cases of exposure in newborns and toddlers in the US, targets specific cell types (neurons, not oligodendrocytes or glia) in neurogenic stem cell cultures and the adverse effects can be suppressed with xenon gas (Liu et al., 2020). The *in vitro* optimization protocols can then inform targeted animal studies needed for translatability.

### *TSCA reauthorization*

The Toxic Substances Control Act amended in 2016 (LCSA) explicitly states the need to consider risks from chemical exposures to potentially exposed and susceptible subpopulations that

include infants, children, and pregnant women (Koman et al., 2019; USA, 2016). USEPA's Policy on Evaluating Health Risks to Children has been in place since 1995 and requires consistent and explicit consideration of children's health risks as part of EPA risk assessments (USEPA, 1995). A conceptual framework for assessing health risk of environmental exposures to children addressed the need for a systematic approach for organizing, evaluating, and incorporating the available data on children's susceptibilities in risk assessments and identified at each phase the questions and issues of particular importance for characterizing risks to the developing organism, from conception through organ maturation and the complementary perspectives of toxicokinetics and toxicodynamics (Daston et al., 2004); see also USEPA's framework https://cfpub.epa.gov/ncea/risk/recordisplay.cfm?deid=158363. As was illustrated for several organ systems and end points in that report, an understanding of the timing and cross-species comparison of developmental processes at various stages could inform the hazard characterization processes by identifying potentially unique pathways that could support specific toxicity testing and risk assessment requirements.

USEPA developed a strategic plan to promote the development and implementation of alternative test methods within TSCA (USEPA, 2018). Amended TSCA requires USEPA to develop a process for evaluating the risks of existing chemicals. The first step is prioritization, which culminates in a finding of a chemical substance designated as either high or low priority for risk evaluation. A designation of high priority immediately initiates a risk evaluation by USEPA to determine potential risk to human health and the environment. USEPA document titled "*Procedures for Prioritization of Chemicals for Risk Evaluation Under the Toxic Substances Control Act*" provides the criteria used to prioritize chemicals (USA, 2017). This includes use of reasonably available information to screen the

candidate chemical substances against the following criteria and considerations: the hazard and exposure potential of the chemical substance; persistence and bioaccumulation; potentially exposed or susceptible subpopulations; storage near significant sources of drinking water; conditions of use or significant changes in the conditions of use of the chemical substance; the chemical substance's production volume or significant changes in production volume; and other risk-based criteria that USEPA determines to be relevant to the designation of the chemical substance's priority. The initial identification of 20 high-priority and 20 low-priority chemicals required that at least half be drawn from a 2014 Update to the TSCA Work Plan (USEPA, 2019b). The screening process for identifying chemicals on the 2014 Update to the TSCA Work Plan included using a combination of information available on hazard, exposure, persistence and bioaccumulation characteristics. The USEPA conducted screening level reviews of available information, including information from the Safer Chemicals Ingredients List (SCIL) (USEPA/OCSPP, 2013), to identify the initial set of low priority chemicals. Low-priority substance designations provide an indication of those chemical substances for which the hazard and/or exposure potential is anticipated to be low and do not meet the standard for high-priority, thus are low-priority for risk evaluation at this time. Over the long term, USEPA may explore ways to reduce the size of the inventory to manageable chunks that could be based on potential child exposure, genotoxicity, DART and DNT, or bioaccumulation/persistence (USEPA, 2019b; USEPA/OCSPP, 2013).

## 4. FUTURE NEEDS FOR REGULATORY USE

The capacity to quantitatively link a molecular initiating event (MIE) with the probability of an apical effect (adverse outcome) of regulatory concern is an important goal of AOP elucidation, where

the weight of evidence linking these phenomena through plausible key events can help inform nuances in developmental hazard identification such as cross-species extrapolation, gestational stage, and sex-related differences in a concept recently addressed as 'evolution versus revolution' in DART testing (Scialli et al., 2018). Determining putative MIE targets from *in vitro* profiling is complicated. The nature of chemical-biological interactions varies across *in vitro* bioactivity profiles that may reveal off-target (e.g., drugs) or promiscuous (e.g., chemicals) targets as the concentration increases. Furthermore, whether an event measured *in vitro* represents a MIE or a downstream key event closer to an apical endpoint observed *in vivo* is a consideration worthy of future exploration. The level of uncertainty accompanying decisions made on the basis of an event measured *in vitro* relative to the 'distance along the AOP' separating that key event from the apical endpoint has important mechanistic implications for translatability. As more AOPs are elucidated, it becomes more feasible to identify key event 'hubs' in AOP networks. These nodes may be common to multiple developmental outcomes and can be targeted for *in vitro* assay development (Scialli et al., 2018). Common AOP nodes also suggest that timing/location of chemical exposure is important consideration to resolving the specific apical outcome. Because the hallmark of multicellular organization is the ability of cells to interact with one another, a basic ontology of cell-signal and response pathways for homeostatic and stress-response pathways is needed. The question then becomes one of robustness: how much of a change can the AOP experience, at some or all stages of gestation, before invoking a critical effect and how do we identify the point-of-departure. Teratogenesis is considered a threshold phenomenon at the individual level although looking across a population, and considering susceptibility and co-exposures, the effect of the perturbation may ultimately be manifest as a distribution of likelihoods that the perturbation will result in an adverse effect. Statistical approaches (Paul Friedman et al.,

2020) and dynamical models that address system state dynamics mathematically (Leung et al., 2016) have shown utility in this regard and can be tested using human cell-based microphysiological systems (Knudsen et al., 2017). Various *in silico* tools can assist AOP elucidation for DART, including literature mining, structure-activity read across, and phenotypic barcoding. AOP-based ontologies are a key component to move NAMs forward in predictive developmental toxicology.

