Using In Vitro High-throughput Screening Assays to Identify Potential Endocrine Disrupting Chemicals

Daniel M. Rotroff [1,2], David J. Dix[2], Keith A. Houck[2], Thomas B. Knudsen[2], Matthew T. Martin[2], Keith W. McLaurin[2], David M. Reif[2], Kevin M. Crofton[3], Amar V. Singh[4], Menghang Xia[5], Ruili Huang[5], Richard S. Judson[2,¶]

[1]Department of Environmental Sciences and Engineering, University of North Carolina, Chapel Hill, NC 27514

[2]United States Environmental Protection Agency, Office of Research and Development, National Center for Computational Toxicology, Research Triangle Park, NC 27711

[3]United States Environmental Protection Agency, Office of Research and Development, National Health and Environmental Effects Research Laboratory, Research Triangle Park, NC 27711

[4]Lockheed Martin, RTP NC

[5]National Center for Advancing Translational Sciences, National Institutes of Health, Bethesda, MD 20892, USA

Key Words: Androgen, Endocrine, Estrogen, High-Throughput, In Vitro, ToxCast

Short Title: High-Throughput Endocrine Disruptor Screening

Disclaimer: The views expressed in this article are those of the authors and do not necessarily reflect the views of policies of the U.S. Environmental Protection Agency. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

¶Corresponding Author
Phone: (919) 541-3085
Fax: (919) 541-3513
Email: judson.richard@epa.gov

Competing Financial Interest Declaration: The authors declare they have no actual or potential competing financial interests.

Acknowledgments

Amar Singh was employed by Lockheed Martin Corporation, Research Triangle Park, NC, as a contractor the EPA/NCCT.

List of Abbreviations

AC50 – Concentration at half-maximum activity

BA – Balanced Accuracy

CASRN – CAS registry number

EDSP – Endocrine Disruptor Screening Program

EPA – Environmental Protection Agency

FN – False Negative

FP – False Positive

FFDCA – Federal Food, Drug, and Cosmetic Act

FQPA – Food Quality Protection Act

HTS – High Throughput Screening

HTS-A – High throughput screening assays for androgen relevant pathways

HTS-E - High throughput screening assays for estrogen relevant pathways

HTS-S - High throughput screening assays for relevant steroidogenesis pathways

HTS-T - High throughput screening assays for thyroid relevant pathways

Guideline-A – EDSP or OECD guideline studies for androgen relevant pathways

Guideline-E - EDSP or OECD guideline studies for estrogen relevant pathways

Guideline-S - EDSP or OECD guideline studies for relevant steroidogenesis pathways

Guideline-T - EDSP or OECD guideline studies for thyroid relevant pathways

MIE – Molecular Initiating Event

MOA- Mode of Action

NRC – National Research Council

OECD – Organization for Economic Co-operation and Development

PMID – PubMed Identification

T1S – Tier 1 Screening

TN – True Negative

TP – True Positive

Abstract

Background: Over the past 20 years, an increased focus on detecting environmental chemicals posing a risk of adverse effects due to endocrine disruption has driven the creation of the U.S. EPA Endocrine Disruptor Screening Program (EDSP). Thousands of chemicals are subject to the EDSP, which could require millions of dollars and decades to process using current test batteries. A need for increased throughput and efficiency motivated the development of methods using *in vitro* high-throughput screening (HTS) assays to prioritize chemicals for EDSP Tier 1 screening (T1S).

Objective: Here we investigate using EPA ToxCast HTS assays for estrogen, androgen, steroidogenic, and thyroid disrupting mechanisms to classify compounds, and compare ToxCast results to *in vitro* and *in vivo* data from EDSP T1S assays.

Method: An iterative model was implemented that optimized the ability of HTS endocrine-related assays to predict components of EDSP T1S and related results. Balanced accuracy was used as a measure of model performance.

Results: ToxCast estrogen and androgen receptor assays predicted the results of relevant EDSP T1S assays with balanced accuracies of 0.91 ($P < 0.001$) and 0.92 ($P < 0.001$), respectively. Uterotrophic and Hershberger assay results were predicted with balanced accuracies of 0.89 ($P < 0.001$) and 1 ($P < 0.001$), respectively. Models for steroidogenic and thyroid-related effects could not be developed with the currently published ToxCast data.

Conclusions: Overall, results suggest that current ToxCast assays can accurately identify chemicals with potential to interact with the estrogenic and androgenic pathways, and could help prioritize chemicals for EDSP T1S assays.

Introduction

Endocrine hormones regulate a diverse set of physiological responses, some of which include sexual dimorphism, reproductive capacity, glucose metabolism, and blood pressure (Cooper and Kavlock 1997; de Mello et al. 2011; Dupont et al. 2000; Lodish et al. 2009; Ng et al. 2001). The wide role of responses regulated by hormones makes them of particular concern for disruption by xenobiotics (Ankley et al. 1998; Colborn et al. 1992; Soto and Sonnenschein 2010; Tilghman et al. 2010). Endocrine disruption can lead to many adverse consequences, some of which include altered reproductive performance, and hormonally mediated cancers (Birnbaum and Fenton 2003; Kavlock et al. 1996; Soto and Sonnenschein 2010; Spencer et al. 2011). Endocrine disruption can also have adverse effects on the fetus or newborn due to the delicate balance of hormones required during critical developmental windows (Bigsby et al. 1999; Chandrasekar et al. 2011; Cooper and Kavlock 1997; Mahoney and Padmanabhan 2010). For example, studies have demonstrated that thyroid hormone insufficiency during pregnancy may lead to adverse neurological outcomes in children (Haddow et al. 1999).

The Federal Food, Drug, and Cosmetic Act (FFDCA), as amended by the Food Quality Protection Act (FQPA), and the Safe Drinking Water Act (SDWA) requires the U.S. Environmental Protection Agency (EPA) to determine whether certain substances may have an effect in humans that is similar to an effect produced by a naturally occurring estrogen, or other such endocrine effects. In response, the EPA formed the Endocrine Disruptor Screening Program (EDSP) (www.epa.gov/endo/) (U.S. EPA 2012a). EDSP is a two-tiered program that requires chemical manufacturers to submit or generate data on a suite of both *in vivo* and *in vitro* assays. The first phase of EDSP assays are designated as the Tier 1 screening battery (T1S) (www.epa.gov/endo/pubs/assayvalidation/tier1battery.htm) (U.S. EPA 2012b). These tests

identify chemicals with the potential to interact with endocrine pathways or mechanisms, and focus on disruption of estrogen, androgen, and thyroid hormone pathways. Based on a weight-of-evidence approach, chemicals showing positive activity in T1S assays could then be subject to more complex Tier 2 tests (U.S. EPA 2011a). The European Commission is continuing the implementation of the European Union's Community Strategy for endocrine disruptors, which includes the establishment of a priority list of substances for further evaluation and assay development and validation. (http://ec.europa.eu/environment/endocrine/index_en.htm) (European Commission 2012). In addition, the European Commission is working towards defining specific criteria to identify endocrine disruptors within a legislative framework, drawing on current scientific opinion (Kortenkamp et al. 2011).

The EPA estimates that the statutory requirements and discretionary authorities through passage of the FQPA and its amendments, and SDWA will require the EDSP to screen as many as 9700 environmental chemicals, and generating this data required under the current testing guidelines will be expensive, time-consuming and will require significant animal resources (U.S. EPA 2011b). To date, chemicals have been nominated for EDSP T1S on the basis of exposure potential, or registration status. Due to these fiscal and time constraints, EPA is considering using endocrine-related *in vitro* high-throughput screening (HTS) assays and *in silico* models to prioritize chemicals for testing in T1S (U.S. EPA 2011b). There has been a significant improvement in HTS technologies since EPA began work on developing and implementing the EDSP. In 2007, the National Research Council (NRC) Report ''Toxicity Testing in the 21st Century: A Vision and a Strategy'' (NRC 2007) acknowledged these advances and recommended that the Agency develop a strategy to use modern molecular-based screening methods to reduce, and ultimately replace, the reliance on whole-animal toxicity testing. The US

151 EPA's ToxCast program (http://epa.gov/ncct/toxcast/) (U.S. EPA 2012c), and the U.S.
152 government's cross-agency Tox21 program (http://www.epa.gov/ncct/Tox21/) are using HTS
153 assays and developing computational tools to predict chemical hazard, to characterize a diverse
154 set of toxicity pathways, and to prioritize the toxicity testing of environmental chemicals (Huang
155 et al. 2011, U.S. EPA 2012d). Included in these programs are assays that cover toxicity pathways
156 involving estrogen, androgen and thyroid hormone receptors, and targets within the
157 steroidogenesis pathway. The current ToxCast chemical library covers approximately 17% of
158 chemicals subject to EDSP, and the larger Tox21 chemical library covers approximately 53% of
159 chemicals subject to EDSP. Assay technologies include competitive binding, reporter gene, and
160 enzyme inhibition assays. The comparison of HTS assays, endocrine-related modes of action
161 (MOA) and EDSP T1S is shown in Figure 1. An endocrine MOA consists of a series of
162 molecular initiating events (MIE) relevant for estrogen, androgen, thyroid, or steroidogenic
163 pathways. These assays do not represent their respective MOA in its entirety, but used to detect
164 chemicals capable of perturbing a particular MOA. The present study investigated the predictive
165 ability of ToxCast HTS assays for endpoints tested in EDSP T1S, and tested the hypothesis that
166 if a chemical activates the estrogen or androgen receptor *in vitro* that estrogen and androgen
167 related effects will occur in *in vivo* bioassays. Ideally, HTS tests should be highly reproducible,
168 and yield a minimal number of false positive (specificity) and false negative (sensitivity)
169 chemicals.

