Article

SCREENING CHEMICALS FOR ESTROGEN RECEPTOR BIOACTIVITY USING A COMPUTATIONAL MODEL

Patience Browne, Richard S. Judson, Warren Casey, Nicole Kleinstreuer, and Russell S. Thomas

Environ. Sci. Technol., **Just Accepted Manuscript** • DOI: 10.1021/acs.est.5b02641 • Publication Date (Web): 12 Jun 2015

SCREENING CHEMICALS FOR ESTROGEN RECEPTOR BIOACTIVITY USING A COMPUTATIONAL MODEL

Patience Browne[1*], Richard S. Judson[2], Warren Casey[3], Nicole Kleinstreuer[4], Russell S. Thomas[2]

[1]US EPA, Office of Chemical Safety and Pollution Prevention, Washington DC; [2]US EPA, Office of Research and Development, Research Triangle Park, NC; [3]National Toxicology Program, Interagency Center for the Evaluation of Alternative Toxicological Methods, Research Triangle Park, NC; [4]ILS/NICEATM, Research Triangle Park, NC

The views expressed in this paper are those of the authors and do not necessarily reflect the views or policies of the US Environmental Protection Agency.

*Address correspondence to P. Browne, US EPA, Office of Science Coordination and Policy, 1200 Pennsylvania Ave NW (7203M), Washington, DC 20460, Telephone: 202-564-8474, Fax: 202-564-8482, E-mail: browne.patience@epa.gov.

Abstract

The U.S. Environmental Protection Agency (EPA) is considering high-throughput and computational methods to evaluate the endocrine bioactivity of environmental chemicals. Here we describe a multi-step, performance-based validation of new methods and demonstrate these new tools are sufficiently robust to be used in the Endocrine Disruptor Screening Program (EDSP). Results from 18 estrogen receptor (ER) ToxCast high-throughput screening assays were integrated into a computational model that can discriminate bioactivity from assay-specific interference and cytotoxicity. Model scores range from 0 (no activity) to 1 (bioactivity of 17β-estradiol). ToxCast ER model performance was evaluated for reference chemicals, as well as results of EDSP Tier 1 screening assays in current practice. ToxCast ER model accuracy was 86 to 93% when compared to reference chemicals, and predicted results of EDSP Tier 1 guideline and other uterotrophic studies with 84 to 100% accuracy. Performance of high-throughput assays and ToxCast ER model predictions demonstrates these methods correctly identify active and inactive reference chemicals, provide a measure of relative ER bioactivity, and rapidly identify chemicals with potential endocrine bioactivities for additional screening and testing. EPA is accepting ToxCast ER model data for 1812 chemicals as alternatives for EDSP Tier 1 ER binding, ER transactivation, and uterotrophic assays.

Introduction

The US Environmental Protection Agency's (EPA) Endocrine Disruptor Screening Program (EDSP) was established in 1999 for the purpose of evaluating potential risk of endocrine disruption in humans and wildlife from exposure to pesticide chemicals and drinking water contaminants. To screen environmental chemicals for potential endocrine bioactivity, EPA developed a battery of five *in vitro* and six *in vivo* Tier 1 screening assays[1]. To test for potential endocrine disruption, EPA developed *in vivo* multigenerational Tier 2 tests that include apical endpoints to identify adverse effects, and establish quantitative dose response relationships[2]. In 2009, EPA published a final list of 67 pesticide chemicals (List 1) and issued EDSP Tier 1 test orders on these chemicals[3]. Fifteen List 1 chemicals were voluntarily withdrawn from the pesticide market. EPA is currently reviewing results of EDSP Tier 1 screening assays, along with other scientifically relevant information, and developing

weight of evidence evaluations of potential endocrine bioactivity with a determination of further testing that may be required for the remaining 52 of the 67 chemicals. A second list of pesticide and high production volume chemicals (List 2) proposed for Tier 1 screening was published in June 2013[4]; however, test orders have yet to be issued. The remaining EDSP universe of pesticide chemicals and drinking water contaminants includes approximately 10,000 environmental chemicals to be screened for potential endocrine bioactivity in humans and wildlife[5].

In response to the US National Academy of Sciences report, Toxicity Testing in the 21st Century[6], and the US President's 2012 proposed budget[7], EPA began a multi-year transition from existing EDSP test methods towards utilizing more rapid, cost-effective, computational models and high-throughput assays. The transition to using computational toxicology approaches to prioritize and screen thousands of EDSP chemicals has been outlined by the Agency in two strategic planning documents[5, 8]. However, to use new computational toxicology approaches in the existing EDSP screening and testing framework, new methods must be validated and perform as well or better than existing methods currently in practice.

A variety of pesticides and environmental chemicals act as estrogen receptor (ER) agonists[9-12]. Although the scope of the program was expanded to consider androgen and thyroid active environmental chemicals, the EDSP was originally established in response to statutory mandates in the Federal Food Quality Protection and Safe Water Drinking Acts compelling EPA to evaluate potential xenoestrogens. As a result, the Tier 1 screening battery was weighted towards assays that detect potential ER interactions. EDSP Tier 1 assay endpoints that measure interaction with the ER only measure agonism and environmental chemicals that act through the ER are primarily expected to act as agonists, therefore we focused this demonstration using computational toxicology tools to evaluate ER agonist bioactivity. Future work will address estrogen antagonism and other endocrine pathways.

In this manuscript, we present a performance-based approach to validate computational toxicology tools for evaluating a chemical's potential ER agonist bioactivity and as alternative data for existing EDSP Tier 1 *in vitro* and *in vivo* assays. High-throughput *in vitro* screening assay data from EPA's ToxCast program[13] were integrated into a computational model of the ER pathway, and model performance was compared with *in vitro* and *in vivo* reference chemicals identified from the peer-reviewed literature[14]. A variety of publications have previously described the ToxCast program[15, 16], the validation of high-throughput screening assays[17], and endocrine-relevant ToxCast assays[18, 19]; however, this paper is the first description of using ToxCast high-throughput screening data as an alternative for regulatory guideline studies. Integrating 21st century toxicology in the EDSP allows for rapid evaluation of potential endocrine bioactivity of thousands of chemicals to which humans and wildlife may be exposed and this approach is consistent with recommendations of the 2007 NAS report[6] and EPA's strategic plan for evaluating the toxicity of chemicals[8], specifically to: 1) provide broad coverage of chemicals examined; 2) reduce the cost and time of toxicity testing; 3) reduce animal use; and 4) develop a robust scientific basis for assessing health effects of environmental agents.

Methods

In Vitro Assays

Details of the *in vitro* assays are described on EPA's ToxCast website[13] and in a variety of publications[15, 20]. Briefly, potential ER bioactivity was measured in 18 high-throughput *in vitro* assays run in EPA's ToxCast

program. The suite of high throughput assays measure the molecular initiating event (*i.e.,* receptor binding), in addition to several key events (*e.g.*, receptor dimerization, DNA binding, transactivation, gene expression, and cell proliferation) in an adverse outcome pathway. The 18 ER assays include three cell-free biochemical radioligand ER binding assays[21, 22]; three protein complementation assays that measure formation of ER dimers and test for activity against ERα and ERβ, each measured at two time points; an assay measuring interaction of the mature transcription factor with DNA at two time points; two reporter gene assays measuring RNA transcript levels[23]; two assays measuring reporter protein levels[24]; an ER-sensitive cell proliferation assay[25]; and two transactivation antagonist assays[24] (Table 1).

This combination of biochemical and cell-based *in vitro* assays relies on different technologies and probes different key events in the ER signaling pathway (Table 1). Though the assays are primarily human proteins and/or cell types, the suite of 18 assays include human, murine, and bovine ER binding assays and ER pathway interactions in a variety of human tissue types (Table 1). Every *in vitro* assay is potentially subject to technology-specific interference (*e.g.*, chemicals that denature the receptor protein, are luminescent, are cytotoxic, etc.) that can be mistakenly interpreted as bioactivity. However, combining data from multiple assays and integrating data in a network model of the entire ER pathway allows for detection of false positives and a more confident assessment of the "true" *in vitro* estrogenic bioactivity of the tested chemical.

Concentration Response Analysis and Computational Modeling

Chemicals were run in concentration-response format in all assays except the cell-free binding assays. Cell-free competitive binding assays were initially run at a single screening concentration (25 μM), and if the test chemical was active in the assay (*i.e.,* radioligand was displaced), the assay was run in concentration-response format. All *in vitro* assays except the assays measuring RNA transcript were normalized to 17β-estradiol. RNA transcript data were normalized as a fold-change from the solvent (DMSO) control. Concentration-response data from *in vitro* assays were fit to three models that included a four parameter Hill model, a modified Hill model with gain-loss at high concentrations, or a constant (no concentration-response) model[26]. The best model was statistically selected using the Akaike Information Criteria value. All concentration-response data were analyzed using the ToxCast data analysis pipeline, which automates the processes of baseline correction, normalization, curve-fitting, hit-calling and detection of a variety of potential confounders[27]. To integrate results of the different *in vitro* assays, concentration-response curves were generated for each assay across 14 concentrations from 0.01-100 μM.