ADME considerations can inform what additional studies or data are needed in a decision context. New tools for rapid evaluation of toxicokinetics (httk) can prioritize development of full physiologically-based pharmacokinetic (PBPK) models, but both approaches should give consideration to temporality of exposure, internal concentrations that may be more or less dynamic relative to the time since last exposure, and whether the exposure has persisted long enough to produce steady state kinetics. Transplacental and lactational exposure and individual variability are data gaps that could be addressed with *in vitro* data and *in silico* models. Machine-learning models and structure-activity relationship (SAR) decision trees (Luechtefeld et al., 2018; Wu et al., 2013) can begin to fill in such information across large batteries of assays where a comprehensive pathway- or signature-based strategy for sensitive endpoints in DART might emerge (Scialli et al., 2018). Read across based on chemical structure similarity and ADME information may be important for adjuncts to AOPs across different lifestages and periods of exposure (Wu et al., 2013). Where *in vivo* data is lacking, a positive DART outcome in NAMs can play a role in identification of a developmental hazard; however, a negative response *in vitro* will be difficult to prove, raising the question as to the level of confidence needed for a health-protective evaluation that is as good or better than what can be achieved in a

traditional pregnant animal bioassay. The reference compounds for validation of such *in vitro* assays are extremely important.

## 5. FINAL THOUGHTS

With 50 years of developmental toxicity testing behind us, vast HTS datasets now in hand and substantial knowledge of molecular embryology can begin to show how NAM performance-based models compare for replacement of animal tests. Risk assessment decisions are typically driven by observational toxicology. Alternative assays focus on mechanisms but AOPs can be used to integrate the *in vivo* and *in vitro* and modeling information into one organizational construct. Challenges remain in elucidating relevant mechanistic information and determining appropriate AOP coverage to enable efficient use in risk assessment particularly for early life-stages.

The science of toxicology has advanced significantly in its ability to model complex developmental processes in simpler systems and *in silico*, and in identifying modes of action that lead to developmental toxicity. Mode of action-based HTS methods like ToxCast and toxicogenomics already can detect biological activity known to adversely affect development, and these methods are already being applied to prioritize the large number of chemicals for which human exposure is possible. These methods, put into an AOP framework, have the potential to be a valuable adjunct to, and potential replacement for, *in vivo* testing. To realize this potential, more work needs to be done to better catalog the mode-of-action landscape, and to develop case studies where hypothesis-driven selection of models is used to investigate the potential of agents to affect development. The results can be compared, both qualitatively and quantitatively, to existing data that underlie regulatory

decisions. Although use of NAMs for regulatory decision making is still on the horizon, the conference highlighted novel testing platforms and computational models that cover multiple levels of biological organization, with the unique temporal dynamics of embryonic development, and novel approaches for estimating toxicokinetic parameters essential in supporting *in vitro* to *in vivo* extrapolation. The ultimate goal is to provide decision makers with better, more human-relevant tools for regulatory decisions to protect children's health.

## 6. ACKNOWLEDGEMENTS

FutureTox IV was conceived by the Scientific Liaison Coalition of the Society of Toxicology (SOT) and was held under the auspices of the SOT Contemporary Concepts in Toxicology (CCT) Conferences Committee. The FutureTox IV Organizing Committee co-authored this manuscript. Discussion and Break-out group leads made a significant contribution to developing consensus among workshop participants and in defining challenges and directions for the future. We thank our session chairs and discussion leaders for Plenary Session I: Unique Susceptibilities at Developmental Life Stages, co-chaired by Thad Schug and Dana Dolinoy; Plenary Session II: *In Vitro* Data and *In Silico* Models for Children's Health Research, co-chaired by Nadira De Abrew and Barbara Wetmore; Plenary Session III: Microphysiological Systems and Data Models for Children's Health Issues: Applications for Organ Specific Assessments, co-Chaired by Donna Mendrick and Suzanne Fitzpatrick; and Plenary Session IV: Challenges and Opportunities to Address Regulatory Gaps, chaired by Kimberly White. We thank break-out group leaders for Break Out Group 1: Novel *In vitro* and *In silico* Strategies for Assessing Developmental Health Effects, led by Jill Franzosa, Nicole Kleinstreuer, and Kamel Mansouri. Break Out Group 2: Novel Strategies for Early Lifestage Toxicokinetic and Exposure

Assessment, led by Germaine Louis and Barbara Wetmore; Break Out Group 3: Regulatory Gaps and Consideration of Uncertainty and Variability for Developmental Risk Assessment and Decision-Making, led by Ruthann Rudel and Sue Euling; and Break Out Group 4: DART Mechanisms: Strategies to Target *in Vitro* and i*n Silico* Models to Address Regulatory Needs, led by Oki Noffisat and Susan Makris. The Dine and Dash[board] Breakfast was hosted by Suzanne Fitzpatrick and Monica Linnenbrink. Finally, the success of this conference would not have been possible without the gracious support from the AIM Staff: Marcia Lawson, Amy Willis, and Clarissa Wilson.

