



EPA Public Access

Author manuscript

Reprod Toxicol. Author manuscript; available in PMC 2020 April 21.

About author manuscripts

Submit a manuscript

Published in final edited form as:

Reprod Toxicol. 2018 October ; 81: 272–280. doi:10.1016/j.reprotox.2018.08.017.

EVALUATION OF ANDROGEN ASSAY RESULTS USING A CURATED HERSHBERGER DATABASE

NC Kleinstreuer^a, P Browne^b, X Chang^c, R Judson^d, W Casey^a, P Ceger^c, C Deisenroth^d, N Baker^e, K Markey^f, RS Thomas^d

^aNIH/NIEHS/DNTP/NICEATM

^bOECD/EHS Paris, France

^cILS, RTP, NC, USA, RTP, NC, USA

^dEPA/ORD/NCCT, RTP, NC, USA

^eLeidos, RTP, NC, USA

^fEPA/OSCP, Washington, DC, USA

Abstract

A set of 39 reference chemicals with reproducible androgen pathway effects in vivo, identified in the companion manuscript [1], were used to interrogate the performance of the ToxCast/Tox 21 androgen receptor (AR) model based on 11 high throughput assays. Cytotoxicity data and specificity confirmation assays were used to distinguish assay loss-of-function from true antagonistic signaling suppression. Overall agreement was 66% (19/29), with ten additional inconclusive chemicals. Most discrepancies were explained using in vitro to in vivo extrapolation to estimate equivalent administered doses. The AR model had 100% positive predictive value for the in vivo response, i.e. there were no false positives, and chemicals with conclusive AR model results (agonist or antagonist) were consistently positive in vivo. Considering the lack of reproducibility of the in vivo Hershberger assay, the in vitro AR model may better predict specific AR interaction and can rapidly and cost-effectively screen thousands of chemicals without using animals.

Keywords

androgen receptor; in vitro; high-throughput screening; ToxCast/Tox21; Hershberger; androgenic; anti-androgenic; endocrine disruptor

The opinions expressed and arguments employed herein are those of the authors, and do not necessarily reflect the official views of the OECD or of the governments of its member countries, nor the official policy of any federal agency.

Disclaimer. The authors declare no competing financial interests. This manuscript and the view expressed herein are those of the authors and do not necessarily reflect the views or policies of the US EPA, NIH, or OECD.

I. Introduction

Screening environmental chemicals for endocrine disrupting effects is of high importance to ensure protection of human health and ecological systems. Several global initiatives aim to implement rapid, cost-effective approaches to detect potential chemical interactions with key steroid hormone pathways, i.e. estrogen, androgen, and thyroid. One such example is the US Environmental Protection Agency's (EPA) Endocrine Disruptor Screening Program (EDSP), which was established in 1998 and in 2015 began accepting data generated from high throughput screening (HTS) assays and computational tools in lieu of guideline methods [2] (US EPA 2015a). The EDSP initially implemented a two-tiered test battery system, where Tier 1 contains 11 in vitro and in vivo assays of varying complexity, designed to screen for potential estrogen, androgen and thyroid hormone activity [3] (US EPA 2015b). Tier 1 consists of five in vitro assays, covering estrogen receptor (ER) binding, ER transcriptional activation, androgen receptor (AR) binding, steroidogenesis, and aromatase activity, as well as six in vivo assays, namely the rodent uterotrophic, Hershberger, and male and female pubertal studies, and the frog amphibian metamorphosis and fish short-term reproduction tests. Tier 2 in vivo studies are designed to test for adverse apical effects and characterize a dose-response relationship, but do not provide substantive mechanistic insight. The first set of 52 "List 1" EDSP chemicals went through Tier 1 testing from 2009–2015, and the weight of evidence for certain chemicals indicated the need for additional Tier 2, or targeted, studies. Given the approximately 10,000 chemicals in the EDSP universe [4] (US EPA 2012), the EPA recognized the need to implement novel approaches that would provide equivalent insight to the existing Tier 1 assays, but would allow for screening large numbers of chemicals in a fraction of the time.

Since the publication of the National Academies of Sciences report, *Toxicity Testing in the 21st Century* [5] (NRC 2007) and the establishment of the federal Tox21/ToxCast research programs, significant technological and computational advancements have resulted in effective tools for screening and prioritization of chemicals, some of which are acceptable replacements for animal-based test methods (e.g. [3,6–8] US EPA 2015b; Browne et al. 2015; Ezendam et al. 2016; Kleinstreuer et al. 2018). The US EPA developed a high throughput model of ER bioactivity that integrated results from 18 high throughput ER assays that were part of the Tox21/ToxCast programs and covered multiple key events in the ER pathway including receptor-binding, dimerization, chromatin binding, transcriptional activation, and ER-dependent cell proliferation [9] (Judson et al. 2015). A set of in vivo reference chemicals was derived by Kleinstreuer and colleagues [10] (2016a), who compiled and curated a large database of rodent uterotrophic studies, the Tier 1 in vivo screen for estrogenic activity. These studies were evaluated for their adherence to internationally accepted test guideline protocols, and chemicals with reproducible results in high-quality "guideline-like" studies were considered in vivo ER pathway reference chemicals. The accuracy and sensitivity of the ER model predictions for 65 in vitro and in vivo reference chemicals was greater than 85% [6,9] (Browne et al. 2015; Judson et al. 2015). Following a series of scientific advisory panel meetings and peer-reviewed publications, in 2015 the EPA decided to adopt the ToxCast ER model as an alternative for three current Tier 1 assays,