170 Previous studies have suggested the use of HTS assays for identifying endocrine
171 disrupting potential. For example, the ReProTect project developed within the 6th European
172 Framework Program tested 14 *in vitro* assays using 10 prototype compounds to determine
173 feasibility for a reproductive screening program (Schenk et al. 2010). Those *in vitro* assays were

grouped into three segments of the reproductive cycle: endocrine disruption, fertility, and embryonic development. The results of ReProTect showed, at least for a limited number of reference chemicals, that an appropriate *in vitro* assay selection can effectively group compounds based on known reproductive toxicity (Schenk et al. 2010).

HTS assays are useful for identifying toxicological molecular initiating events (MIE) which represent a series of key events relevant for a given biological or toxicity pathway. Through a combination of competitive ligand binding, reporter gene, and enzyme inhibition assays, the ability to predict the potential for chemicals to cause endocrine disruption can be determined through investigating associations with results from guideline EDSP T1S assays and other relevant endocrine studies. The aim of the current study was to determine the feasibility of using HTS *in vitro* assays for prioritizing chemicals for T1S by investigating the associations with data captured from guideline and non- guideline endocrine related study reports, and to determine which endocrine MOA are suitable for more sophisticated model development.

Methods

Chemical Selection. This study used data on the Phase I ToxCast chemical library with data for 309 unique chemical structures (U.S. EPA 2008). The majority of these chemicals are either current or former food-use pesticide active ingredients designed to be bioactive, or industrial chemicals and of environmental relevance. Details of the chemical library are given in Judson et al. (2009). Data on an additional 23 reference chemicals were included that were tested in a separate study (Judson et al. 2010), 17 of which were not in the ToxCast Phase I library. CAS registry numbers (CASRN) for the ToxCast Phase 1 and the additional 17 chemicals are provided as supplemental material (See Supplemental Material, File 1).

196 ***Guideline and Non-Guideline Endocrine Assays:*** Data from guideline endocrine-related *in vitro*
197 and *in vivo* studies were extracted from EDSP Tier 1 validation reports from the EPA EDSP
198 website (http://www.epa.gov/endo/pubs/assayvalidation/status.htm) (U.S. EPA 2012e). Non-
199 guideline studies were obtained from open literature by querying PubMed and Google Scholar
200 resources using the terms {any chemical name or CASRN in the 309} AND {in vitro or in vivo}
201 AND {estrogen or androgen or uterotrophic or Hershberger or steroidogenesis or thyroid
202 hormone}. A wide variety of studies were returned from the automated search. The list of studies
203 was manually curated to remove studies that did not contain data usable for the current analysis
204 (e.g. studies of mixtures without testing compounds individually, studies that mentioned the
205 chemical but did not test it in a bioassay, studies measuring bioaccumulation). Studies that
206 identified their methods as following: The Organization for Economic Co-operation and
207 Development (OECD) guidelines (Kanno et al. 2001; Kanno et al. 2003; OECD 1999; OECD
208 2001; OECD 2003; OECD 2007) or EDSP protocols were grouped together with EDSP T1S data
209 for the guideline analysis. When available, PubMed identifiers (PMID) were used as unique
210 annotations for each report. For the few instances when no PMID was available or for each
211 EDSP T1S validation report, a unique identifying number was generated. The citation
212 information for all documents used in the analysis is provided in Supplemental Material, File 2.

213 Guideline endocrine related assays gathered from EDSP validation reports and OECD
214 guideline studies were categorized according to whether they tested estrogen, androgen,
215 steroidogenesis, or thyroid related MOA (Guideline-E, Guideline-A, Guideline-S, Guideline-T,
216 respectively). Additional information captured included: study type (e.g. amphibian
217 metamorphosis, reporter gene, etc.), assay type (e.g. serum levels, organ weight, etc.), species,
218 strain, cell type, target, and whether or not it was an EDSP/OECD guideline study. Chemical

219 potency (e.g. AC50, LEC) for a given endpoint was captured in whatever way it was represented
220 in the study report along with the maximum concentration/dose tested. Additionally, agonist or
221 antagonist responses were noted when applicable. Data from guideline and non-guideline studies
222 were dichotomized as either active if a response was observed, or inactive if no response was
223 observed. If a study investigated multiple endpoints for a given endocrine MOA and produced at
224 least one statistically significant endpoint, then that study-chemical-MOA combination was
225 considered active. Activity/Inactivity was determined based on the presence of a statistically
226 significant response or was based on the study author's conclusion. Data was further annotated
227 as either a hit value of 1 or 0, if it was active or inactive, respectively. All guideline and non-
228 guideline literature studies were combined so as to have a single hit value for each study-
229 chemical-MOA combination. Data that was conflicting or otherwise unclear was included in the
230 data table, but was annotated as such, and removed from analyses. The data obtained from
231 guideline endocrine-related studies and other non-guideline literature reports is included in
232 Supplemental Material, File 3.

233 ***ToxCast In Vitro Assays.*** HTS Competitive binding, enzyme inhibition, and reporter gene assays
234 representing estrogen, androgen, steroidogenesis, or thyroid related endpoints (HTS-E, HTS-A,
235 HTS-S, HTS-T, respectively) were selected as a subset of the >500 HTS assays generated by the
236 ToxCast program (ToxCastDB v.17) (http://www.epa.gov/ncct/toxcast/data.html) (See
237 Supplemental Material, File 1). The details and a description of each assay are reported in Table
238 1.

239 For chemicals that produced a statistically-significant and concentration-dependent
240 response in a given assay, the concentration at half-maximum activity (AC50) was recorded.
241 The criteria for determining the activity of a compound are assay platform dependent and further

details can be found in Supplemental Material. The data was then dichotomized so that if an AC50 was present for a given chemical-endpoint concentration a 1 was reported or if no response was observed then a 0 was reported. Triplicated chemicals were designated 1 or 0 on a majority basis. Chemicals run in duplicate with at least one sample producing an AC50 were designated as a 1. Experimental methods for each assay used are provided in the Supplemental Material.

Model Development. An iterative, balanced optimization analysis was performed to determine the ability of ToxCast HTS assays to correctly classify the results of guideline endocrine-related assays, while maintaining balance between sensitivity and specificity. The process by which this was performed is illustrated in Figure 2. Because each HTS endocrine MOA may have multiple ToxCast HTS assays, disjunctive logic employing varied "weight-of-evidence" thresholds were used to determine optimal predictive performance. This model tested variable thresholds for the HTS ToxCast assay results represented as un-weighted binary data, while the guideline or non-guideline endocrine-related assay results remained static. Initially, the model began with a threshold criterion of 1 positive ToxCast HTS assay out of the total number of ToxCast HTS assays for a chemical to be considered to perturb a given MOA. Once calculated, the model was then re-run with increasing increments of 1 assay until all ToxCast HTS assays for a given endocrine MOA were required to be positive for a chemical to be considered to perturb the given MOA. As the threshold for a positive call was increased, a larger weight of evidence was required for a chemical to be considered a "hit" for perturbing the given endocrine MOA. An exception was made for guideline pubertal studies and the ToxCast NVS_NR_hAR assay. Guideline pubertal studies test for effects that can arise through multiple different endocrine related pathways. For this reason, if a chemical was considered positive in the pubertal assay and

the result conflicted with other guideline studies (e.g. receptor binding, reporter gene), the pubertal assay was not included in the weight-of-evidence. The ToxCast NVS_NR_hAR assay is a human androgen receptor binding assay in the LNCaP prostatic cell line. The androgen receptor in this cell line is known to bind to steroid hormones other than androgens (Veldscholte et al. 1992). For this reason, if a compound was negative in all other HTS-A assays, the result for the NVS_NR_hAR assay was not included in the weight-of-evidence.