## 7. FUNDING and DISCLAIMERS

FutureTox IV Sponsors included Society of Toxicology, Scientific Liaison Coalition, the American Chemistry Council, the National Institute of Environmental Health Sciences, US Environmental Protection Agency, Chemical Safety for Sustainability Research Program (for trainee travel grant), US Food and Drug Administration, Office of the Chief Scientist and the National Center for Toxicological Research, American Academy of Clinical Toxicology, the American College of Toxicology, Johns Hopkins Center for Alternatives to Animal Testing (CAAT), the Society of Toxicologic Pathology, The Teratology Society (now the Society for Birth Defects Research and Prevention), and the Foundation for Chemistry.

DISCLAIMERS: The views expressed in this article are those of the authors and do not necessarily reflect the views or policies of the US Environmental Protection Agency, the US 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. The authors have no conflicts of interest to disclose.

# References

Allard P., Kleinstreuer N.C., Knudsen T.B., Colaiacovo M.P. (2013). A c. Elegans screening platform for the rapid assessment of chemical disruption of germline function. Environ Health Perspect. **121**(6):717-724.

Baker N.C., Sipes N.S., Franzosa J., Belair D.G., Abbott B.D., Judson R.S., Knudsen T.B. (2020). Characterizing cleft palate toxicants using toxcast data, chemical structure, and the biomedical literature. Birth Defects Res. **112**(1):19-39.

Bal-Price A., Crofton K.M., Leist M., Allen S., Arand M., Buetler T., Delrue N., FitzGerald R.E., Hartung T., Heinonen T., Hogberg H., Bennekou S.H., Lichtensteiger W., Oggier D., Paparella M., Axelstad M., Piersma A., Rached E., Schilter B., Schmuck G., Stoppini L., Tongiorgi E., Tiramani M., Monnet-Tschudi F., Wilks M.F., Ylikomi T., Fritsche E. (2015). International stakeholder network (istnet): Creating a developmental neurotoxicity (dnt) testing road map for regulatory purposes. Arch Toxicol. **89**(2):269-287.

Belair D.G., Wolf C.J., Moorefield S.D., Wood C., Becker C., Abbott B.D. (2018). A three-dimensional organoid culture model to assess the influence of chemicals on morphogenetic fusion. Toxicol Sci. **166**(2):394-408.

Boyle C.A., Boulet S., Schieve L.A., Cohen R.A., Blumberg S.J., Yeargin-Allsopp M., Visser S., Kogan M.D. (2011). Trends in the prevalence of developmental disabilities in us children, 1997-2008. Pediatrics. **127**(6):1034-1042.

Browne P., Judson R.S., Casey W.M., Kleinstreuer N.C., Thomas R.S. (2015). Screening chemicals for estrogen receptor bioactivity using a computational model. Environ Sci Technol. **49**(14):8804-8814.

Camacho J., Truong L., Kurt Z., Chen Y.W., Morselli M., Gutierrez G., Pellegrini M., Yang X., Allard P. (2018). The memory of environmental chemical exposure in c. Elegans is dependent on the jumonji demethylases jmjd-2 and jmjd-3/utx-1. Cell Rep. **23**(8):2392-2404.

Ciallella H.L., Zhu H. (2019). Advancing computational toxicology in the big data era by artificial intelligence: Data-driven and mechanism-driven modeling for chemical toxicity. Chem Res Toxicol. **32**(4):536-547.

Cohn B.A., Cirillo P.M., Terry M.B. (2019). Ddt and breast cancer: Prospective study of induction time and susceptibility windows. J Natl Cancer Inst. **111**(8):803-810.

Collins F.S., Gray G.M., Bucher J.R. (2008). Toxicology. Transforming environmental health protection. Science. **319**(5865):906-907.

Daston G., Faustman E., Ginsberg G., Fenner-Crisp P., Olin S., Sonawane B., Bruckner J., Breslin W., McLaughlin T.J. (2004). A framework for assessing risks to children from exposure to environmental agents. Environ Health Perspect. **112**(2):238-256.

Dias A.S., de Almeida I., Belmonte J.M., Glazier J.A., Stern C.D. (2014). Somites without a clock. Science. **343**(6172):791-795.

Dishaw L.V., Macaulay L.J., Roberts S.C., Stapleton H.M. (2014). Exposures, mechanisms, and impacts of endocrine-active flame retardants. Curr Opin Pharmacol. **19**:125-133.