The concentration-response curves for all 18 assays were included in a computational model, referred to here as the ToxCast ER model for bioactivity[14]. The computational model integrates data from the 18 *in vitro* assays measuring ER agonist and antagonist responses in an unweighted manner, while subtracting background and other non-specific assay interference including cytotoxicity. The model output includes separate agonist and antagonist area under the curve (AUC) scores, though only the agonist response was considered in this analysis. The ER agonist AUC model scores range from 0 (no activity) to 1 (bioactivity of 17β-estradiol). The computational model is a simple linear additive model and assumes that the value for a given concentration is a linear sum of the contributions from the ER interaction (*e.g.,* receptor binding, transactivition, mRNA production, etc.) measured in each assay. Each assay contributes equally to the overall score for ER pathway activity if there is direct molecular interaction between a chemical and an assay, and if the assay is not hit by the chemical it provides no contribution to the score. Therefore, the model assumes lossless transmission of signals from the key events in the signaling pathway to the assays. For each chemical-concentration pair, a constrained least-squares minimization approach is used to reconcile the predicted assay values and the measured values, taking

into account both potency and efficacy (mathematical details and a complete description of the model can be found in EPA 2015)[14].

Model scores were truncated at values < 0.001, considered to have no ER bioactivity and given a score of 0, as a value <0.001 implies an concentration required to elicit 50% of the maximal response (AC_{50}) greater than 10 millimolar which is several orders of magnitude greater than the highest concentrations tested in ToxCast assays. ToxCast ER agonist scores $\geq$ 0.1 were considered positive; a model score of 0.1 equates to an AC_{50} of about 100 µM and approximates the upper limit of bioactivity detected in this approach. Model scores of 0.1> AUC >0 were considered inconclusive for this validation because these chemicals were active in only one or two ToxCast ER assays and activity was limited to the highest concentrations tested. Activity in only a few of the 18 assays may be due to differences (and thus differential sensitivities) in the various *in vitro* assays, though all chemicals were tested up to at least 100 µM. It is unlikely that environmental exposures to most chemicals would result in an internal dose at this level, and therefore such limited activity and low potency (*i.e.*, interaction with the ER occurring only at high concentration of test chemical) has questionable relevance to *in vivo* bioactivity. We evaluated model performance against all chemicals and after excluding chemicals with inconclusive scores.

Performance-Based Validation of the ToxCast ER Model

To assess the strengths and limitations of the ToxCast ER model, we adopted a performance-based validation approach consistent with the Organization for Economic and Cooperation and Development (OECD) conceptual framework for assessing potential endocrine disrupting chemicals[28]. In principle, this method can be used to assess the applicability of any test method or set of methods that meets defined performance standards. The ToxCast ER model was validated using two approaches: comparison of ToxCast ER model scores against sets of reference chemicals with independently confirmed ER bioactivity in a validated test method, and comparison of ToxCast ER model scores with results of EDSP Tier 1 guideline assays currently used to screen chemicals for endocrine bioactivity.

In vitro ER reference chemicals, previously identified using multiple validated low throughput *in vitro* ER assays, were identified by the Interagency Coordinating Committee on the Validation of Alternative Test Methods (ICCVAM) and OECD[29] for the express purpose of validating novel *in vitro* assays. Forty chemicals (28 agonists of differing potencies indicated by a range in AC_{50} and 12 inactive chemicals) were selected for reproducible results in *in vitro* ER binding and transactivation assays, and to include a diverse set of chemical structures (Table 2)[29, 30]. All *in vitro* reference chemicals were run in the 18 high-throughput ToxCast ER assays and the resulting ToxCast ER model scores for agonist bioactivity were compared with results anticipated from low or medium throughput *in vitro* assays[29].

In vivo reference chemicals were established from a literature search of short-term rodent assays that were methodologically similar to the OECD[31] and EDSP Tier 1 battery uterotrophic[32] assays. A comprehensive search and review of uterotrophic studies published in peer-reviewed literature was performed[14]. Chemical name and chemical abstract services registry number (CASRN) were used to search PubMed, the EPA's Aggregated Computational Toxicology Resource (ACToR)[33], and the US Food and Drug Administration's Endocrine Disruptor Knowledge Base (EDKB)[34] for the 1812 chemicals run in the ToxCast *in vitro* ER assays with "uterotrophic assay", "uterotrophic", "uterotropic", and "uterine weight" as modifier terms. Articles identified were reviewed for methodological consistency with the EDSP uterotrophic assay guidelines[32] based on: 1) age and species of animals used (immature rat or ovariectomized mouse or rat); 2) number of animals per treatment group; 3) number of treatment groups; 4) route of chemical administration; 5) length of dosing; and 6) time of necropsy.

Over 1000 articles were identified, entered into a database, and independently reviewed by two scientists. Of the articles identified, 442 studies of 103 chemicals met all six minimum criteria[14], were considered "guideline-like", and were used in this analysis.

Chemical data from guideline-like uterotrophic studies were considered with two levels of stringency. First, a chemical was considered positive for potential *in vivo* ER agonist bioactivity if a significant increase in uterine weight among treated animals was reported and negative if no significant increase in uterine weight was reported in any guideline-like study. Second, a subset of chemicals with reproducible results from two or more independent, guideline-like studies were referred to as *in vivo* reference chemicals. Chemicals that resulted in a significant increase in uterine weight in two or more studies were considered active reference chemicals, while those chemicals that showed negative results in all studies (two or more) were considered inactive reference chemicals (Table 3).

ToxCast ER model agonist scores were compared with results of the EDSP Tier 1 assays that directly assess a test chemicals' ER bioactivity. The Tier 1 *in vitro* ER binding assay[35] uses rat uterine cytosol (primarily ERα) and cannot distinguish between potential agonist or antagonist bioactivity. The competitive binding assay measures test chemical displacement of radioligand ([^{3}H]17β-estradiol) from the ER across a range of concentrations in three independent runs. Results of the assay are "positive" if the test chemical displaces >50% of radioligand (and Log(IC_{50}) is calculated), "equivocal" if test chemical displaces <50% but >25% of radioligand, and "negative" if test chemical displaces <25% of radioligand[35]. The Tier 1 *in vitro* Estrogen Receptor Transcriptional Activation (ERTA) assay[36] measures chemiluminescence in response to an ERα-mediated increase in luciferase gene expression (*i.e.,* agonist activity). A test chemical is "positive" if the maximum response induced by the test chemical is $\geq$ 10% of the maximum response of the positive control (17β-estradiol; RPC_{max}) in at least two of three assay runs (*i.e.,* $RPC_{max} \geq 10$). If the test chemical fails to achieve at least 10% of the response of the positive control, a negative response is recorded for the test chemical. The Tier 1 uterotrophic assay[32] is a short-term, *in vivo* assay designed to detect exogenous estrogen agonist activity indicated by an increase in uterine weight in animals in prepubertal or ovariectomized rodents, in which the hypothalamic-pituitary-gonadal axis is not functional.

Application of ToxCast ER model to EDSP Chemicals

ER model agonist scores were examined for 1812 chemicals evaluated in all 18 ToxCast ER assays. These chemicals include 62 EDSP List 1 chemicals for which EPA issued Tier 1 test orders, and 57 List 2 chemicals identified by EPA as candidates to receive the next group of Tier 1 test orders. The 1812 chemicals include 387 pesticide active ingredients, and 364 pesticide inerts; most of the remaining chemicals are relevant to the EDSP, contingent on potential for exposure of substantial human populations through sources of drinking water.

Results

Concentration-response curves of the 18 high-throughput *in vitro* ER assays were integrated into a model of ER pathway bioactivity[14]. Model outputs include an integrated measure of agonist bioactivity, antagonist bioactivity, as well as "false positive" signaling due to cytotoxicity or technology-specific interference[14]. For reasons described previously in this paper, only ER agonist bioactivity was considered in these analyses. Performance metrics (true positives, true negatives, false positives, false negatives, balanced accuracy, sensitivity, and specificity)[37] were calculated for ToxCast ER model score performance against reference

chemicals and guideline studies. Performance metrics were calculated for chemicals with any indication of ToxCast ER agonist bioactivity (AUC>0) or no activity (AUC=0) and again excluding inconclusive model scores (0<AUC<0.1) for which no call of bioactivity could be determined (Table 4).