For a specific set of criteria across all overlapping chemicals, sensitivity, specificity, and balanced accuracy (BA) were calculated as measures of model performance to compare (see the contingency table in Figure 2B). The guideline analysis was performed comparing ToxCast HTS assays and guideline endocrine assays gathered from EDSP validation reports/OECD guideline studies. A separate non-guideline analysis comparing ToxCast HTS assays with assays from non-guideline studies was also conducted. Many studies in the EDSP/OECD guideline studies and non-guideline literature have multiple studies/assays for each chemical-MOA combination. Because separate studies are not always in agreement relative to a chemical-MOA perturbation, the model was run using two scenarios: 1) any positive report for a chemical resulted in a positive call for the chemical-MOA combination, or 2) greater than 50% (threshold > 0.50) of guideline or non-guideline endocrine-related studies/assays must report the chemical to be active for a given endocrine MOA.

For each threshold criteria the number of true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) were calculated. A TP was any chemical that was determined to be positive with the ToxCast HTS assays and was also positive in guideline endocrine reports. A FP was any chemical determined to be positive in ToxCast but reported as negative in the guideline endocrine reports. If a chemical was determined negative in the

ToxCast HTS assays and positive in the guideline endocrine reports then it was recorded as a FN. Lastly, a TN was any chemical that was determined to be negative in the ToxCast HTS assays and reported to be negative in the guideline endocrine reports. At each threshold combination, all of the available chemicals were classified as TP, FP, TN, or FN and were used to calculate sensitivity, specificity, and BA as a measure of model performance.

Statistical Analysis. In order to identify statistically significant BA values, a permutation test was performed. The test randomized which ToxCast assays were associated with guideline endocrine studies or biomedical literature for each endocrine MOA in order to determine whether or not a randomly chosen set of assays from the >500 ToxCast endpoints would likely produce a similar association. The BA calculation based on random assay associations was performed using the same number of ToxCast assays as the model and with the same threshold criteria. Assays were permuted 10,000 times to build the random BA population distribution, and the percentile where the model BA fell among this distribution was calculated to provide a *P* value. A *P* value of < 0.01 was considered statistically significant. The distributions developed from the permutation tests were used to define the confidence intervals in Figures 3 and 4.

Results

Data Collection. Data covering guideline endocrine-related *in vitro* and *in vivo* assays was extracted from documents used in EDSP Tier 1 validation or conducted according to OECD guidelines. There were a total of 40 studies covering 154 unique chemicals, resulting in a total of 1246 captured endpoints. Table 2 shows the chemical overlap between the ToxCast chemical library and the chemicals captured from guideline and non-guideline studies. There were 21 chemicals available from EDSP validation documents and other OECD guideline studies

covering the Guideline-E MOA that overlapped with the ToxCast HTS-E assays. There were 13 chemicals overlapping in the corresponding Guideline-A assays, 8 in the T assays and 17 in the S assays. Additional data used in a separate analysis was extracted from a total of 215 non-guideline studies (See Supplemental Material, File 3).

Model Results. The results presented in Figure 3 demonstrate the predictive ability of ToxCast HTS-E and HTS-A assays for corresponding endocrine MOA in the guideline endocrine-related studies. Detailed results from the univariate model with guideline studies are provided in Supplemental Material, File 4.

HTS and Guideline Endocrine Assay Comparisons: For HTS-E endpoints, an optimal BA of 0.91 ($P < 0.001$) was obtained with a sensitivity of 0.89 and specificity of 0.92 with a threshold of 2 positives for ToxCast HTS-E assays and >50% for Guideline-E studies (Fig. 3). This means a minimum of 2 ToxCast HTS-E assays must report an AC50 value for a chemical to be considered positive; and greater than 50% of Guideline-E assays must be reported as positive in the EDSP validation reports or OECD guideline studies. A table of overlapping HTS-E and HTS-A chemicals and corresponding performance in the HTS and guideline studies is provided in Supplemental Material, Tables 3 and 4. There were 21 Guideline-E related chemicals that overlapped with the ToxCast Phase I chemicals. One chemical, chlorpyrifos-methyl (5598-13-0), was misclassified as a positive (FP) and one chemical, prochloraz (67747-09-5), was misclassified as a negative (FN) by this set of ToxCast assays. If the goal was to optimize sensitivity, threshold criteria of 1 ToxCast HTS-E assay and >50% of Guideline-E would produce a perfect sensitivity of 1 but specificity drops to 0.5 across this set of ToxCast HTS-E assays (See Supplemental Material, File 4). An additional analysis was conducted lowering the threshold criteria for the Guideline-E assays from >50% to any single positive report resulted in

a positive call. This lowers the sensitivity from 0.89 to 0.5 and the overall BA drops to 0.75 (Fig. 3).

Figure 3 demonstrates the predictive ability of the ToxCast HTS-A assays with the Guideline-A results. The optimal predictive ability of the ToxCast HTS-A assays was reached with a threshold of 1 HTS-A assay and a threshold > 50% for the Guideline-A assays. This set of criteria produced a BA of 0.92 ($P < 0.001$) with a sensitivity of 0.83 and specificity of 1 (See Supplemental Material, Table 4). The results for HTS-S and HTS-T were not statistically significant among any of the analyses with BA of .56 ($P > 0.01$) and .50 ($P > 0.01$), respectively (See Supplemental Material, File 4).

HTS and Uterotrophic and Hershberger Comparisons: A separate analysis was conducted to determine the predictive capability of the ToxCast HTS-E assays to detect positive and negative chemicals reported in EDSP/OECD guideline uterotrophic assays (Fig. 3). 18 chemicals were available for comparison and the optimal thresholds for HTS-E produced a BA of 0.9 ($P < 0.001$) with a sensitivity and specificity of 0.88 and 0.9, respectively.

Additionally, the predictive ability of ToxCast HTS-A assays for EDSP/OECD guideline Hershberger results was determined. Although, only 6 chemicals were available for comparison, the analysis resulted in a BA of 1 ($P < 0.001$) with perfect measure of sensitivity and specificity with thresholds of 1 positive assay required for both HTS-A and EDSP/OECD guideline Hershberger reports (Fig. 3).

HTS and Non-Guideline Study Comparisons: Predictive modeling results for non-guideline studies in the biomedical literature are presented in Figure 4. All results from the analysis with non-guideline studies are provided in supplemental material, File 5. The HTS-E MOA produced

a maximum BA of 0.74 ($P < 0.01$) with at least one ToxCast assay being positive (ToxCast HTS-E threshold of 1) and a literature threshold of >50%. These criteria produced a sensitivity of 0.75 and a specificity of 0.72. Due to the wide range of test conditions, assay technologies, and species present in the open-literature there was a loss of sensitivity compared to the guideline studies. This is apparent due to the model optimization occurring with only a single HTS-E assay required for a positive classification, as opposed to optimizing at two assays in the guideline analysis. There was an overall concordance of 0.7 between the Guideline-E assay results and the estrogen-related literature results given the stated thresholds (Data not shown).

The optimal BA reached 0.65 ($P > 0.01$) with ToxCast HTS-A assays threshold of 1, and an androgen-related literature threshold >50%. At these thresholds, there was a low sensitivity (0.3), but a perfect specificity of 1 (Fig. 4). There was a concordance between chemical classifications for Guideline-A reports and non-guideline reports of 0.77 at the reported thresholds of >50% (Data not shown).

Discussion

The results of this study demonstrate that ToxCast *in vitro* assays perform adequately to prioritize chemicals for further EDSP T1S for estrogen and androgen activity, and these HTS assays are predictive of the likelihood of a positive or negative finding in more resource-intensive assays. Additional HTS assays will be needed to predict steroidogenic and thyroid activity of chemicals. Methods for prioritizing chemicals based on a broad range of ToxCast HTS assays, in combination with physical-chemical properties, have been previously developed (Reif et al. 2010). Other efforts are also underway to develop more sophisticated, pathway-based, predictive models that would be more suitable to support regulatory decision making.

This study demonstrates what MOA these models would be expected to be successful, and which areas need additional technologies before a sufficient screening tool would be expected to succeed. This information could be used for more focused follow-up efforts which can identify some endocrine related MOAs for prioritization.

The HTS-E and HTS-A assays demonstrate a high degree of association with the Guideline-E and Guideline-A assays. The two types of misclassifications, FP and FN, are important because they highlight shortcomings in the model or further specify the domain of applicability. FP are compounds predicted to be active, but were not active in this analysis based on the threshold of EDSP/OECD reports or literature data. These are significant because a FP could lead to unnecessary testing in more resource intensive assays, and a FN is of concern because they represent potentially active chemicals that would have gone undetected.