Donovan K.A., An J., Nowak R.P., Yuan J.C., Fink E.C., Berry B.C., Ebert B.L., Fischer E.S. (2018). Thalidomide promotes degradation of sall4, a transcription factor implicated in duane radial ray syndrome. Elife. **7**.

European Parliament, Council of the European Union (2006). Regulation (ec) no 1907/2006 of the european parliament and of the council of 18 december 2006 concerning the registration, evaluation, authorisation and restriction of chemicals (reach), establishing a european chemicals agency, amending directive 1999/45/ec and repealing council regulation (eec) no 793/93 and commission regulation (ec) no 1488/94 as well as council directive 76/769/eec and commission directives 91/155/eec, 93/67/eec, 93/105/ec and 2000/21/ec. [https://eur-lex.Europa.Eu/legal-content/en/txt/?Uri=celex:32006r1907].

European Commission, Directorate-General (2018). Setting priorities for further development and validation of test methods and testing approaches for evaluating endocrine disrupters. Brussels. (https://op.europa.eu/en/publication-detail/-/publication/6b464845-4833-11e8-be1d-01aa75ed71a1.

European Chemical Agency (ECHA) and European Food Safety Authority (EFSA) (2018). Guidance for the identification of endocrine disruptors in the context of regulations (eu) no 528/2012 and (ec)no 1107/2009. EFSA Journal 2018;16(6):5311, doi: 10.2903/j.efsa.2018.5311.

Fritsche, E. (2016). Report of the oecd/efsa workshop on developmental neurotoxicity (dnt): The use of non-animal test. Oecd environ. Heal. Saf. Publications ser. Test. Assess. 242. Report on integrated testing strategies for the identification and evaluation of chemical hazards associated with the developmental neurotoxicity (DNT).

Geng Y., Peterson R.T. (2019). The zebrafish subcortical social brain as a model for studying social behavior disorders. Dis Model Mech. **12**(8).

Ghadery C., Best L.A., Pavese N., Tai Y.F., Strafella A.P. (2019). Pet evaluation of microglial activation in non-neurodegenerative brain diseases. Curr Neurol Neurosci Rep. **19**(7):38.

Ghyselinck N.B., Duester G. (2019). Retinoic acid signaling pathways. Development. **146**(13).

Giedd J. (1999). Brain development, ix: Human brain growth. Am J Psychiatry. **156**(1):4.

Glen C.M., Kemp M.L., Voit E.O. (2019). Agent-based modeling of morphogenetic systems: Advantages and challenges. PLoS Comput Biol. **15**(3):e1006577.

Green D.J., Burnham J.M., Schuette P., Liu X.I., Maas B.M., Yao L., McCune S.K., Chen J., van den Anker J.N., Burckart G.J. (2018). Primary endpoints in pediatric efficacy trials submitted to the us fda. J Clin Pharmacol. **58**(7):885-890.

Grignard E., Hakansson H., Munn S. (2020). Regulatory needs and activities to address the retinoid system in the context of endocrine disruption: The european viewpoint. Reprod Toxicol. **93**:250-258.

Harris S., Hermsen S.A., Yu X., Hong S.W., Faustman E.M. (2015). Comparison of toxicogenomic responses to phthalate ester exposure in an organotypic testis co-culture model and responses observed in vivo. Reprod Toxicol. **58**:149-159.

Hutson M.S., Leung M.C.K., Baker N.C., Spencer R.M., Knudsen T.B. (2017). Computational model of secondary palate fusion and disruption. Chem Res Toxicol. **30**(4):965-979.

Hwang W.Y., Fu Y., Reyon D., Maeder M.L., Kaini P., Sander J.D., Joung J.K., Peterson R.T., Yeh J.R. (2013). Heritable and precise zebrafish genome editing using a crispr-cas system. PLoS One. **8**(7):e68708.

ICCVAM (Interagency Coordinating Committee on the Validation of Alternative Methods) (2018). A strategic roadmap for establishing new approaches to evaluate the safety of chemicals and medical products in the united states. https://dx.doi.org/10.22427/NTP-ICCVAM-ROADMAP2018. Progress report (2018-19) available: https://ntp.niehs.nih.gov/iccvamreport/2019/index.html?utm_source=direct&utm_medium=prod&utm_campaign=ntpgolinks&utm_term=2019iccvamreport. .

Iimura T., Pourquie O. (2007). Hox genes in time and space during vertebrate body formation. Dev Growth Differ. **49**(4):265-275.

Johnson S.B., Blum R.W., Giedd J.N. (2009). Adolescent maturity and the brain: The promise and pitfalls of neuroscience research in adolescent health policy. J Adolesc Health. **45**(3):216-221.

Jones G. (2012). Metabolism and biomarkers of vitamin d. Scand J Clin Lab Invest Suppl. **243**:7-13.

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., Willett C.E., Andersen M.E., Borgert C.J., Coady K.K., Dourson M.L., Fowle J.R., 3rd, Gray L.E., Lamb J.C., Ortego L.S., Schug T.T., Toole C.M., Zorrilla L.M., Kroner O.L., Patterson J., Rinckel L.A., Jones B.R. (2014). Lessons learned, challenges, and opportunities: The u.S. Endocrine disruptor screening program. ALTEX. **31**(1):63-78.