The performance-based assessment of the ToxCast ER model for agonist bioactivity relied on *in vitro* reference chemicals, *in vivo* reference chemicals, guideline-like uterotrophic studies, and results of EDSP Tier 1 assays (Figure SI-1). For the 40 *in vitro* agonist reference chemicals, the ToxCast ER model performed very well, with an overall balanced accuracy of 93% (Table 4). Of the 28 active reference chemicals, 26 of 28 had ToxCast ER model bioactivity (Table 2). ToxCast ER agonist scores were positive ($\geq$ 0.1) for all strong, moderate and weak agonist reference chemicals (Table 2). ToxCast ER model bioactivity was inconclusive (0<AUC<0.1) for one very weak active chemical (di-n-butyl phthalate). Two very weak reference chemicals (diethylhexyl phthalate (DEHP) and dicofol) were false negatives (model score = 0). Of the 12 inactive reference chemicals, 11 chemicals had no ER model agonist bioactivity. One inactive chemical (haloperidol) had an inconclusive ToxCast model score (0<AUC<0.1). If the two chemicals with inconclusive ToxCast ER agonist model scores are excluded from performance metrics, the overall accuracy is 95% (Table 4).

The overall accuracy of the ToxCast ER model agonist bioactivity evaluated for 43 *in vivo* reference chemicals (Table 3) with independently verified results in two or more guideline-like uterotrophic studies was 86% (Table 4). Of 30 active reference chemicals, 29 had positive ToxCast ER agonist model scores (AUC $\geq$ 0.1; Table 3). The potential false negative chemical, octamethylcyclotetrasiloxane (D4), was positive in multiple uterotrophic studies run in independent labs but negative in the ToxCast ER bioactivity model (Table 3). Due to the volatility of the chemical (157 Pa /1.18 mmHg at 25°C), it is possible that the concentration of the compound actually tested in the high-throughput assays was lower than the calculated nominal concentration with these considerations. Of 13 inactive *in vivo* reference chemicals, eight had no ToxCast ER agonist bioactivity (AUC=0; Table 3). Kaempferol was negative in uterotrophic studies but had modest ER agonist model bioactivity (AUC=0.25, Table 3) and scored as a false positive, though the positive result was consistent with other lower throughput *in vitro* ER assays[29, 38]. Four inactive *in vivo* reference chemicals (dibutyl phthalate, dicyclohexyl phthalate, dihexyl phthalate, and fenvalerate) had very low ToxCast ER model scores associated with inconclusive calls (0<AUC<0.1; Table 3), This result supports the hypothesis that ToxCast ER model bioactivity in this range has limited *in vivo* relevance. If the four inconclusive chemicals were excluded from calculations of performance metrics, the resulting overall accuracy for model performance for *in vivo* reference chemicals was 95% (37/39; Table 4).

To expand the evaluation of the ToxCast ER model, we compared model agonist bioactivity with 103 chemicals run in at least one guideline-like uterotrophic study. This larger set of chemicals included the 43 *in vivo* reference chemicals as defined above, plus 60 additional chemicals that did not meet the stringent criteria for an *in vivo* reference chemical because they were only run in one study or had discordant results that could not be resolved (*e.g.*, only one guideline-like positive and at least one guideline-like negative study for the same chemical; Figure SI-1). Of the 55 chemicals with at least one positive uterotrophic response, 49 had ToxCast model bioactivity; nine of the 49 had inconclusive scores (Table SI-1). Six chemicals with positive uterotrophic studies (methylparaben, triclosan, reserpine, permethrin, octamethylcyclotetrasiloxane, and gibberellic acid) had no reported bioactivity in the ToxCast model (*i.e.,* AUC = 0). Forty-eight chemicals had no significant effect on uterine weight in any study examined. Of these, 37 had no ToxCast model bioactivity, seven were inconclusive, and four chemicals with ToxCast model scores $\geq$0.1 (phenolphthalein, benzoic acid, kaempferol, and benzylbutylphthalate) were potentially false positives. It is worth noting that three of the four potential

false positives (phenolphthalein, kaempferol, and benzylbutylphthalate) were identified as *in vitro* positive reference chemicals[29]. The overall accuracy of the ToxCast ER model when compared to uterotrophic guideline-like studies was 84% (86/103; Table 4). If inconclusive calls are excluded from analyses, concordance between the ToxCast ER model bioactivity and the *in vivo* guideline-like uterotrophic studies for 75 chemicals with ToxCast positive ($\geq$0.1) or negative (0) model scores was 88% (75/85; Table 4).

Comparing ToxCast model scores and EDSP Tier 1 results, three List 1 chemicals did not have ToxCast assay data and none of the remaining 49 chemicals had ToxCast ER model scores $\geq$ 0.1. Similarly, none of the chemicals had clear positive agonist activity in the Tier 1 ER *in vitro* assays (ER binding and ERTA) or *in vivo* (uterotrophic) assays. ToxCast ER model scores were inconclusive for eight List 1 Tier 1 chemicals, all of which had limited signal in the EDSP Tier 1 assays, but none of which would be considered positive based on the Tier 1 response (Table 4; Table SI-1). All ToxCast ER assay responses for these chemicals were detected at concentrations similar to those that resulted in cytotoxicity and may be explained by cell-stress or cytotoxicity-related false positive activity. Although there were both positive and negative Tier 1 ERTA assays reported for chemical codes 41 and 43, there were not clear indications of a positive Tier 1 ER binding, ERTA, or uterotrophic study (or any study submitted to EPA to satisfy a Tier 1 test order) for any chemical. Similarly ToxCast model scores were negative for the remaining 41 chemicals. Comparison between computational methods and Tier 1 assays is biased by the lack of positive results, but for this analysis the ToxCast model accuracy against List 1 chemicals with Tier 1 data is 84% (41/49) and 100% if inconclusive results are not included (Table 4).

ToxCast ER model scores were used to evaluate potential agonist activity in the 1812 chemicals with data for all 18 *in vitro* ToxCast ER assays, including 57 of 107 List 2 chemicals (Table SI-2). Concentration-response data, curve fits, and AC_{50} for all 18 ER assay as well as ToxCast ER model agonist scores are available for all 1812 chemicals (http://actor.epa.gov/edsp21/). All of the 57 List 2 chemicals lacked ToxCast ER agonist bioactivity (*i.e.*, model scores < 0.1; Figure SI-2)[39]. However among the remaining chemicals run in ToxCast high-throughput ER assays, about 7% (133) chemicals had ToxCast ER model scores indicating positive agonist bioactivity (*i.e.*, scores $\geq$ 0.1), 15% (276) were inconclusive, and 77% (1403) have no observed ER agonist bioactivity (Figure SI-2)[39].

Discussion

Before new computational toxicology tools can be used to screen for potential endocrine bioactivity, their utility should be adequately demonstrated for the proposed purpose. The key aspect of the analysis presented in this paper is the performance-based validation approach which uses multiple sets of well-studied reference chemicals to establish specificity and sensitivity of the ToxCast ER computational model, and comparisons of model scores with existing test methods used for regulatory decision making. Our analyses focused on ER agonism because the EDSP Tier 1 battery assays that measure ER interactions are only capable of detecting agonism, and most estrogen-active environmental chemicals act as agonists[9-12]. The ToxCast ER model accurately predicted the bioactivity of reference chemicals across a range of structures and potencies[14], and for a relatively large set of 193 chemicals with bioactivities independently confirmed by another test method (*i.e.*, *in vitro* and *in vivo* reference chemicals as well as results of guideline-like uterotrophic studies, results of List 1/Tier 1 *in vitro* and uterotrophic studies; Figure SI-1). Together, these analyses provide a high degree of confidence in ToxCast ER model predictions and demonstrate the utility of using these computational tools to meet the intended objectives which were to: 1) contribute to the weight of evidence evaluation of a chemical's ER agonist

activity; and 2) provide an alternative source of data for specific EDSP Tier 1 endpoints measuring *in vitro* and *in vivo* ER interaction.

The time and resource intensive multi-laboratory approaches traditionally used to validate toxicology assays require seven to 10 years even for simple assays and are not suited to the rapid inclusion of new high-throughput tools and emerging technologies. A number of groups have proposed more rapid performance-based validation approach for new assays[40-45], including a large number of reference chemicals that spans a range of structures and potencies. The inclusion of relatively large sets of reference chemicals provides a high degree of confidence in the validation, greatly increases knowledge of which chemicals are active in a given test, helps to define the chemical space for which the assays can accurately predict an outcome (*i.e.,* the domain of applicability)[46, 47], and illustrates a major strength of high-throughput assays capable of screening a variety of chemical classes. Furthermore, the performance-based validation identifies a set of reference chemicals which can be used to validate any assay or group of assays that accurately detect the reference chemicals, and such an approach can and will be adopted for other endocrine pathways. Given the thousands of chemicals to be screened for potential endocrine bioactivity by the EDSP, validating high throughput approaches to be used in a tiered testing strategy is the only practical path forward.