The HTS-E model correctly classified 90% of chemicals, and only two out of 21 chemicals were misclassified as FP or FN. Chlorpyrifos-methyl was a FP, meaning that it was predicted to be estrogenic by ToxCast HTS-E assays but was not positive in the only Guideline-E report, which was a uterotrophic study by Kang et al. (Kang et al. 2004) (See Supplemental Material, Table 3). This same chemical was reported to be inactive in all of the extracted non-guideline-E literature data (active in 0/4 available assays). Chlorpyrifos-methyl was inactive in all ToxCast HTS-E assays except for the Attagene ERα TRANS and CIS reporter gene assays, which resulted in the subsequent positive call.

Non-guideline estrogen-related literature for prochloraz reported observations of ERα antagonism in some reporter gene and proliferation assays (Bonefeld-Jorgensen et al. 2005; Kjaerstad et al. 2010), but other studies did not observe activity in reporter gene assays

(Andersen et al. 2002; Kojima et al. 2004; Lemaire et al. 2006; Petit et al. 1997) or proliferation assays (Andersen et al. 2002; Vinggaard et al. 1999) (See Supplemental Material, File 3). Prochloraz was a FN in this analysis, as it was active in the NCGC ERα antagonist assay, but negative in all other ToxCast HTS-E binding and reporter gene assays (See Supplemental Material, File 1). Prochloraz tested positive in the only Guideline-E assay available (See Supplemental Material, Table 3). This EDSP/OECD fathead minnow assay showed altered fecundity, vitellogenin, and oocyte atresia after prochloraz treatment (U.S. EPA 2007). Prochloraz is known to disrupt steroidogenesis through CYP 17 hydroxylase and aromatase inhibition, preventing the critical conversion of progesterone to 17α-hydroxyprogesterone and testosterone to 17β-estradiol, respectively (Blystone et al. 2007; Sanderson et al. 2002). The fathead minnow assay likely detected this non-receptor mediated mechanism of estrogen disruption- and this mechanism of action would not have been expected to be detected in the current set of ToxCast HTS-E assays. Prochloraz was the only compound misclassified in the HTS-A analysis, and the effects observed in the male fish reproductive study are likely a result of the same steroidogenic perturbations. Prochloraz was correctly identified by ToxCast aromatase enzyme inhibition assay which was grouped with the HTS-S related MOA.

Although a limited number of chemicals were available for comparison, we found a strong association between the ToxCast HTS-E and HTS-A assays with EDSP/OECD guideline uterotrophic and Hershberger studies. 18 chemicals were available for comparison between ToxCast HTS-E and guideline uterotrophic assays and only 2 were misclassified (See Supplemental Material, Table 3). There were 6 chemicals available for analysis between ToxCast HTS-A assays and Hershberger responses, and all chemicals were classified correctly for a perfect BA of 1 (See Supplemental Material, Table 4).

There are several explanations for why a chemical may be misclassified by the ToxCast HTS models. In some scenarios a chemical may not have been tested to very high concentrations, at which they may exhibit a response in ToxCast assays. Inconsistencies could also result from species, tissue, or cell-type differences between the ToxCast versus guideline studies. Most of the ToxCast assays use human cell lines or reporter constructs, and some areas of misclassification may be due to species differences between these assays and the rodent bioassays. Not only should interspecies differences be taken into consideration, but the intraspecies differences may also be quite substantial. For example, studies have highlighted not only the importance of tissue and cell distribution and context within an organism for both ER and AR (Kolasa et al. 2003; Zhou et al. 2002), but also the presence of ERα and ERβ splice variants (Saunders et al. 2002). Most *in vitro* assays are limited in their metabolic capabilities, so chemicals that require metabolic activation in order to be active may not be detected. However, methoxychlor and vinclozolin, which become more active with metabolism, were both detected in the HTS-E (See Supplemental Material, Table 3) and HTS-A (See Supplemental Material, Table 4) assays, respectively. Furthermore, *in vivo* assays may detect chemicals that perturb endocrine related endpoints elicited via toxicity in other organs, such as the liver (Leffert and Alexander 1976; Masuyama et al. 2000; Xie et al. 2003). The assays selected for this study comprise only a small portion of the overall endocrine pathway domain. Alterations in neuroendocrine or other pathways, as well as some feedback mechanisms, could be affected by a compound and would not be detected by these assays. The methods used to classify compounds in this study may result in different conclusions than those obtained by EDSP (U.S. EPA 2011a). Despite these limitations, evidence from this study indicates that very few chemicals that are active in EDSP T1S go undetected by ToxCast HTS-E and HTS-A assays. The majority of

misclassifications appear to be from downstream estrogenic and androgenic effects caused by alterations of upstream steroidogenic enzymes. The majority of active Guideline-E and Guideline-A chemicals in this dataset appear to operate through receptor mediated pathways and are detectable *in vitro*.

The non-guideline literature analysis demonstrated that ToxCast HTS assays are also predictive of a broader range of endocrine-related assays. As expected, there was a loss of accuracy in predicting the non-guideline literature analysis when compared to the EDSP/OECD guideline studies due to the wide variety of species tested, assay protocols, and technologies implemented in the non-guideline literature reports. An additional factor that led to the loss of sensitivity in the HTS-A non-guideline analysis was the imbalance of positive to negative reports. The guideline study had 6 positives out of 13 total chemicals (46%) at >50% threshold and the non-guideline report had 47 positives out of 59 total chemicals (80%) at the same threshold. The sensitivity would be expected to improve with a more balanced dataset.

Based on this analysis, there is a clear need to develop HTS assays capable of detecting steroidogenesis and thyroid disrupting compounds. The current HTS-S related assay within ToxCast is limited to a single cell-free aromatase enzyme activity assay. Aromatase is a key enzyme in the biosynthesis of estrogens from androgens (Schuurmans et al. 1991; Stoker et al. 2000a). However, in addition to aromatase inhibition there are other mechanisms of steroidogenesis that may be impacted by environmental chemicals that are not tested in our current HTS battery (Stoker et al. 2000a; Stoker et al. 2000b). Additional assay technologies that may provide a more comprehensive set of steroidogenesis endpoints are currently being assessed.

The ToxCast HTS-T assays used in this analysis are composed of thyroid hormone receptor binding and reporter gene assays. Only a limited number of chemicals were available for comparison between the HTS-T assays and the guideline studies. The inability of the ToxCast HTS-T assay results to associate with compounds thought to disrupt thyroid homeostasis in EDSP/OECD guideline studies, suggest that most of these compounds are not acting through thyroid hormone receptor-mediated mechanisms (Paul et al. 2010; Zorrilla et al. 2009). Thyroid hormone homeostasis has been shown to be altered through enhanced or suppressed clearance of thyroid hormone by metabolic enzymes (Saghir et al. 2008; Zorrilla et al. 2009) . ToxCast contains HTS assays measuring nuclear receptor activation and metabolic enzyme activity, which could be relevant for thyroid hormone metabolism. However, many chemicals that activated these in vitro ToxCast assays were not associated with adverse outcomes in the in vivo literature captured by this study, and the subsequent lack of specificity for thyroid active chemicals led to their exclusion from this analysis (Data not shown).

From these findings, we can conclude that most chemicals chosen to validate EDSP T1S assays alter estrogen and/or androgen related endpoints through nuclear receptor-mediated mechanisms and are capable of being efficiently detected by the ToxCast HTS assays. For the purpose of prioritization, it is important to establish sufficient confidence that the assays being utilized are specific and sensitive so that chemicals prioritized for EDSP T1S include those most likely to be active. Although further efforts are needed to improve the ability to detect steroidogenic and thyroid-disrupting chemicals with *in vitro* test systems; these results indicate that ToxCast endocrine assays are highly predictive of chemicals with estrogenic and androgenic receptor-based endocrine MOA, and their use for endocrine testing would allow efficient prioritizing of chemicals for further testing.

References

Andersen HR, Vinggaard AM, Rasmussen TH, Gjermandsen IM, Bonefeld-Jorgensen EC. 2002. Effects of currently used pesticides in assays for estrogenicity, androgenicity, and aromatase activity in vitro. Toxicol Appl Pharmacol 179(1):1-12.

Ankley, G. T., and Giesy, J. P. 1998. Endocrine disruptors in wildlife: a weight of evidence perspective. In Principles and Processes for Evaluating Endocrine Disruption in Wildlife (R. Kendall, R. Dickerson, J. Giesy, and W. Suk, Eds.), pp. 349–367. SETAC Press, Pensacola.

Bigsby R, Chapin RE, Daston GP, Davis BJ, Gorski J, Gray LE, et al. 1999. Evaluating the effects of endocrine disruptors on endocrine function during development. Environ Health Perspect 107 Suppl 4:613-618.