Juberg D.R., Knudsen T.B., Sander M., Beck N.B., Faustman E.M., Mendrick D.L., Fowle J.R., 3rd, Hartung T., Tice R.R., Lemazurier E., Becker R.A., Fitzpatrick S.C., Daston G.P., Harrill A., Hines R.N., Keller D.A., Lipscomb J.C., Watson D., Bahadori T., Crofton K.M. (2017). Futuretox iii: Bridges for translation. Toxicol Sci. **155**(1):22-31.

Kleinstreuer N., Dix D., Rountree M., Baker N., Sipes N., Reif D., Spencer R., Knudsen T. (2013). A computational model predicting disruption of blood vessel development. PLoS Comput Biol. **9**(4):e1002996.

Knight G.T., Lundin B.F., Iyer N., Ashton L.M., Sethares W.A., Willett R.M., Ashton R.S. (2018). Engineering induction of singular neural rosette emergence within hpsc-derived tissues. Elife. **7**.

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., Watkins P.B., Whelan M. (2015). Futuretox ii: In vitro data and in silico models for predictive toxicology. Toxicol Sci. **143**(2):256-267.

Knudsen T.B., Klieforth B., Slikker W., Jr. (2017). Programming microphysiological systems for children's health protection. Exp Biol Med (Maywood). **242**(16):1586-1592.

Koman P.D., Singla V., Lam J., Woodruff T.J. (2019). Population susceptibility: A vital consideration in chemical risk evaluation under the lautenberg toxic substances control act. PLoS Biol. **17**(8):e3000372.

La Merrill M.A., Vandenberg L.N., Smith M.T., Goodson W., Browne P., Patisaul H.B., Guyton K.Z., Kortenkamp A., Cogliano V.J., Woodruff T.J., Rieswijk L., Sone H., Korach K.S., Gore A.C., Zeise L., Zoeller R.T. (2020). Consensus on the key characteristics of endocrine-disrupting chemicals as a basis for hazard identification. Nat Rev Endocrinol. **16**(1):45-57.

Lee J.S., Romero R., Han Y.M., Kim H.C., Kim C.J., Hong J.S., Huh D. (2016). Placenta-on-a-chip: A novel platform to study the biology of the human placenta. J Matern Fetal Neonatal Med. **29**(7):1046-1054.

Leung M.C., Hutson M.S., Seifert A.W., Spencer R.M., Knudsen T.B. (2016). Computational modeling and simulation of genital tubercle development. Reprod Toxicol. **64**:151-161.

Lippmann E.S., Williams C.E., Ruhl D.A., Estevez-Silva M.C., Chapman E.R., Coon J.J., Ashton R.S. (2015). Deterministic hox patterning in human pluripotent stem cell-derived neuroectoderm. Stem Cell Reports. **4**(4):632-644.

Liu F., Liu S., Patterson T.A., Fogle C., Hanig J.P., Wang C., Slikker W., Jr. (2020). Protective effects of xenon on propofol-induced neurotoxicity in human neural stem cell-derived models. Mol Neurobiol. **57**(1):200-207.

Luechtefeld T., Marsh D., Rowlands C., Hartung T. (2018). Machine learning of toxicological big data enables read-across structure activity relationships (rasar) outperforming animal test reproducibility. Toxicol Sci. **165**(1):198-212.

Lundby Z., Camacho J., Allard P. (2016). Fast functional germline and epigenetic assays in the nematode caenorhabditis elegans. Methods Mol Biol. **1473**:99-107.

Maenner M.J., Shaw K.A., Baio J., EdS, Washington A., Patrick M., DiRienzo M., Christensen D.L., Wiggins L.D., Pettygrove S., Andrews J.G., Lopez M., Hudson A., Baroud T., Schwenk Y., White T., Rosenberg C.R., Lee L.C., Harrington R.A., Huston M., Hewitt A., PhD, Esler A., Hall-Lande J., Poynter J.N., Hallas-Muchow L., Constantino J.N., Fitzgerald R.T., Zahorodny W., Shenouda J., Daniels J.L., Warren Z., Vehorn A., Salinas A., Durkin M.S., Dietz P.M. (2020). Prevalence of autism spectrum disorder among children aged 8 years - autism and developmental disabilities monitoring network, 11 sites, united states, 2016. MMWR Surveill Summ. **69**(4):1-12.

Mahony C., Ashton R.S., Birk B., Boobis A.R., Cull T., Daston G.P., Ewart L., Knudsen T.B., Manou I., Maurer-Stroh S., Margiotta-Casaluci L., Muller B.P., Nordlund P., Roberts R.A., Steger-Hartmann T., Vandenbossche E., Viant M.R., Vinken M., Whelan M., Zvonimir Z., Cronin M.T.D. (2020). New ideas for non-animal approaches to predict repeated-dose systemic toxicity: Report from an epaa blue sky workshop. Regul Toxicol Pharmacol. **114**:104668.