The performance-based validation approach described in this study used a larger set of reference chemicals than were used to validate the EDSP Tier 1 ER assays and the bioactivities were independently confirmed. The *in vitro* reference chemicals were active (or inactive) in several different types of ER assays, with reported potencies ranging five orders of magnitude, the lower end of which can be used to establish a highly sensitive *in vitro* assay but may have limited *in vivo* biological relevance. The ToxCast ER model identified all strong, moderate, and weak reference chemicals (*i.e.*, AC50<1 μM); the only ambiguous results were inconclusive bioactivity for one negative chemical and inconclusive (1 chemical) or no detected bioactivity (2 chemicals) for three very weak reference chemicals. In contrast to the 40 *in vitro* reference chemicals used in this analysis, the Tier 1 ER *in vitro* transactivation and binding assays were initially validated with 12 and 23 chemicals, respectively[28, 48], and results for 35% of the reference chemicals tested in the Tier 1 *in vitro* binding assay were not consistent with the expected outcomes, either because of lack of agreement among assay results from different labs or disagreement with observed results and anticipated activity of the selected chemicals[48]. The 43 *in vivo* reference chemicals included a similarly diverse range of structures and potencies, and greatly exceeded the seven chemicals examined in the OECD validation of the *in vivo* uterotrophic assay[49]. The relatively short duration and limited number of animals employed in the standard uterotrophic study design has been reported to potentially elicit false negative results due to variability in uterus weights of control animals[50, 51]. The ToxCast ER model also performed better against uterotrophic results than lower throughput *in vitro* assays[52] which showed 66% agreement between results of competitive ER binding and uterotrophic assays results for 65 chemicals. Though the ToxCast ER model results also agreed with the List 1 Tier 1 uterotrophic assay results, the guideline uterotrophic study includes a single endpoint (uterine weight) and may actually be less sensitive and reproducible than *in vitro* assays[52], particularly given that 18 ToxCast ER assays targeting multiple key events along the pathway are likely to reduce false negative responses.

Although comparing the ToxCast ER model scores with *in vivo* uterotrophic results was a critical part of the validation approach, further analysis of the full set of 103 guideline-like uterotrophic studies indicated a moderate degree of *in vivo* inter-study variability[13]. Evaluation of uterotrophic study results for any single chemical often differed with animal model, strain, dose of test chemical, and delivery route used in the study, and highlighted the inherent variability in uterotrophic "guideline" method[14, 53]. Of chemicals with >1 guideline-

like study, 26% had contradictory results with at least one positive and one negative study[14], which puts into perspective the 84% accuracy of the ToxCast model when compared with all guideline-like uterotrophic studies in this analysis. While much of the variability in the uterotrophic assay can be explained by differences in the experimental design, among the 24 guideline-like uterotrophic assays conducted for bisphenol A (BPA) delivered by subcutaneous injection to the immature rat, discordant results ranged over three orders of magnitude (*e.g.,* 4 mg/kg/d produced a positive response in one study and 1000 mg/kg/d failed to do so in another; Figure 1). Recognizing this variability is important because it sets realistic expectations as to the performance of any alternative method. For the uterotrophic assay, it is unrealistic that an alternative method should predict both the true positives as well as account for the associated *in vivo* experimental variability.

The ToxCast ER model incorporates the 18 high throughput *in vitro* assays in an unweighted manner. During the development of the model, multiple approaches for weighting the assays in a variety were investigated, including performance against reference chemicals and levels of baseline or background noise, both of which are indicators of overall reliability. Analyses indicated there was little change in the model output (AUC score), and the primary consequence was minor shifts in the ranking of chemicals at the expense of adding 18 free parameters to the model whose exact structures and values that could not be well justified. As a result, the mathematically simpler, unweighted model was used in this analysis.

For analyses presented in this manuscript, AUC$\geq$0.1 (equivalent to an AC_{50} of about 100 µM) was considered bioactive, AUC truncated at <0.001 was considered inactive, and chemicals with model scores 0<AUC<0.1 were considered inconclusive. When all bioactivities (*e.g.*, model scores <0.001) of reference chemical and List 1 Tier 1 chemicals (which all have independent corroboration of the estrogen agonist activity) are included in the evaluation of model performance, the number of false positive calls increases (Table 2), supporting the hypothesis that very low bioactivity scores in the inconclusive range are not biologically relevant. Similarly, if the threshold for positive calls was changed from 0.1 to 0.01, two more *in vitro* reference chemicals (di-n-butyl phthalate and haloperidol; Table 3) and four more *in vivo* reference chemicals (dibutyl phthalate, dicyclohexyl phthatlate, dihexyl phthatlate, and fenvalerate; Table 4) were predicted to be bioactive. Though this adjustment would detect one very weak active *in vitro* reference chemical, it would also result in one additional false positive *in vitro* reference chemical and four additional false negative *in vivo* reference chemicals. Further, an additional 139 chemicals of the 1812 run in ToxCast would be considered "active" (Table SI-2). None of the eight List 1 Tier 1 chemicals with inconclusive ToxCast model scores had consistent positive outcomes in the EDSP Tier 1 assays, further supporting the absence of biologically relevant bioactivity associated with ToxCast ER model scores in this range. At this point, quantitative uncertainties in the *in vitro* data and qualitative uncertainties in the uterotrophic assay data reduce the value of over interpreting inconclusive ToxCast ER agonist bioactivities. It should be noted that inconclusive chemicals would not escape additional testing, but additional data, including results of other EDSP Tier 1 battery assays, would be used to help interpret inconclusive ToxCast model scores (0.1>AUC>0). Eventually, the goal is to resolve inconclusive calls and distinguish “true” negative from very weak bioactivity, though this resolution may require additional data (e.g., assays, QSARs, or *in vivo* testing).

Performance-based validation of computational data demonstrates that the ToxCast ER model performs as well or better than the EDSP Tier 1 ER binding, ERTA, and uterotrophic assays. Accuracy of the ToxCast ER model is 84 to 93% when compared with all data sets described in this analysis (*i.e.*, 40 *in vitro* reference chemicals, 43 *in vivo* reference chemicals, the 63 guideline-like uterotrophic studies, and results of Tier 1 battery assays for 49 chemicals), and when inconclusive model scores are removed from the analyses, the accuracy ranges from 88 to 100% (Table 4). The high sensitivity of the model is critical for screening environmental chemicals, since it

means that few false negatives were observed (Table 4). Given the redundancy of coverage among the 18 ToxCast ER assays (*e.g.*, multiple receptor binding and transactivation assays), it is unlikely that running a single EDSP Tier 1 guideline ER binding or ERTA assay would provide additional insight into a chemical's potential ER bioactivity. In addition, the 18 assays included in the network model provide a more comprehensive pathway coverage for the biology of the ER signaling pathway (*e.g.*, receptor binding, dimerization, mRNA production, protein production, cell proliferation) and a more robust estimate of a chemical's potential ER bioactivity than existing Tier 1 ER binding and ERTA assays because the ToxCast model is capable of detecting false positives due to assay-specific interference or cytotoxicity that can be discriminated from "true" bioactivity[14]. When the performance of the ToxCast ER model is considered in the context of our stated objectives, it is clear that the model has demonstrated utility for contributing to the weight of evidence of a chemical's potential interaction with the ER pathway, and if ToxCast *in vitro* assay data were available for a given chemical, no additional information on potential ER bioactivity would be gained by requiring a EDSP Tier 1 ER binding, ERTA or uterotrophic assay.

The ToxCast high throughput screening assays and models built on the output of the assays do have some limitations, though many of these are also limitations of the EDSP Tier 1 counterparts. As with any *in vitro* assay, highly volatile chemicals or chemicals with low solubility (in DMSO) are difficult to assess. The current high-throughput ER assays have limited capacity to address chemicals that may be biotransformed to active or inactive metabolites, though the ToxCast ER bioactivity model does detect both methoxychlor (AUC=0.254) and its more potent metabolite 2,2-bis(4-hydroxyphenyl)-1,1,1-trichloroethane (HPTE; AUC=0.568). Additionally, this is a similar limitation of the EDSP Tier 1 ER *in vitro* and uterotrophic assays, as the preferred method of administration for the guideline *in vivo* uterotrophic assay is subcutaneous injection which bypasses first pass hepatic metabolism. The EDSP Tier 1 ER *in vitro* assays are specific to ERα and only the ToxCast dimerization and protein production assays detect ERβ bioactivity, therefore both screens may not detect chemicals that are ERβ-selective or act through non-classical ER mechanisms.