Birnbaum LS, Fenton SE. 2003. Cancer and developmental exposure to endocrine disruptors. Environ Health Perspect 111(4):389-394.

Blystone CR, Furr J, Lambright CS, Howdeshell KL, Ryan BC, Wilson VS, et al. 2007. Prochloraz inhibits testosterone production at dosages below those that affect androgen-dependent organ weights or the onset of puberty in the male Sprague Dawley rat. Toxicol Sci 97(1):65-74.

Bonefeld-Jorgensen EC, Grunfeld HT, Gjermandsen IM. 2005. Effect of pesticides on estrogen receptor transactivation in vitro: a comparison of stable transfected MVLN and transient transfected MCF-7 cells. Mol Cell Endocrinol 244(1-2):20-30.

Chandrasekar G, Arner A, Kitambi SS, Dahlman-Wright K, Lendahl MA. 2011. Developmental toxicity of the environmental pollutant 4-nonylphenol in zebrafish. Neurotoxicol Teratol 33(6):752-64.

Colborn T., Clement C. 1992. Chemically-induced alterations in sexual and functional development: the wildlife/human connection. Advances in modern environmental toxicology 21:403.

Cooper RL, Kavlock RJ. 1997. Endocrine disruptors and reproductive development: a weight-of-evidence overview. J Endocrinol 152(2):159-166.

de Mello WG, de Morais SR, Dornelles RC, Kagohara Elias LL, Antunes-Rodrigues J, Bedran de Castro JC. 2011. Effects of neonatal castration and androgenization on sexual dimorphism in bone, leptin and corticosterone secretion. Bone 50(4):893-900.

Dupont S, Krust A, Gansmuller A, Dierich A, Chambon P, Mark M. 2000. Effect of single and compound knockouts of estrogen receptors alpha (ERalpha) and beta (ERbeta) on mouse reproductive phenotypes. Development 127(19):4277-4291.

529 European Commission. 2012. Endocrine Disruptors Website: How the European Commission
530 Addresses Challenges Posed by Endocrine Disruptors. Available:
531 http://ec.europa.eu/environment/endocrine/index_en.htm [accessed July 3, 2012].

532
533 Haddow JE, Palomaki GE, Allan WC, Williams JR, Knight GJ, Gagnon J, et al. 1999. Maternal
534 thyroid deficiency during pregnancy and subsequent neuropsychological development of the
535 child. N Engl J Med 341(8):549-555.

536 Huang R, Xia M, Cho MH, Sakamuru S, Shinn P, Houck KA, et al. 2011. Chemical genomics
537 profiling of environmental chemical modulation of human nuclear receptors. Environ Health
538 Perspect 119(8):1142-8.

539 Judson RS, Martin MT, Reif DM, Houck KA, Knudsen TB, Rotroff DM, et al. 2010. Analysis of
540 eight oil spill dispersants using rapid, in vitro tests for endocrine and other biological activity.
541 Environ Sci Technol 44(15):5979-5985.

542 Judson RS, Richard A, Dix DJ, Houck KA, Martin MT, Kavlock RJ, et al. 2009. The toxicity
543 data landscape for environmental chemicals. Environ Health Perspect. 117(5):685-95.

544 Kang HG, Jeong SH, Cho JH, Kim DG, Park JM, Cho MH. 2004. Chlropyrifos-methyl shows
545 anti-androgenic activity without estrogenic activity in rats. Toxicology 199(2-3):219-230.

546 Kanno J, Onyon L, Haseman J, Fenner-Crisp P, Ashby J, Owens W. 2001. The OECD program
547 to validate the rat uterotrophic bioassay to screen compounds for in vivo estrogenic responses:
548 phase 1. Environ Health Perspect 109(8):785-794.

549 Kanno J, Onyon L, Peddada S, Ashby J, Jacob E, Owens W. 2003. The OECD program to
550 validate the rat uterotrophic bioassay. Phase 2: dose-response studies. Environ Health Perspect
551 111(12):1530-1549.

552 Kavlock RJ, Daston GP, DeRosa C, Fenner-Crisp P, Gray LE, Kaattari S, et al. 1996. Research
553 needs for the risk assessment of health and environmental effects of endocrine disruptors: a
554 report of the U.S. EPA-sponsored workshop. Environ Health Perspect 104 Suppl 4:715-740.

555 Kjaerstad MB, Taxvig C, Nellemann C, Vinggaard AM, Andersen HR. 2010. Endocrine
556 disrupting effects in vitro of conazole antifungals used as pesticides and pharmaceuticals. Reprod
557 Toxicol 30(4):573-582.

558 Kojima H, Katsura E, Takeuchi S, Niiyama K, Kobayashi K. 2004. Screening for estrogen and
559 androgen receptor activities in 200 pesticides by in vitro reporter gene assays using Chinese
560 hamster ovary cells. Environ Health Perspect 112(5):524-531.

561 Kolasa A, Wiszniewska B, Marchlewicz M, Wenda-Rozewicka L. 2003. Localisation of
562 oestrogen receptors (ERalpha and ERbeta) in the human and rat epididymides. Folia Morphol
563 (Warsz) 62(4):467-469.

564 Kortencamp A, Martin O, Faust M, Evans R, McKinlay R, Orton F, et al. 2011.State of the Art
565 Assessment of Endocrine Disruptors: Final Report. European Commission.

http://ec.europa.eu/environment/endocrine/documents/4_SOTA%20EDC%20Final%20Report%20V3%206%20Feb%2012.pdf [accessed 07 April 2012].

Leffert HL, Alexander NM.1976. Thyroid hormone metabolism during liver regeneration in rats. Endocrinology 98(5):1241-1247.

Lemaire G, Mnif W, Mauvais P, Balaguer P, Rahmani R. 2006. Activation of alpha- and beta-estrogen receptors by persistent pesticides in reporter cell lines. Life Sci 79(12):1160-1169.

Lodish MB, Sinaii N, Patronas N, Batista DL, Keil M, Samuel J, Moran J, et al. 2009. Blood pressure in pediatric patients with Cushing syndrome. J Clin Endocrinol Metab. 94(6):2002-8.

Mahoney MM, Padmanabhan V. 2010. Developmental programming: impact of fetal exposure to endocrine-disrupting chemicals on gonadotropin-releasing hormone and estrogen receptor mRNA in sheep hypothalamus. Toxicol Appl Pharmacol.1;247(2):98-104.

Masuyama H, Hiramatsu Y, Kunitomi M, Kudo T, MacDonald PN. 2000. Endocrine disrupting chemicals, phthalic acid and nonylphenol, activate Pregnane X receptor-mediated transcription. Mol Endocrinol 14(3):421-428.

National Research Council.2007. Toxicity Testing in the 21st Century: A Vision and a Strategy. The National Academies of Science. Washington, D.C.

Ng PC, Lam CW, Fok TF, Lee CH, Ma KC, Chan IH, et al. 2001. Refractory hypotension in preterm infants with adrenocortical insufficiency. Arch Dis Child Fetal Neonatal Ed 84(2): F122-124.

OECD. 2003. OECD Draft Report of the Validation of the Rat Uterotrophic Bioassay: Phase 2.Testing of Potent and Weak Oestrogen Agonists by Multiple Laboratories. Environment Directorate Joint Meeting of the Chemicals Committee and the Working Party on Chemicals, Pesticides and Biotechnology. Available: http://www.epa.gov/endo/pubs/edmvac/uterotrophic_oecd_rodent_validation_p2_3_5_2003.pdf [accessed 19 April 2012].

OECD. 2007. Test Guideline 440: OECD Guideline for Testing of Chemicals. Uterotrophic Bioassay in Rodents: A short-term screening test for oestrogenic properties. Available: http://www.epa.gov/endo/pubs/uterotrophic_OECD_guideline.pdf [accessed 19 April 2012].

Organization for Economic Cooperation and Development. 1999. Validation protocol for the uterotrophic assay. Paris: Task Force on Endocrine Disruptor Testing and Assessment, OECD.

Organization for Economic Cooperation and Development. 2001. Progress of work on the validation of the rodent uterotrophic bioassay: Phase two status report (including statistical evaluation of the results). Paris: Endocrine Disruptor Testing and Assessment, OECD.

Paul KB, Hedge JM, DeVito MJ, Crofton KM. 2010. Short-term exposure to triclosan decreases thyroxine in vivo via upregulation of hepatic catabolism in Young Long-Evans rats. Toxicol Sci 113(2):367-379.