Mansouri K., Abdelaziz A., Rybacka A., Roncaglioni A., Tropsha A., Varnek A., Zakharov A., Worth A., Richard A.M., Grulke C.M., Trisciuzzi D., Fourches D., Horvath D., Benfenati E., Muratov E., Wedebye E.B., Grisoni F., Mangiatordi G.F., Incisivo G.M., Hong H., Ng H.W., Tetko I.V., Balabin I., Kancherla J., Shen J., Burton J., Nicklaus M., Cassotti M., Nikolov N.G., Nicolotti O., Andersson P.L., Zang Q., Politi R., Beger R.D., Todeschini R., Huang R., Farag S., Rosenberg S.A., Slavov S., Hu X., Judson R.S. (2016). Cerapp: Collaborative estrogen receptor activity prediction project. Environ Health Perspect. **124**(7):1023-1033.

Mansouri K., Kleinstreuer N., Abdelaziz A.M., Alberga D., Alves V.M., Andersson P.L., Andrade C.H., Bai F., Balabin I., Ballabio D., Benfenati E., Bhhatarai B., Boyer S., Chen J., Consonni V., Farag S., Fourches D., Garcia-Sosa A.T., Gramatica P., Grisoni F., Grulke C.M., Hong H., Horvath D., Hu X., Huang R., Jeliazkova N., Li J., Li X., Liu H., Manganelli S., Mangiatordi G.F., Maran U., Marcou G., Martin T., Muratov E., Nguyen D.T., Nicolotti O., Nikolov N.G., Norinder U., Papa E., Petitjean M., Piir G., Pogodin P., Poroikov V., Qiao X., Richard A.M., Roncaglioni A.,

Ruiz P., Rupakheti C., Sakkiah S., Sangion A., Schramm K.W., Selvaraj C., Shah I., Sild S., Sun L., Taboureau O., Tang Y., Tetko I.V., Todeschini R., Tong W., Trisciuzzi D., Tropsha A., Van Den Driessche G., Varnek A., Wang Z., Wedebye E.B., Williams A.J., Xie H., Zakharov A.V., Zheng Z., Judson R.S. (2020). Compara: Collaborative modeling project for androgen receptor activity. Environ Health Perspect. **128**(2):27002.

Markov D.A., Lu J.Q., Samson P.C., Wikswo J.P., McCawley L.J. (2012). Thick-tissue bioreactor as a platform for long-term organotypic culture and drug delivery. Lab Chip. **12**(21):4560-4568.

Masjosthsumann S., Blum J., Bartmann K., Stürzl L.-C., Keßel E., Förster N., Dönmez A., Klose J., Pahl M., Waldmann T., Bendt F., Dolde X., Holzer A.-K., Hübenthal U., Mosig A., Leist M., Fritsche E. (2020). Establishment of an a priori protocol for the implementation and interpretation of an in-vitro testing battery for the assessment of developmental neurotoxicity. . EFSA Journal (in press).

McCollum C.W., Conde-Vancells J., Hans C., Vazquez-Chantada M., Kleinstreuer N., Tal T., Knudsen T., Shah S.S., Merchant F.A., Finnell R.H., Gustafsson J.A., Cabrera R., Bondesson M. (2017). Identification of vascular disruptor compounds by analysis in zebrafish embryos and mouse embryonic endothelial cells. Reprod Toxicol. **70**:60-69.

Momper J.D., Mulugeta Y., Burckart G.J. (2015). Failed pediatric drug development trials. Clin Pharmacol Ther. **98**(3):245-251.

Morgan M.M., Johnson B.P., Livingston M.K., Schuler L.A., Alarid E.T., Sung K.E., Beebe D.J. (2016). Personalized in vitro cancer models to predict therapeutic response: Challenges and a framework for improvement. Pharmacol Ther. **165**:79-92.

National Academies of Sciences, Engineering,Medicine (2017). Using 21st century science to improve risk-related evaluations. National Academies Press, Washington D.C.

. https://doi.org/10.17226/24635.

National Research Council (2000). Scientific frontiers in developmental toxicology and risk assessment. National Academies Press, Washington D.C.

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 D.C.

National Research Council (2012). Exposure science in the 21st century: A vision and a strategy. . National Academies Press, Washington D.C.

Pamies D., Barreras P., Block K., Makri G., Kumar A., Wiersma D., Smirnova L., Zang C., Bressler J., Christian K.M., Harris G., Ming G.L., Berlinicke C.J., Kyro K., Song H., Pardo C.A., Hartung T., Hogberg H.T. (2017). A human brain microphysiological system derived from induced pluripotent stem cells to study neurological diseases and toxicity. ALTEX. **34**(3):362-376.

Paul Friedman K., Gagne M., Loo L.H., Karamertzanis P., Netzeva T., Sobanski T., Franzosa J.A., Richard A.M., Lougee R.R., Gissi A., Lee J.J., Angrish M., Dorne J.L., Foster S., Raffaele K., Bahadori T., Gwinn M.R., Lambert J., Whelan M., Rasenberg M., Barton-Maclaren T., Thomas R.S. (2020). Utility of in vitro bioactivity as a lower bound estimate of in vivo adverse effect levels and in risk-based prioritization. Toxicol Sci. **173**(1):202-225.