Our intention was to demonstrate the validity and utility of the ToxCast ER model for screening thousands of environmental chemicals for potential ER agonist bioactivity, though these analyses do not support substitution of all EDSP Tier 1 assays relevant to the estrogen pathway. In the absence of an *in vivo* uterotrophic assay, the EDSP Tier 1 *in vivo* rat pubertal assay[54], which includes a variety of endpoints in addition to uterine weight, could detect chemicals that are metabolized to active metabolites through the oral administration of test chemical. The rat pubertal assay exposes a larger sample of animals (n=16 versus n=6) for a longer duration (20 days versus three days), which may increase the sensitivity of the assay compared with the uterotrophic test. The EDSP Tier 1 battery is intended to screen potential endocrine effects in the highly conserved ER signally pathway of humans *and* wildlife. ToxCast *in vitro* ER assays measure effects using cells derived from the kidney, cervix, liver, ovary, uterus, and breast; use rodent, bovine, and human receptor proteins; and detect interaction using a variety of technologies (Table 1). This diversity of the 18 high-throughput ER assays accounts for estrogenic effects more broadly across cell types, organs, and species than the single human ovarian and two rodent uterine ER assays in the existing Tier 1 battery and may have greater relevance to wildlife.

We demonstrated the ToxCast ER model ability to predict ER bioactivity of *in vitro* and *in vivo* reference chemicals, the utility of using the ToxCast ER model bioactivity as an alternative to the EDSP Tier 1 ER binding, ERTA, and uterotrophic endpoints, and the application of ToxCast ER model scores to prioritize chemicals in the EDSP universe for additional screening and testing. Results of the ER model indicate only about 7% of the 1812 chemicals run in ToxCast have potential significant ER agonist bioactivity and this subset does not include any

List 1 or List 2 chemicals, and the absence of predicted bioactivity among EDSP List 1 chemicals is consistent with the initial review of Tier 1 battery data for a subset of List 1 chemicals[55, 56]. The ToxCast ER model can be used to rapidly screen thousands of chemicals in the EDSP universe, allowing EPA to move away from screening lists of few chemicals with relatively low or no potential endocrine activity, reduce reliance on animal-based assays, and identify chemicals with the greatest potential endocrine bioactivity that may be high priority candidates for further screening and testing[55, 56]. This approach for using computational toxicology tools in the EDSP only evaluated ER-mediated bioactivity and it should be noted that while List 1 and List 2 chemicals appear to have limited ER bioactivity, these chemicals may be active in other endocrine pathways. In the future, we plan to use a performance-based validation approach of high-throughput ToxCast ER assays and other computational toxicology tools to compare with the existing "guideline-like" Tier 1 assays, including the fish and pubertal rat, to determine how well high-throughput models predict estrogen bioactivity in neuroendocrine-intact animals. In addition, EPA will use this performance-based approach for validating new computational tools to screen for androgen and thyroid effects, taking advantage of both existing and innovative, emerging technologies to implement a scientifically robust and comprehensive chemical prioritization process for EDSP. The application of these innovative tools to screening chemicals for endocrine bioactivity represents the first step in a paradigm shift for chemical safety testing, a practical approach to rapidly screen thousands of environmental chemicals for potential endocrine bioactivity in humans and wildlife, and the first systematic application of ToxCast data in an EPA regulatory program.

Acknowledgements

We thank D. Bergfelt, A. Dixon, T. Green, K. Hamernik, S. Lynn, K. Markey, L. Touart, and W. Wooge from OSCP; K. Houck, K. Crofton, T. Knudsen, M. Martin, A. Richard, W. Setzer, I. Shah, J. Wambaugh, and T. Bahadori from ORD for their contributions. We are grateful to the NICEATM/ILS team for curation of the uterotrophic database. We also thank R. Hines from ORD for his technical review of the manuscript.

Supporting Information Available

Table SI-1 summarizes List 1 chemical EDSP Tier 1 ER binding, ERTA, and uterotrophic results compared with ToxCast ER model scores. Table SI-2 includes ToxCast ER model scores for all 1812 chemicals, along with uterotrophic results. Figure SI-1 summarizes the various chemical populations used in the performance based validation approach. Figure SI-2 is a graph of ToxCast ER model scores for active and inactive *in vitro* reference chemicals, EDSP List 1 chemicals, EDSP List 2 chemicals, and the remaining positive ToxCast ER agonist scores among the 1812 chemicals for which all 18 ToxCast high throughput ER assay data are available.

This information is available free of charge via the Internet at http://pubs.acs.org .

446 **Table 1.** Summary of the 18 high-throughput *in vitro* estrogen receptor (ER) assays included in the ToxCast ER bioactivity model. For additional
447 information on assays, please reference the EDSP21 dashboard[39]. NA = Not Applicable (cell-free binding assay).

Assay ID	Assay Name	Biological Process Target	Detection Technology	Organism	Tissue	Cell Line
A1	NVS_NR_bER	receptor binding	radioligand	bovine	uterus	NA
A2	NVS_NR_hER	receptor binding	radioligand	human	NA	NA
A3	NVS_NR_mERa	receptor binding	radioligand	mouse	NA	NA
A4	OT_ER_ERaERa_0480	protein complementation	fluorescence	human	kidney	HEK293T
A5	OT_ER_ERaERa_1440	protein complementation	fluorescence	human	kidney	HEK293T
A6	OT_ER_ERaERb_0480	protein complementation	fluorescence	human	kidney	HEK293T
A7	OT_ER_ERaERb_1440	protein complementation	fluorescence	human	kidney	HEK293T
A8	OT_ER_ERbERb_0480	protein complementation	fluorescence	human	kidney	HEK293T
A9	OT_ER_ERbERb_1440	protein complementation	fluorescence	human	kidney	HEK293T
A10	OT_ERa_EREGFP_0120	protein production	fluorescence	human	cervix	HeLa
A11	OT_ERa_EREGFP_0480	protein production	fluorescence	human	cervix	HeLa
A12	ATG_ERa_TRANS_up	mRNA induction	fluorescence	human	liver	HepG2
A13	ATG_ERE_CIS_up	mRNA induction	fluorescence	human	liver	HepG2
A14	Tox21_ERa_BLA_Agonist_ratio	protein production	fluorescence	human	kidney	HEK293T
A15	Tox21_ERa_LUC_BG1_Agonist	protein production	bioluminescence	human	ovary	BG1
A16	ACEA_T47D_80hr_Positive	cell proliferation	electrical impedance	human	breast	T47D
A17	Tox21_ERa_BLA_Antagonist_ratio	protein production	fluorescence	human	kidney	HEK293T
A18	Tox21_ERa_LUC_BG1_Antagonist	protein production	bioluminescence	human	ovary	BG1

448

449 **Table 2:** *In vitro* estrogen receptor (ER) agonist reference chemicals.

CASRN	Chemical Name	Agonist Potency[1]	ToxCast ER Model Score
57-63-6	17alpha-Ethinyl estradiol	Strong	1
84-16-2	meso-Hexestrol	Strong	0.99
56-53-1	Diethylstilbestrol (DES)	Strong	0.94
50-28-2	17beta-Estradiol	Strong	0.94
57-91-0	17alpha-Estradiol	Moderate	1.06
53-16-7	Estrone	Moderate	0.81
140-66-9	4-tert-Octylphenol	Moderate	0.39
446-72-0	Genistein	Weak	0.54
77-40-7	Bisphenol B	Weak	0.49
80-05-7	Bisphenol A	Weak	0.45
486-66-8	Daidzein	Weak	0.44
521-18-6	5alpha-Dihydrotestosterone	Weak	0.40
789-02-6	o,p'-DDT	Weak	0.39
599-64-4	4-Cumylphenol	Weak	0.38
143-50-0	Kepone	Weak	0.17
58-18-4	17alpha-Methyltestosterone	Very Weak	0.50
520-36-5	Apigenin	Very Weak	0.31
72-43-5	Methoxychlor	Very Weak	0.25
520-18-3	Kaempferol	Very Weak	0.25
85-68-7	Butylbenzyl phthalate	Very Weak	0.18
480-40-0	Chrysin	Very Weak	0.13
60168-88-9	Fenarimol	Very Weak	0.11
104-40-5	p-n-Nonylphenol	Very Weak	0.1
120-47-8	Ethylparaben	Very Weak	0.1
72-55-9	p,p'-DDE	Very Weak	0.1
84-74-2	Di-n-butyl phthalate	Very Weak	0.03
115-32-2	Dicofol	Very Weak	0
117-81-7	Diethylhexyl phthalate	Very Weak	0
52-86-8	Haloperidol	Inactive	0.01
52-01-7	Spironolactone	Inactive	0
50-22-6	Corticosterone	Inactive	0
13311-84-7	Flutamide	Inactive	0
1912-24-9	Atrazine	Inactive	0
32809-16-8	Procymidone	Inactive	0
330-55-2	Linuron	Inactive	0
50-55-5	Reserpine	Inactive	0
52806-53-8	Hydroxyflutamide	Inactive	0
57-30-7	Phenobarbital Sodium	Inactive	0
65277-42-1	Ketoconazole	Inactive	0
66-81-9	Cycloheximide	Inactive	0

450 [1]Reference chemical potency, determined by concentration required to elicit 50% of the maximal response
451 (AC_{50}), in low throughput *in vitro* ER assays[28, 29]. Strong = AC_{50} <0.0001 µM, moderate = AC_{50} < 0.1 µM, weak =
452 AC_{50} < 1 µM, very weak = all other activities, inactive = no detected activity[29].