602 Petit F, Le Goff P, Cravedi JP, Valotaire Y, Pakdel F. 1997. Two complementary bioassays for
603 screening the estrogenic potency of xenobiotics: recombinant yeast for trout estrogen receptor
604 and trout hepatocyte cultures. J Mol Endocrinol 19(3):321-335.

605 Reif DM, Martin MT, Tan SW, Houck KA, Judson RS, Richard AM, et al. 2010. Endocrine
606 profiling and prioritization of environmental chemicals using ToxCast data. Environ Health
607 Perspect 118(12):1714-1720.

608 Saghir SA, Charles GD, Bartels MJ, Kan LH, Dryzga MD, Brzak KA, et al. 2008. Mechanism of
609 trifluralin-induced thyroid tumors in rats. Toxicol Lett 180(1):38-45.

610 Sanderson JT, Boerma J, Lansbergen GW, van den Berg M. 2002. Induction and inhibition of
611 aromatase (CYP19) activity by various classes of pesticides in H295R human adrenocortical
612 carcinoma cells. Toxicol Appl Pharmacol 182(1):44-54.

613 Saunders PT, Millar MR, Macpherson S, Irvine DS, Groome NP, Evans LR, et al. 2002. ERbeta1
614 and the ERbeta2 splice variant (ERbetacx/beta2) are expressed in distinct cell populations in the
615 adult human testis. J Clin Endocrinol Metab 87(6):2706-2715.

616 Schenk B, Weimer M, Bremer S, van der Burg B, Cortvrindt R, Freyberger A, et al. 2010. The
617 ReProTect Feasibility Study, a novel comprehensive in vitro approach to detect reproductive
618 toxicants. Reprod Toxicol 30(1):200-218.

619 Schuurmans AL, Bolt J, Veldscholte J, Mulder E. 1991. Regulation of growth of LNCaP human
620 prostate tumor cells by growth factors and steroid hormones. J Steroid Biochem Mol Biol 40(1-
621 3):193-197.

622 Soto AM, Sonnenschein C. 2010. Environmental causes of cancer: endocrine disruptors as
623 carcinogens. Nat Rev Endocrinol 6(7):363-70.

624 Spencer TE, Dunlap KA, Filant J. 2011. Comparative developmental biology of the uterus:
625 Insights into mechanisms and developmental disruption. Mol Cell Endocrinol 354(1-2):34-53.

626 Stoker TE, Laws SC, Guidici DL, Cooper RL. 2000a. The effect of atrazine on puberty in male
627 wistar rats: an evaluation in the protocol for the assessment of pubertal development and thyroid
628 function. Toxicol Sci 58(1):50-59.

629 Stoker TE, Parks LG, Gray LE, Cooper RL. 2000b. Endocrine-disrupting chemicals: prepubertal
630 exposures and effects on sexual maturation and thyroid function in the male rat. A focus on the
631 EDSTAC recommendations. Endocrine Disrupter Screening and Testing Advisory Committee.
632 Crit Rev Toxicol 30(2):197-252.

633 Tilghman SL, Nierth-Simpson EN, Wallace R, Burow ME, McLachlan JA. 2010. Environmental
634 hormones: Multiple pathways for response may lead to multiple disease outcomes. Steroids
635 75(8-9):520-3.
636
637 U.S. EPA. 2007. Validation of the Fish Short-Term Reproduction Assay: Integrated Summary
638 Report. Available: http://www.epa.gov/endo/pubs/assayvalidation/status.htm [accessed March 3,
639 2011].

U.S. EPA. 2008. TOXCAST: Research Chemical Inventory for EPA's ToxCastTM Program: Structure-Index File Available: http://www.epa.gov/ncct/dsstox/sdf_toxcst.html [accessed 12 January 2012].

U.S. EPA. 2011a. Weight-of-Evidence: Evaluating Results of EDSP Tier 1 Screening to Identify the Need for Tier 2 Testing. Office of Chemical Safety and Pollution Prevention. Available: http://www.regulations.gov/#!documentDetail;D=EPA-HQ-OPPT-2010-0877-0021 [accessed 17 April 2012].

U.S. EPA. 2011b. Endocrine Disruptor Screening Program for the 21st Century: EDSP21 Work Plan. Office of Chemical Safety and Pollution Prevention. Available: http://www.epa.gov/endo/pubs/edsp21_work_plan_summary%20_overview_final.pdf [accessed 12 January 2012].

U.S. EPA. 2012a. Endocrine Disruptor Screening Program. Office of Chemical Safety and Pollution Prevention. Available: http:// www.epa.gov/endo/ [accessed 03 July 2012].

U.S. EPA. 2012b. Endocrine Disruptor Screening Program: Tier 1 Screening Battery. Office of Chemical Safety and Polution Prevention. Available: http:// www.epa.gov/endo/pubs/assayvalidation/tier1battery.htm [accessed 03 July 2012].

U.S. EPA. 2012c. ToxCast: Screening Chemicals to Predict Toxicity Faster and Better. National Center for Computational Toxicology. Available: http:// www. http://epa.gov/ncct/toxcast/ [accessed 03 July 2012].

U.S. EPA. 2012d. Computational Toxicology Research: Tox21. Office of Research and Development. Available: http://www.epa.gov/ncct/Tox21/ [accessed 03 July 2012].

U.S. EPA. 2012e. Endocrine Disruptor Screening Program. Office of Chemical Safety and Pollution Prevention. Available: http:// http://www.epa.gov/endo/pubs/assayvalidation/status.htm [accessed 03 July 2012].

Veldscholte J, Berrevoets CA, Ris-Stalpers C, Kuiper GG, Jenster G, Trapman J, et al. 1992.The androgen receptor in LNCaP cells contains a mutation in the ligand binding domain which affects steroid binding characteristics and response to antiandrogens. J Steroid Biochem Mol Biol 41(3-8):665-9.

Vinggaard AM, Breinholt V, Larsen JC. 1999. Screening of selected pesticides for oestrogen receptor activation in vitro. Food Addit Contam 16(12):533-542.

Xie W, Yeuh MF, Radominska-Pandya A, Saini SP, Negishi Y, Bottroff BS, et al. 2003. Control of steroid, heme, and carcinogen metabolism by nuclear pregnane X receptor and constitutive androstane receptor. Proc Natl Acad Sci U S A 100(7):4150-4155.

Zhou Q, Nie R, Prins GS, Saunders PT, Katzenellenbogen BS, Hess RA. 2002. Localization of androgen and estrogen receptors in adult male mouse reproductive tract. J Androl 23(6):870-881.

675 Zorrilla LM, Gibson EK, Jeffay SC, Crofton KM, Setzer WR, Cooper RL, et al. 2009. The
676 effects of triclosan on puberty and thyroid hormones in male Wistar rats. Toxicol Sci 107(1):56-
677 64.

Table 1

Summary of Endocrine Literature Survey

ToxCast Assay	Assigned MOA	Species	Assay Target	Assay Technology	No. of Unique Chemicals Tested	No. of Chemicals Overlapping with EDSP/OECD Reports	No. of overlapping active chemicals in ToxCast
ATG_AR_TRANS	HTS-A	Human	Androgen Receptor-Agonist	Multiplexed reporter gene assay	309[a]	13	0
NCGC_AR_Agonist	HTS-A	Human	Androgen Receptor-Agonist	GAL4 BLAM Reporter gene assay	309	13	0
NCGC_AR_Antagonist	HTS-A	Human	Androgen Receptor-Antagonist	GAL4 BLAM Reporter gene assay	309	13	5
NVS_NR_hAR	HTS-A	Human	Androgen Receptor	Competitive Binding	309	13	6
NVS_NR_rAR	HTS-A	Rat	Androgen Receptor	Competitive Binding	309	13	1
ATG_ERa_TRANS	HTS-E	Human	Estrogen Receptor-alpha	Multiplexed reporter gene assay	326[b]	21	12
ATG_ERE_CIS	HTS-E	Human	Estrogen Receptor Response Element	Multiplexed reporter gene assay	326[b]	21	11
ATG_ERRa_TRANS	HTS-E	Human	Estrogen Related Receptor-alpha	Multiplexed reporter gene assay	326[b]	21	0
ATG_ERRg_TRANS	HTS-E	Human	Estrogen Related Receptor-gamma	Multiplexed reporter gene assay	326[b]	21	0
NCGC_ERalpha_Agonist	HTS-E	Human	Estrogen Receptor-alpha-Agonist	GAL4 BLAM Reporter gene assay	326[b]	21	7
NCGC_ERalpha_Antagonist	HTS-E	Human	Estrogen Receptor-alpha-Antagonist	GAL4 BLAM Reporter gene assay	309	15	4
NVS_NR_bER	HTS-E	Bovine	Estrogen Receptor	Competitive Binding	316[b]	17	1
NVS_NR_hER	HTS-E	Human	Estrogen Receptor	Competitive Binding	326[b]	21	4
NVS_NR_mERa	HTS-E	Mouse	Estrogen Receptor-alpha	Competitive Binding	316[b]	17	1
NVS_ADME_hCYP19A1	HTS-S	Human	Aromatase	Enzyme Inhibition	309	17	1
NCGC_TRbeta_Agonist	HTS-T	Human	Thyroid Hormone Receptor-beta-Agonist	GAL4 BLAM Reporter gene assay	309	8	0
NCGC_TRbeta_Antagonist	HTS-T	Human	Thyroid Hormone Receptor-beta-Antagonist	GAL4 BLAM Reporter gene assay	309	8	0
NVS_NR_hTRa	HTS-T	Human	Thyroid Hormone Receptor-alpha-Antagonist	Receptor Activation	309	8	0