Pearce R.G., Setzer R.W., Strope C.L., Wambaugh J.F., Sipes N.S. (2017). Httk: R package for high-throughput toxicokinetics. J Stat Softw. **79**(4):1-26.

Rowlands J.C., Sander M., Bus J.S., FutureTox Organizing C. (2014). Futuretox: Building the road for 21st century toxicology and risk assessment practices. Toxicol Sci. **137**(2):269-277.

Sagiv S.K., Bruno J.L., Baker J.M., Palzes V., Kogut K., Rauch S., Gunier R., Mora A.M., Reiss A.L., Eskenazi B. (2019). Prenatal exposure to organophosphate pesticides and functional neuroimaging in adolescents living in proximity to pesticide application. Proc Natl Acad Sci U S A. **116**(37):18347-18356.

Schaner C.E., Kelly W.G. 2006. Germline chromatin. Wormbook: The online review of c. Elegans biology [internet]. WormBook.

Schmuck M.R., Temme T., Dach K., de Boer D., Barenys M., Bendt F., Mosig A., Fritsche E. (2017). Omnisphero: A high-content image analysis (hca) approach for phenotypic developmental neurotoxicity (dnt) screenings of organoid neurosphere cultures in vitro. Arch Toxicol. **91**(4):2017-2028.

Scialli A.R., Daston G., Chen C., Coder P.S., Euling S.Y., Foreman J., Hoberman A.M., Hui J., Knudsen T., Makris S.L., Morford L., Piersma A.H., Stanislaus D., Thompson K.E. (2018). Rethinking developmental toxicity testing: Evolution or revolution? Birth Defects Res. **110**(10):840-850.

Sturla S.J., Boobis A.R., FitzGerald R.E., Hoeng J., Kavlock R.J., Schirmer K., Whelan M., Wilks M.F., Peitsch M.C. (2014). Systems toxicology: From basic research to risk assessment. Chem Res Toxicol. **27**(3):314-329.

Thomas R.S., Bahadori T., Buckley T.J., Cowden J., Deisenroth C., Dionisio K.L., Frithsen J.B., Grulke C.M., Gwinn M.R., Harrill J.A., Higuchi M., Houck K.A., Hughes M.F., Hunter E.S., Isaacs K.K., Judson R.S., Knudsen T.B., Lambert J.C., Linnenbrink M., Martin T.M., Newton S.R., Padilla S., Patlewicz G., Paul-Friedman K., Phillips K.A., Richard A.M., Sams R., Shafer T.J., Setzer R.W., Shah I., Simmons J.E., Simmons S.O., Singh A., Sobus J.R., Strynar M., Swank A., Tornero-Valez R., Ulrich E.M., Villeneuve D.L., Wambaugh J.F., Wetmore B.A., Williams A.J. (2019). The next generation blueprint of computational toxicology at the u.S. Environmental protection agency. Toxicol Sci. **169**(2):317-332.

Tollefsen K.E., Scholz S., Cronin M.T., Edwards S.W., de Knecht J., Crofton K., Garcia-Reyero N., Hartung T., Worth A., Patlewicz G. (2014). Applying adverse outcome pathways (aops) to support integrated approaches to testing and assessment (iata). Regul Toxicol Pharmacol. **70**(3):629-640.

Truskey G.A. (2018). Human microphysiological systems and organoids as in vitro models for toxicological studies. Front Public Health. **6**:185.

United States, Public Law 114-182 (2016). [https://www.Congress.Gov/114/plaws/publ182/plaw-114publ182.Pdf] and [https://www.Epa.Gov/assessing-and-managing-chemicals-under-tsca/alternative-test-methods-and-strategies-reduce].

Federal Register (2017). Procedures for prioritization of chemicals for risk evaluation under the toxic substances control act. Docket identification (id) number epa-hq-oppt-2016-0636.

United States Environmental Protection Agency Tsca chemical substance inventory. https://www.epa.gov/tsca-inventory/about-tsca-chemical-substance-inventory (accessed June 2020).

United States Environmental Protection Agency (1995). (reaffirmation 2013 and 2018) epa's policy on evaluating risk to children. Https://www.Epa.Gov/children/epas-policy-evaluating-risk-children (accessed june 2020). https://www.epa.gov/tsca-inventory/about-tsca-chemical-substance-inventory (accessed June 2020).

U.S. Environmental Protection Agency (EPA) (2015). Issue paper on physiological and behavioral changes in pregnant and lactating women and available exposure factors. National Center for Environmental Assessment, Washington, D.C. EPA/600/R-15/169. Available from the National Technical Information Service, Springfield, VA and online at http://www.epa.gov/ncea.

U.S. Environmental Protection Agency (EPA) (2018). Strategic plan to promote the development and implementation of alternative test methods within the tsca program. . https://www.epa.gov/sites/production/files/2018-06/documents/epa_alt_strat_plan_6-20-18_clean_final.pdf.

United States Environmental Protection Agency (2019a). Memo directive to prioritize efforts to reduce animal testing. https://www.epa.gov/research/administrator-memo-prioritizing-efforts-reduce-animal-testing-september-10-2019.