namely the ER binding in vitro assay, the ER transcriptional activation in vitro assay (ERTA), and the in vivo uterotrophic assay [3,6] (Browne et al. 2015; US EPA 2015b).

More recently, an AR pathway model was similarly constructed, integrating 11 high throughput assays from the Tox21/ToxCast programs that measure various molecular and cellular events (e.g. receptor-binding, coregulator recruitment, chromatin-binding of the mature transcription factor, and gene transcription) involved in AR agonism and/or antagonism [11] (Kleinstreuer et al. 2016b). For both the ER and AR models, a substantial strength of the high throughput approach is inclusion of assays that use different technologies and measure chemical interactions at different points in the signaling pathway [9,11] (Judson et al. 2015; Kleinstreuer et al. 2016b). Thus, the integrative high throughput models are able to distinguish between “true” agonists or antagonists and chemicals that may interfere with these various assay technologies through false signals. The model outputs were combined with additional information from a range of cell stress and cytotoxicity assays to provide predictions, and confidence scores, for true AR pathway activity for >1800 chemicals. The AR model demonstrated excellent performance against agonist (95.2% accuracy) and antagonist (97.5% accuracy) in vitro reference chemicals derived from a systematic review of the in vitro literature [11] (Kleinstreuer et al. 2016b).

The current EDSP Tier 1 screening assays for AR pathway activity are the AR in vitro binding assay and the in vivo rodent Hershberger, a short-term bioassay for identifying (anti)androgenic effects of chemicals on accessory sex tissue (AST) weights [12,13] (US EPA 2009a, OECD 2009). The database described in the companion paper [1] (Browne et al. in prep) includes 602 Hershberger studies on 216 unique chemicals, and details on test chemical, experimental design, and experimental results were recorded and assessed using a priori criteria to ensure quality of results. Reference chemicals were identified based on in vivo effects on the androgen pathway that were reproducible in more than one Hershberger study, or in a Hershberger study and another in vivo study design with androgen responsive endpoints. These reference chemicals were intended for use in performance-based validation of new methods and predictive models, such as the ToxCast AR model. Each of these reference chemicals have associated doses at which in vivo androgen pathway activity was observed (or not). To facilitate comparison between in vitro and in vivo systems, in vitro activity concentrations can be translated into predicted in vivo exposures using toxicokinetic modeling and in vitro to in vivo extrapolation (IVIVE) [14–16] (Wetmore et al. 2015, Pearce et al. 2017, Chang et al. 2018).

In this manuscript, we compared the ToxCast AR model performance against a set of in vivo reference chemicals identified from a database of chemicals tested in the Hershberger assay and other in vivo guideline tests measuring androgen-responsive endpoints [1] (Browne et al. in prep). We extrapolated from the in vitro activity concentrations to predicted in vivo doses to provide an equivalent basis for comparison, and discussed the results in the context of the biology covered by the AR model and the Hershberger study, respectively.

2. Methods

2.1. Reference chemical identification

A comprehensive literature search was conducted as described in [1] Browne et al. (in prep) to identify Hershberger studies for ~3200 discrete chemicals in the EDSP universe. Briefly, the full text of each publication was reviewed for relevance and when appropriate, study details were extracted into a database. Details of each chemical/experiment/publication combination (i.e. “study”) were recorded separately and information such as chemical name, purity, source, Chemical Abstract Service Reference Number (CASRN); rat strain, age, intact/castrated, post-surgical recovery duration, number of animals per treatment group; dosing duration, route of test chemical administration, dose levels; positive and negative controls for androgenic and anti-androgenic portions of the assays; ASTs measured; and significant effects were recorded [1] (Browne et al. in prep). Results from other endocrine assays designed to screen for (anti)androgenic effects in male rodents were identified for the chemicals with Hershberger studies, and included for comparison to Hershberger data. Results from the EDSP Tier 1 male pubertal rat screening assay [17] (US EPA 2009b), results of other in vivo male rat study designs with androgen-responsive endpoints (e.g. 28 day repeated dose oral toxicity, 90 day repeated dose oral toxicity, or extended one generation studies in rats) that were accepted by the US EPA in lieu of a male pubertal, or the results of another in vivo study reported in the same publication as the Hershberger study and intended to support results observed in the Hershberger study results were included in the database. The results of male pubertal assays (or other experimental protocols as defined above) were very briefly summarized, and details such as the animal model or test guideline, age of animals, dosing route, NOEL, LOEL, effect, and comments were recorded (Supplemental Table 2). The Hershberger assay is intended to respond to chemicals that alter androgen signaling in male rodent ASTs, but may also detect chemicals that interfere with activity of the enzyme 5 α -reductase responsible for the conversion of testosterone to the more potent androgen dihydrotestosterone. An independent literature search was also conducted to identify a reference list for reported 5 α -reductase inhibitors and the resulting list of compounds was cross-referenced to the Hershberger database [1] (Browne et al. in prep; Supplemental Table 5).