a Additional reference compounds from Judson et al. 2010 were run but not included because this is the only androgen-related HTS assay that tested these chemicals

b Includes additional reference compounds from Judson et al. 2010

678

Table 2

Summary of Endocrine Literature Survey

Endocrine Modes of Action	No. of Documents*	No. of Data Points*	No. of Unique Chemicals from Literature Survey*	No. Chemicals Overlapping with ToxCast for Comparison*
Estrogenecity	18 (108)	410 (979)	104 (158)	21 (143)
Androgenecity	22 (54)	571 (301)	60 (73)	13 (59)
Steroidogenesis	10 (32)	123 (251)	44 (61)	17 (55)
Thyroid	7 (48)	142 (190)	27 (57)	8 (47)
ALL	40 (215)	1246 (1721)	154 (182)	35 (157)

* Guideline (Non-Guideline)

Figures Legends:

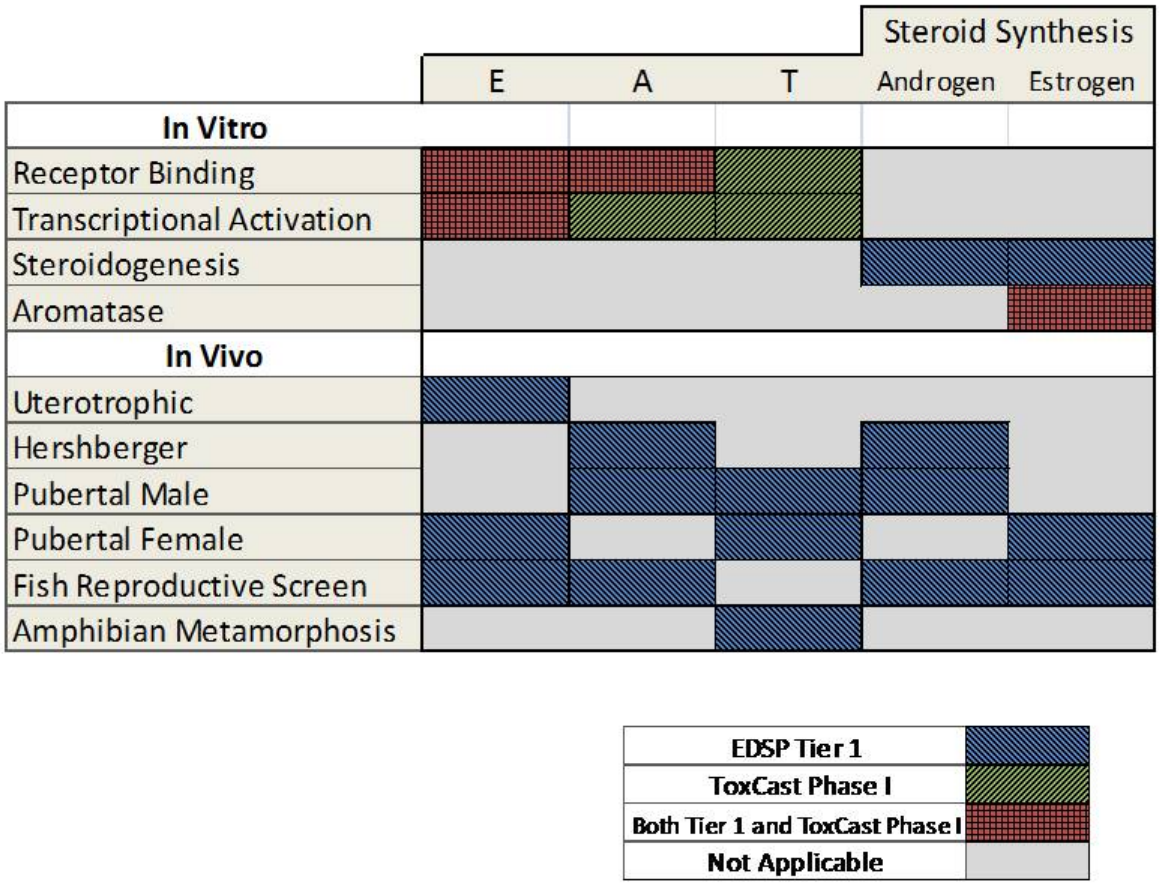

Figure 1. Overlap between EDSP Tier 1 assays and ToxCast Phase I assays by endocrine modes of action (MOA). Colors illustrate the type of endocrine MOA data (top) that was captured from the various study types (side).