U.S. Environmental Protection Agency (EPA) (2019b). New approach methods work plan: Reducing use of animals in chemical testing. https://www.epa.gov/chemical-research/epa-new-approach-methods-work-plan-reducing-use-animals-chemical-testing#:~:text=EPA's%20New%20Approach%20Methods%20(NAMs,human%20health%20and%20the%20environment (accessed December 2020).

U.S. Environmental Protection Agency (EPA) (2013). Safer chemical ingredients list. Data and tools. . https://www.epa.gov/saferchoice/safer-ingredients (accessed June 2020).

Vasquez S., Shah N., Hoberman A. 2012. Small animal imaging and examination by micro-ct. Teratogenicity testing: Methods and protocols. Paul C. Barrow (ed.), Methods in Molecular Biology 947: 223-232.

Villeneuve D.L., Crump D., Garcia-Reyero N., Hecker M., Hutchinson T.H., LaLone C.A., Landesmann B., Lettieri T., Munn S., Nepelska M., Ottinger M.A., Vergauwen L., Whelan M. (2014). Adverse outcome pathway (aop) development i: Strategies and principles. Toxicol Sci. **142**(2):312-320.

Wang A., Gerona R.R., Schwartz J.M., Lin T., Sirota M., Morello-Frosch R., Woodruff T.J. (2018a). A suspect screening method for characterizing multiple chemical exposures among a demographically diverse population of pregnant women in san francisco. Environ Health Perspect. **126**(7):077009.

Wang A., Padula A., Sirota M., Woodruff T.J. (2016). Environmental influences on reproductive health: The importance of chemical exposures. Fertil Steril. **106**(4):905-929.

Wang Y., Liu M., Johnson S.B., Yuan G., Arriba A.K., Zubizarreta M.E., Chatterjee S., Nagarkatti M., Nagarkatti P., Xiao S. (2019). Doxorubicin obliterates mouse ovarian reserve through both primordial follicle atresia and overactivation. Toxicol Appl Pharmacol. **381**:114714.

Wang Y., Liu M., Zhang J., Liu Y., Kopp M., Zheng W., Xiao S. (2018b). Multidrug resistance protein 1 deficiency promotes doxorubicin-induced ovarian toxicity in female mice. Toxicol Sci. **163**(1):279-292.

Watson D.E., Hunziker R., Wikswo J.P. (2017). Fitting tissue chips and microphysiological systems into the grand scheme of medicine, biology, pharmacology, and toxicology. Exp Biol Med (Maywood). **242**(16):1559-1572.

Wegner S., Hong S., Yu X., Faustman E.M. (2013). Preparation of rodent testis co-cultures. Curr Protoc Toxicol. **Chapter 16**:Unit 16 10.

Wetmore B.A., Allen B., Clewell H.J., 3rd, Parker T., Wambaugh J.F., Almond L.M., Sochaski M.A., Thomas R.S. (2014). Incorporating population variability and susceptible subpopulations into dosimetry for high-throughput toxicity testing. Toxicol Sci. **142**(1):210-224.

Wetmore B.A., Wambaugh J.F., Allen B., Ferguson S.S., Sochaski M.A., Setzer R.W., Houck K.A., Strope C.L., Cantwell K., Judson R.S., LeCluyse E., Clewell H.J., Thomas R.S., Andersen M.E. (2015). Incorporating high-throughput exposure predictions with dosimetry-adjusted in vitro bioactivity to inform chemical toxicity testing. Toxicol Sci. **148**(1):121-136.

Woodruff T.J., Zota A.R., Schwartz J.M. (2011). Environmental chemicals in pregnant women in the united states: Nhanes 2003-2004. Environ Health Perspect. **119**(6):878-885.

Wu S., Fisher J., Naciff J., Laufersweiler M., Lester C., Daston G., Blackburn K. (2013). Framework for identifying chemicals with structural features associated with the potential to act as developmental or reproductive toxicants. Chem Res Toxicol. **26**(12):1840-1861.

Xiao S., Coppeta J.R., Rogers H.B., Isenberg B.C., Zhu J., Olalekan S.A., McKinnon K.E., Dokic D., Rashedi A.S., Haisenleder D.J., Malpani S.S., Arnold-Murray C.A., Chen K., Jiang M., Bai L., Nguyen C.T., Zhang J., Laronda M.M., Hope T.J., Maniar K.P., Pavone M.E., Avram M.J., Sefton E.C., Getsios S., Burdette J.E., Kim J.J., Borenstein J.T., Woodruff T.K. (2017). A microfluidic culture model of the human reproductive tract and 28-day menstrual cycle. Nat Commun. **8**:14584.

Zhong X., Harris G., Smirnova L., Zufferey V., Sa R., Baldino Russo F., Baleeiro Beltrao Braga P.C., Chesnut M., Zurich M.G., Hogberg H.T., Hartung T., Pamies D. (2020). Antidepressant paroxetine exerts developmental neurotoxicity in an ipsc-derived 3d human brain model. Front Cell Neurosci. **14**:25.