453 **Table 3:** *In vivo* estrogen receptor (ER) agonist reference chemicals with at least two independent active or
454 inactive guideline-like uterotrophic studies[14]. The numbers of guideline-like active and inactive study results are
455 reported for each chemical.

CASRN	Name	Active	Inactive	Bioactivity	ToxCast ER Model Score
57-91-0	17alpha-Estradiol	2	0	Active	1.06
57-63-6	Ethinyl Estradiol	59	0	Active	1
56-53-1	Diethylstilbestrol (DES)	8	1	Active	0.94
50-28-2	Estradiol	25	0	Active	0.94
474-86-2	Equilin	2	0	Active	0.82
53-16-7	Estrone	9	0	Active	0.81
50-27-1	Estriol	4	0	Active	0.79
72-33-3	Mestranol	3	0	Active	0.74
17924-92-4	Zearalenone	4	0	Active	0.71
1478-61-1	Bisphenol AF	4	0	Active	0.55
446-72-0	Genistein	27	1	Active	0.54
68-22-4	Norethindrone	2	0	Active	0.52
58-18-4	Methyltestosterone	3	0	Active	0.50
77-40-7	Bisphenol B	2	0	Active	0.49
80-05-7	Bisphenol A	37	6	Active	0.45
104-43-8	4-Dodecylphenol	3	0	Active	0.41
521-18-6	Dihydrotestosterone	3	0	Active	0.4
131-55-5	Benzophenone-2	6	0	Active	0.40
140-66-9	4-(1,1,3,3-Tetramethylbutyl)phenol	3	1	Active	0.39
789-02-6	o,p'-DDT	15	1	Active	0.39
599-64-4	p-Cumylphenol	2	0	Active	0.38
5153-25-3	Benzoic acid, 4-hydroxy-, 2-ethylhexyl ester	2	0	Active	0.37
80-46-6	4-(1,1-Dimethylpropyl)phenol	4	0	Active	0.28
131-56-6	2,4-Dihydroxybenzophenone	3	0	Active	0.27
80-09-1	Bisphenol S	2	0	Active	0.26
72-43-5	Methoxychlor	18	1	Active	0.25
94-26-8	Butylparaben	8	2	Active	0.25
98-54-4	p-tert-Butylphenol	2	0	Active	0.16
104-40-5	Nonylphenol	5	4	Active	0.10
556-67-2	Octamethylcyclotetrasiloxane	3	0	Active	0
520-18-3	Kaempferol	0	3	Inactive	0.25
84-74-2	Dibutyl phthalate	0	2	Inactive	0.03
84-61-7	Dicyclohexyl phthalate	0	2	Inactive	0.02
84-75-3	Dihexyl phthalate	0	2	Inactive	0.01
51630-58-1	Fenvalerate	0	2	Inactive	0.01
103-23-1	Bis(2-ethylhexyl)hexanedioate	0	2	Inactive	0
117-81-7	Bis(2-ethylhexyl)phthalate	0	2	Inactive	0

CASRN	Name	Active	Inactive	Bioactivity	ToxCast ER Model Score
1461-22-9	Tributylchlorostannane	0	2	Inactive	0
1912-24-9	Atrazine	0	2	Inactive	0
61-82-5	Amitrole	0	2	Inactive	0
84-66-2	Diethyl phthalate	0	2	Inactive	0
87-86-5	Pentachlorophenol	0	2	Inactive	0
99-96-7	4-Hydroxybenzoic acid	0	2	Inactive	0

456

Table 4. Performance based validation of the ToxCast ER model based on 18 high-throughput *in vitro* assays measuring potential estrogen receptor (ER) agonist activities and *in vitro* reference chemicals (see text for detailed explanation). ToxCast ER model scores ≥0.1 were considered positive, negative scores = 0 (and values <0.001 were truncated as 0), and model scores (0>AUC<0.1) were inconclusive. Performance metrics were calculated with all chemicals and excluding chemicals with inconclusive model scores (values shown in parentheses).

Performance	*In vitro* reference chemicals	*In vivo* reference chemicals	GL uterotrophic studies	Tier 1 studies
# True Pos	26 (25)	29 (29)	49 (38)	0 (0)
# True Neg	11 (11)	8 (8)	37 (37)	41 (41)
# False Pos	1 (0)	5 (1)	11 (4)	8 (0)
# False Neg	2 (2)	1 (1)	6 (6)	0 (0)
Accuracy	0.93 (0.95)	0.86 (0.95)	0.84 (0.88)	0.84 (1.0)
Sensitivity	0.93 (0.93)	0.97 (0.97)	0.89 (0.86)	0 (0)
Specificity	0.92 (1.0)	0.67 (0.89)	0.77 (0.90)	0.84 (1.0)

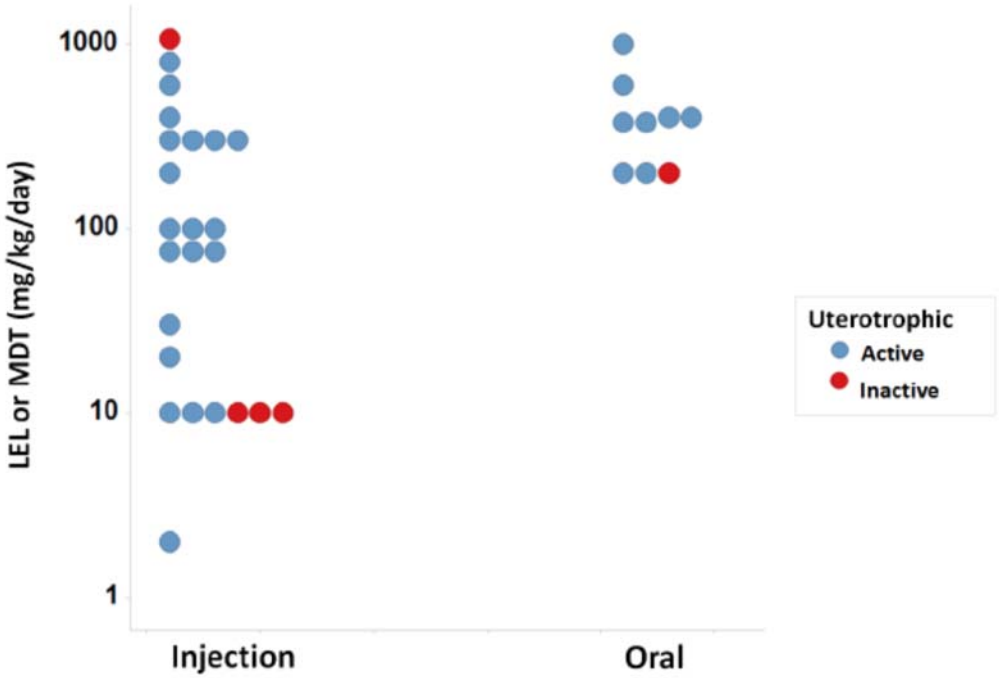


Figure 1. Variability of results for bisphenol A (BPA) in uterotrophic studies conducted in the immature rat model. All studies are methodologically similar to the EDSP Tier 1 guideline and considered "guideline-like", yet have discordant results even with the same route of administration. LEL = lowest effect level; MDT = maximum dose tested.