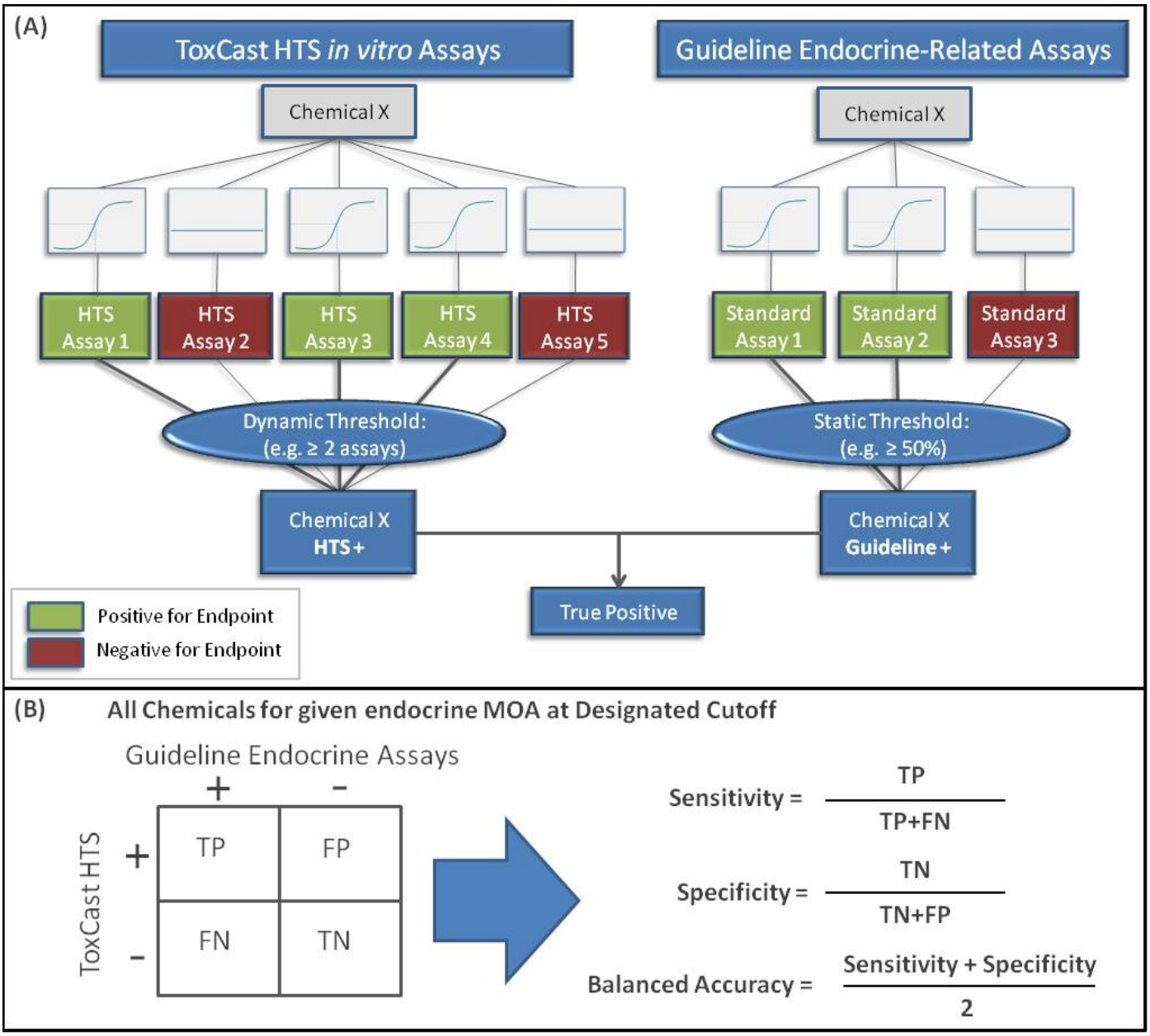

Figure 2. Graphical representation of the balanced optimization model used to analyze predictive capacity of endocrine related ToxCast assays. Multiple assays and study reports were available for each chemical-mode of action (MOA) combination. To determine whether a chemical was classified as active or inactive for ToxCast High-throughput screening (HTS) assays, a variable threshold was iteratively optimized to range from any positive assay to all positive assays required for a chemical to be classified as positive. Two scenarios were used to classify chemicals with data from EDSP/OECD guideline reports with either any single positive report or >50% of reports required for a chemical to be classified as positive. All chemicals were then tabulated in a two-by-two contingency table to calculate sensitivity, specificity, and balanced accuracy were used as a measure of model performance. Panel A provides a snapshot of a step in this modeling/optimization process. Chemical X is positive in 3 of 5 HTS assays and 2 of 3 guideline reports. In this example, the dynamic HTS threshold is "at least 2 positive assays" and the guideline threshold is "at

least 50% positive reports", so Chemical X is considered a true positive (TP). Panel B shows how results for all chemicals are tabulated (e.g. Chemical X would be counted in the true positive portion of the contingency table) to arrive at an estimate of balanced accuracy for each set of threshold parameters.

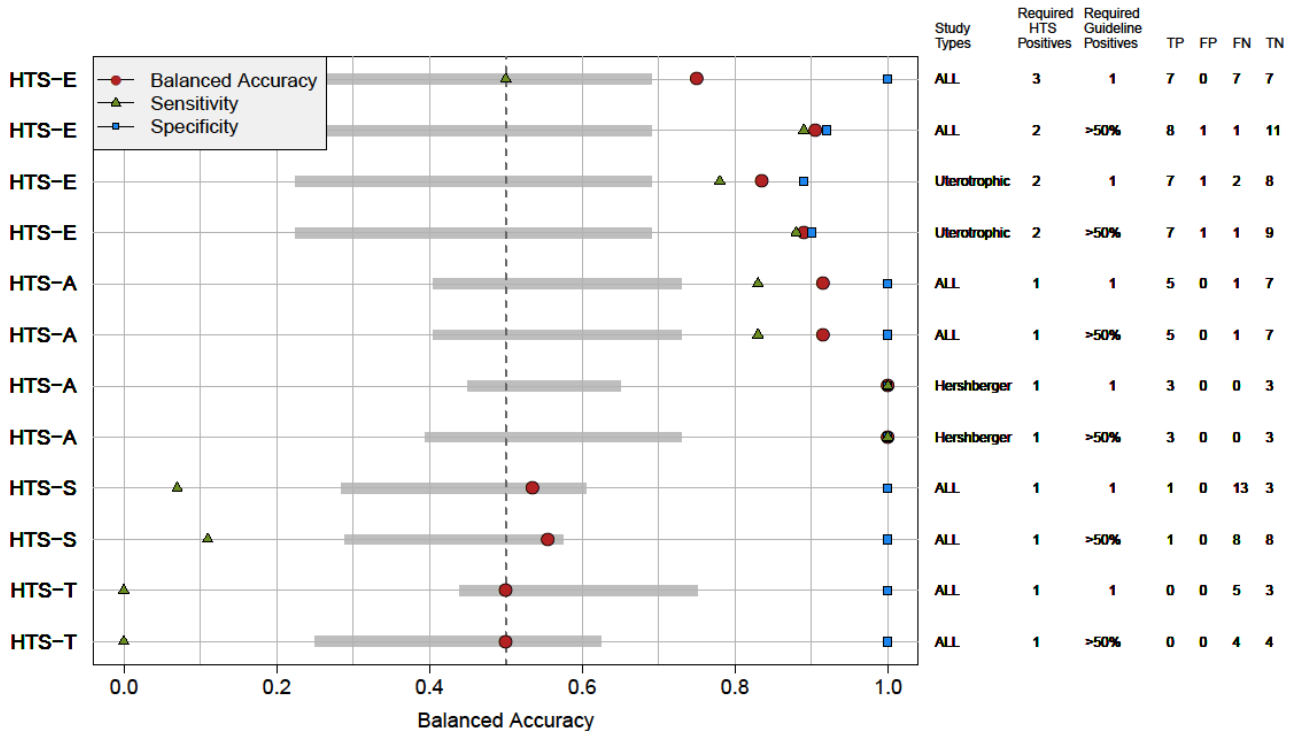

Figure 3. A forest plot illustrating the performance, as measured by sensitivity, specificity, and balanced accuracy (BA), of ToxCast endocrine related assays for predicting outcomes captured in EDSP/OECD guideline studies is graphically represented for comparison. Circles represent the optimal BA obtained across all threshold combinations with the corresponding sensitivity and specificity at the same threshold. Gray boxes indicate 95% confidence intervals around permuted BA distributions. Analyses designated "ALL" include all available assays for the stated endocrine mode of action. If the "Required Guideline Positives" column is >50%, then greater than 50% of the studies had to report a positive result for a chemical to be considered a positive in the analysis. If the "Required Guideline Positives" column is designated "1" then any study reporting a positive resulted in the chemical being considered positive in the analysis. A separate analysis comparing only uterotrophic and Hershberger analyses has also been included. The number of chemicals classified as true positive (TP), false positive (FP), true negative (TN), and false negative (FN) are tabulated along the side.

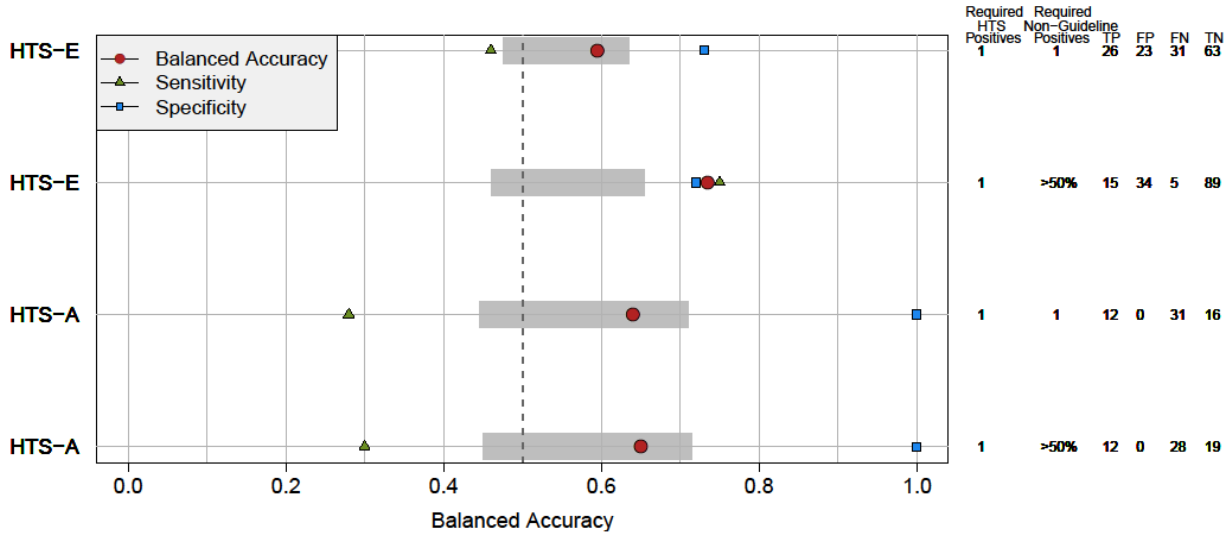

Figure 4. A forest plot illustrating the performance, as measured by sensitivity, specificity, and balanced accuracy (BA), of ToxCast endocrine related assays for predicting outcomes captured in non-guideline endocrine studies is graphically represented for comparison. Circles represent the optimal BA obtained across all threshold combinations with the corresponding sensitivity and specificity at the same threshold. Gray boxes indicate 95% confidence intervals around permuted BA distributions. If the "Required Non-Guideline Positives" column is >50%, then greater than 50% of the studies had to report a positive result for a chemical to be considered a positive in the analysis. If the "Required Non-Guideline Positives" column is designated "1", then any study reporting a positive resulted in the chemical being considered positive in the analysis. The number of chemicals classified as true positive (TP), false positive (FP), true negative (TN), and false negative (FN) are also tabulated.