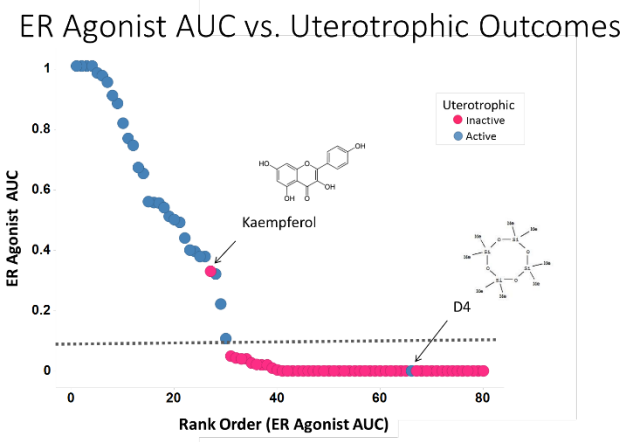
Literature Cited

1. U.S. EPA EDSP Tier 1 Battery of Assays http://www.epa.gov/endo/pubs/assayvalidation/tier1battery.htm (2009),
2. U.S. EPA EDSP Overview. http://www.epa.gov/endo/pubs/edspoverview/components.htm
3. U.S. EPA Final List of Chemicals for Initial Tier 1 Screening. http://www.epa.gov/endo/pubs/prioritysetting/finallist.html
4. U.S. EPA Revised Second List of Chemicals for Tier 1 Screening. http://www.epa.gov/endo/pubs/prioritysetting/revlist2.htm
5. U.S. EPA, Universe of Chemicals and General Validation Principles. In U.S. Environmental Protection Agency: http://www.epa.gov/endo/pubs/edsp_chemical_universe_and_general_validations_white_paper_11_12.pdf, 2012.
6. NRC, *Toxicity testing in the 21st century: A vision and a strategy*. National Academies Press: Washington, DC, 2007; Vol. 25, p 136-138.
7. U.S. EPA, FY 2012 EPA Budget In Brief. In Officer, O. o. t. C. F., Ed. U.S. Environmental Protection Agency: http://nepis.epa.gov/Exe/ZyPURL.cgi?Dockey=P100A5RE.txt, 2011; p 128.
8. U.S. EPA, The U.S. Environmental Protection Agency's Strategic Plan for Evaluating the Toxicity of Chemicals In U.S. Environmental Protection Agency: http://www.epa.gov/spc/toxicitytesting/docs/toxtest_strategy_032309.pdf, 2009.
9. Kojima, H.; Katsura, E.; Takeuchi, S.; Niiyama, K.; Kobayashi, K., Screening for estrogen and androgen receptor activities in 200 pesticides by in vitro reporter gene assays using Chinese hamster ovary cells. *Environmental health perspectives* **2004,** *112*, (5), 524-31.
10. Safe, S. H.; Pallaroni, L.; Yoon, K.; Gaido, K.; Ross, S.; Saville, B.; McDonnellc, D., Toxicology of environmental estrogens. *Reproduction, fertility, and development* **2001,** *13*, (4), 307-15.
11. Sonnenschein, C.; Soto, A. M., An updated review of environmental estrogen and androgen mimics and antagonists. *The Journal of steroid biochemistry and molecular biology* **1998,** *65*, (1-6), 143-50.
12. Diamanti-Kandarakis, E.; Bourguignon, J. P.; Giudice, L. C.; Hauser, R.; Prins, G. S.; Soto, A. M.; Zoeller, R. T.; Gore, A. C., Endocrine-disrupting chemicals: an Endocrine Society scientific statement. *Endocrine reviews* **2009,** *30*, (4), 293-342.
13. U.S. EPA EPA's ToxCast Program. http://www.epa.gov/ncct/toxcast/
14. U.S. EPA, Integrated Bioactivity and Exposure Ranking: A computational Approach for the Prioritization and Screening of Chemicals in the Endocrine Disruptor Screening Program. In U.S. Environmental Protection Agency: http://www.regulations.gov/#!documentDetail;D=EPA-HQ-OPP-2014-0614-0003, 2014.
15. Dix, D. J.; Houck, K. A.; Martin, M. T.; Richard, A. M.; Setzer, R. W.; Kavlock, R. J., The ToxCast program for prioritizing toxicity testing of environmental chemicals. *Toxicol. Sci.* **2007,** *95*, (1), 5-12.
16. Houck, K. A.; Kavlock, R. J., Understanding mechanisms of toxicity: Insights from drug discovery research. *Toxicology and Applied Pharmacology* **2008,** *227*, 163-178.
17. Martin, M. T.; Knudsen, T. B.; Reif, D. M.; Houck, K. A.; Judson, R. S.; Kavlock, R. J.; Dix, D. J., Predictive model of rat reprodcutive toxicity from ToxCast high throughput screening *Biology of Reproduction* **2011,** *85*, (2), 327-339.
18. Rotroff, D. M.; Dix, D. J.; Houck, K. A.; Knudsen, T. B.; Martin, M. T.; McLaurin, K. W.; Reif, D. M.; Crofton, K. M.; Singh, A. V.; Xia, M. H.; Huang, R. L.; Judson, R. S., Using *in vitro* high throughput screening assays to identifiy potential endocrine disrupting chemicals. *Environmental health perspectives* **2013,** *121*, 7-14.
19. Reif, D. M.; Martin, M. T.; Tan, S. W.; Houck, K. A.; Judson, R. S.; Richard, A. M.; Knudsen, T. B.; Dix, D. J.; Kavlock, R. J., Endocrine profile and prioritization of environmental chemicals using ToxCast data. *Environmental health perspectives* **2010,** *118*, (12), 1714-1720.

20. Judson, R. S.; Houck, K. A.; Kavlock, R. J.; Knudsen, T. B.; Martin, M. T.; Mortensen, H. M.; Reif, D. M.; Rotroff, D. M.; Shah, I.; Richard, A. M.; Dix, D. J., In vitro screening of environmental chemicals for targeted testing prioritization: the ToxCast project. *Environmental health perspectives* **2010,** *118*, (4), 485-92.
21. Knudsen, T.; Houck, K.; Judson, R. S.; Singh, A.; Mortensen, H.; Reif, D.; Dix, D. J.; Kavlock, R. J., Biochemical activities of 320 ToxCast chemicals evaluated across 239 functional targets. *Toxicology* **2011,** *282*, 1-15.
22. Sipes, N. S.; Martin, M. T.; Kothiya, P.; Reif, D. M.; Judson, R. S.; Richard, A. M.; Houck, K. A.; Dix, D. J.; Kavlock, R. J.; Knudsen, T. B., Profiling 976 ToxCast chemicals across 331 enzymatic and receptor signaling assays. *Chemical research in toxicology* **2013,** *26*, (6), 878-95.
23. Martin, M. T.; Dix, D. J.; Judson, R. S.; Kavlock, R. J.; Reif, D. M.; Richard, A. M.; Rotroff, D. M.; Romanov, S.; Medvedev, A.; Poltoratskaya, N.; Gambarian, M.; Moeser, M.; Makarov, S. S.; Houck, K. A., Impact of environmental chemicals on key transcription regulators and correlation to toxicity end points within EPA's ToxCast program. *Chemical research in toxicology* **2010,** *23*, (3), 578-90.
24. Huang, R.; Sakamuru, S.; Martin, M. T.; Reif, D. M.; Judson, R. S.; Houck, K. A.; Casey, W.; Hsieh, J.-H.; Shockley, K. R.; Ceger, P.; Fostel, J.; Witt, K. L.; Tong, W.; Rotroff, D. M.; Zhao, T.; Shinn, P.; Simeonov, A.; Dix, D. J.; Austin, C. P.; Kavlock, R. J.; Tice, R. R.; Xia, M., Profiling of the Tox21 10K compound library for agonists and antagonists of the estrogen receptor alpha signaling pathway. *Scientific Reports* **2014,** *4*, (5664).
25. Rotroff, D. M.; Dix, D. J.; Houck, K. A.; Kavlock, R. J.; Knudsen, T. B.; Martin, M. T.; Reif, D. M.; Richard, A. M.; Sipes, N. S.; Abassi, Y. A.; Jin, C.; Stampfl, M.; Judson, R. S., Real-time growth kinetics measuring hormone mimicry for ToxCast chemicals in T-47D human ductal carcinoma cells. *Chemical research in toxicology* **2013,** *26*, (7), 1097-107.
26. U.S. EPA U.S. EPA ToxCast Program Data. http://epa.gov/ncct/toxcast/data.html
27. Filer, D., The ToxCast Analysis Pipeline: An R Package for Processing And Modeling Chemical Screening Data. In U.S. Environmental Protection Agency: http://www.epa.gov/ncct/toxcast/files/MySQL%20Database/Pipeline_Overview.pdf, 2014.
28. OECD, Test No. 455: Performance-Based Test Guideline for Stably Transfected Transactivation In Vitro Assays to Detect Estrogen Receptor Agonists. In Organisation for Economic Co-operation and Development: http://www.oecd-ilibrary.org/environment/test-no-455-performance-based-test-guideline-for-stably-transfected-transactivation-in-vitro-assays-to-detect-estrogen-receptor-agonists_9789264185388-en, 2012.
29. ICCVAM *The LUMI-CELL® ER (BG1Luc ER TA) Test Method: An In Vitro Assay for Identifying Human Estrogen Receptor Agonist and Antagonist Activity of Chemicals*; NEIHS: http://ntp.niehs.nih.gov/pubhealth/evalatm/iccvam/test-method-evaluations/endocrine-disruptors/bg1luc/tmer/index.html, 2011.
30. OECD, OECD Test No. 457: BG1Luc Estrogen Receptor Transactivation Test Method for Identifying Estrogen Receptor Agonists and Antagonists. In http://www.oecd-ilibrary.org/environment/test-no-457-bg1luc-estrogen-receptor-transactivation-test-method-for-identifying-estrogen-receptor-agonists-and-antagonists_9789264185395-en, 2012; Vol. 2012.
31. OECD, Test No. 440: Uterotrophic Bioassay in Rodents A short-term screening test for oestrogenic properties In Organisation for Economic Co-operation and Development: http://www.oecd-ilibrary.org/environment/test-no-440-uterotrophic-bioassay-in-rodents_9789264067417-en, 2007.
32. U.S. EPA, Endocrine Disruptor Screening Program Test Guidelines OPPTS 890. 1600: Uterotrophic Assay. In U.S. Environmental Protection Agency: http://www.regulations.gov/#!documentDetail;D=EPA-HQ-OPPT-2009-0576-0012, 2009.
33. U.S. EPA EPA ACToR. http://actor.epa.gov/actor
34. FDA, U. S. US Food and Drug Administration's Endocrine Disruptor Knowledge Base. http://www.fda.gov/ScienceResearch/BioinformaticsTools/EndocrineDisruptorKnowledgebase/default.htm

35. U.S. EPA, Endocrine Disruptor Screening Program Test Guidelines OPPTS 890. 1250: Rat Uterine Cytosol Estrogen Receptor Binding Assay. In U.S. Environmental Protection Agency: http://www.regulations.gov/#!documentDetail;D=EPA-HQ-OPPT-2009-0576-0005, 2009.
36. U.S. EPA, Endocrine Disruptor Screening Program Test Guidelines OPPTS 890. 1300: Estrogen Receptor Transactivation Assay. In U.S. Environmental Protection Agency: http://www.regulations.gov/#!documentDetail;D=EPA-HQ-OPPT-2009-0576-0006, 2009.
37. Stehman, S. V., Selecting and interpreting measures of thematic classification accuracy. *Remote Sensing of Environment* **1997,** *62*, (1), 13.
38. Zava, D. T., Estrogenic and antiproliferative properties of genestine and other flavonoids in human breast cancer cells *in vitro*. *Nutrition and Cancer* **1997,** *27*, (1), 31-40.
39. U.S. EPA EDSP21 Dashboard. http://actor.epa.gov/edsp21/
40. Becker, R. A.; Friedman, K. P.; Simon, T. W.; Marty, M. S.; Patlewicz, G.; Rowlands, J. C., An exposure:activity profiling method for interpreting high-throughput screening data for estrogenic activity--proof of concept. *Regulatory toxicology and pharmacology : RTP* **2015,** *71*, (3), 398-408.
41. Patlewicz, G.; Ball, N.; Boogaard, P. J.; Becker, R. A.; Hubesch, B., Building scientific confidence in the development and evaluation of read-across. *Regulatory toxicology and pharmacology : RTP* **2015,** *72*, (1), 117-33.
42. Patlewicz, G.; Simon, T. W.; Rowlands, J. C.; Budinsky, R. A.; Becker, R. A., Proposing a scientific confidence framework to help support the application of adverse outcome pathways for regulatory purposes. *Regulatory toxicology and pharmacology : RTP* **2015,** *71*, (3), 463-77.
43. Judson, R.; Kavlock, R.; Martin, M.; Reif, D.; Houck, K.; Knudsen, T.; Richard, A.; Tice, R. R.; Whelan, M.; Xia, M.; Huang, R.; Austin, C.; Daston, G.; Hartung, T.; Fowle, J. R., 3rd; Wooge, W.; Tong, W.; Dix, D., Perspectives on validation of high-throughput assays supporting 21st century toxicity testing. *Altex* **2013,** *30*, (1), 51-6.
44. Sauer, J. M.; Hartung, T.; Leist, M.; Knudsen, T. B.; Hoeng, J.; Hayes, A. W., Systems Toxicology: The Future of Risk Assessment. *International journal of toxicology* **2015**, [Epub ahead of print].
45. Tralau, T.; Oelgeschlager, M.; Gurtler, R.; Heinemeyer, G.; Herzler, M.; Hofer, T.; Itter, H.; Kuhl, T.; Lange, N.; Lorenz, N.; Muller-Graf, C.; Pabel, U.; Pirow, R.; Ritz, V.; Schafft, H.; Schneider, H.; Schulz, T.; Schumacher, D.; Zellmer, S.; Fleur-Bol, G.; Greiner, M.; Lahrssen-Wiederholt, M.; Lampen, A.; Luch, A.; Schonfelder, G.; Solecki, R.; Wittkowski, R.; Hensel, A., Regulatory toxicology in the twenty-first century: challenges, perspectives and possible solutions. *Arch Toxicol* **2015,** *89*, (6), 823-50.
46. OECD, Guidance Document on the Validation of (Quantitative) Structure-Activity Relationships [(Q)SAR] Models. In Organisation for Economic Co-operation and Development: http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?doclanguage=en&cote=env/jm/mono(2007)2, 2007; p 154.
47. OECD, Guidance Document on Developing and Assessing Adverse Outcome Pathways: Series on Testing and Assessment No. 184. In Organisation for Economic Co-operation and Development: http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2013)6&doclanguage=en, 2013.
48. U.S. EPA, Integrated Summary Report for Validation of an Estrogen Receptor Binding Assay Using Rat Uterine Cytosol as Source of Receptor as a Potential Screen in the Endocrine Disruptor Screening Program Tier 1 Battery. In U.S. Environmental Protection Agency: http://www.epa.gov/endo/pubs/assayvalidation/status.htm, 2009.
49. OECD, Series on Testing and Assessment. Number 67. Report of the Validation of the Uterotrophic Bioassay: Additional Data Supporting the TEst Guideline on the Uterotrophic Bioassay in Rodents. . In http://epa.gov/endo/pubs/uterotrophic_OECD_validation_report.pdf, 2007.

50. Ashby, J.; Odum, J., Gene expression changes in the immature rat uterus: effects of uterotrophic and sub-uterotrophic doses of bisphenol A. *Toxicological Sciences* **2004,** *82*, (2), 458-467.
51. Christian, M. S.; Hoberman, A. M.; Bachmann, S.; Hellwig, J., Variability in the uterotrophic response assay (an in vivo estrogenic response assay) in untreated control and positive control (DES-DP, 2.5 microG/kg, bid) Wistar and Sprague-Dawley rats. . *Drug and Chemical Toxicology* **1998,** *21 Supple 1*, 50-101.
52. Akahori, Y.; Nakai, M.; Yamasaki, K.; Takatsuki, M.; Shimohigashi, Y.; Ohtaki, M., Relationship between the results of *in vitro* receptor binding assay to human estrogen receptor α and *in vivo* uterotrophic assay: comparative study with 65 selected chemicals. *Toxicology in Vitro* **2008,** *22*, 225-332.
53. Laws, S. C.; Carey, S. A.; Ferrell, J. M.; Bodman, G. J.; Cooper, R. L., Estrogenic activity of octylphenol, nonylphenol, bisphenol A and methoxychlor in rats. *Toxicological sciences : an official journal of the Society of Toxicology* **2000,** *54*, (1), 154-67.
54. U.S. EPA, Endocrine Disruptor Screening Program Test Guidelines: OPPTS 890.1450. Pubertal Development and Thyroid Function in Intact Juvenile/Peripubertal Female Rats. In U.S. Environmental Protection Agency: http://www.regulations.gov/#!documentDetail;D=EPA-HQ-OPPT-2009-0576-0009, 2009.
55. U.S. EPA, SAP Minutes No: 2013-05 - Weight-of-Evidence: Evaluating Results of EDSP Tier 1 Screening. In U.S. Environmental Protection Agency: http://www.epa.gov/scipoly/sap/meetings/2013/july/073013minutes.pdf, 2013; p 65.
56. U.S. EPA, Final EDSP (Endocrine Disruptor Screening Program) White Paper on Tier I Assay and Battery Performance for May 21-23, 2013 FIFRA SAP Meeting. In U.S. Environmental Protection Agency: http://www.regulations.gov/contentStreamer?objectId=09000064812a3e5d&disposition=attachment&contentType=pdf, 2013; p 160.


ER Agonist AUC vs. Uterotrophic Outcomes
Uterotrophic
Inactive
Active
Kaempferol
D4
ER Agonist AUC
Rank Order (ER Agonist AUC)
0
0.2
0.4
0.6
0.8
1
0
20
40
60
80