Based on the Hershberger database and associated in vivo data presented in [1] Browne et al. (in prep), we derived a list of chemicals with reproducible effects on the androgen pathway in vivo, referred to here as “in vivo reference chemicals”. These reference chemicals were those with confirmed androgen pathway effects in more than one study according to the following criteria:

1. Positive:
 - 2 Hershberger positive **or**
 - 1 Hershberger + 1 other in vivo (both positive) **and** greater number of positive than negative in vivo studies
2. Negative:
 - 2 negative Hershberger (with no positives) **or**

- 1 negative Hershberger + 1 other negative in vivo **and** no positive in vivo studies

Further details of the literature search terms, literature curation, criteria for initial study inclusion, criteria for “reproducible results”, and identification of unique chemicals are contained in [1] Browne et al. (in prep).

2.2. ToxCast AR model

A computational model of AR pathway activity was developed based on HTS assays from the ToxCast/Tox21 program that measure test chemical interaction at multiple points in the AR signaling pathway using a variety of assay technologies [11] (Kleinstreuer et al. 2016b). The AR model integrates data from 11 AR assays including receptor-binding, coregulator recruitment, chromatin-binding of the mature transcription factor, and gene transcription. The model predicts both agonist and antagonist activities by combining information from the assays relevant to each activity into a synthetic concentration-response curve, and providing area-under-the-curve (AUC) scores for agonism, antagonism, and assay interference pathways. For each chemical, AUC scores are normalized to yield a value of 1 for the reference antagonist, hydroxyflutamide, and across ~1800 chemicals tested in ToxCast the AUC scores range from 0 (no activity) to about 1.6 [11] (Kleinstreuer et al. 2016b). A model AUC of 0.1 corresponds to AR pathway activity below 100 μ M, and model scores below 0.001 were truncated. The AR model also yields pathway-level “pseudo” (derived from the synthetic concentration-response curve) activity concentrations, such as AC50s (at half-maximal response) or ACCs (at statistically significant cutoff), for predicted agonist and/or antagonist activities.

Some chemicals interfere with various assay technologies (e.g. via auto-fluorescence or cytostatic mechanisms; [18–21] Inglese et al. 2007, Thorne et al. 2010, Bruns and Watson 2012, Hsieh et al. 2015), and a considerable advantage of the AR model is the reliance on multiple assay technologies that allows discrimination of false positives due to technology-specific interference from “true” signaling [11] (Kleinstreuer et al. 2016b). Further, most environmental chemicals that interact with the AR are expected to do so in an antagonist (or anti-androgenic) fashion, which may be confounded by cytotoxicity in loss-of-function assays. To confirm AR-mediated activity, the AR antagonist transactivation assay (the Tox21 antagonist luciferase assay in the MDAKB2 cell line) was run independently with two different concentrations of the activating ligand (R1881, at the EC₅₀ of 0.5 nM and 20 times the EC₅₀). The higher concentration of the ligand should shift the potency of true competitive antagonists to higher concentrations, or quench the signal of weak antagonists. The data were flagged if a chemical was active in both runs at similar concentrations or if a potency shift was observed in the opposite direction than would be expected. The confirmation data were considered alongside cytostatic activity profiling in 33 viability and proliferation inhibition assays [22] (Judson et al. 2016), and then combined with the results of the AR model to provide an overall confidence score in the prediction of a chemical as a potential AR antagonist; further details can be found in [11] (Kleinstreuer et al. 2016b). For antagonist activity, the following decision logic was applied:

1. Positive:

- AR model AUC ≥ 0.1 **and** confidence score ≥ 1
2. Negative:
 - AR model AUC = 0 **and** confidence score ≥ 1
 3. Inconclusive:
 - AR model AUC ≥ 0.1 **and** confidence score < 1
 - $0 < \text{AR model AUC} < 0.1$, regardless of confidence score
 - AR model AUC = 0 **and** confidence score > 1

In cases where the AUC ≥ 0.1 , and the confidence score ≥ 1 (e.g. because the confirmation assay data demonstrated a shift in the wrong direction), chemicals are potential false positives in vitro and are deemed inconclusive. Similarly, chemicals with a non-zero AUC < 0.1 and confidence score > 1 were also considered inconclusive, as they could represent AR model false negatives. These categorizations are consistent with the logic applied to develop a training set for an international QSAR modeling effort to predict interaction of untested environmental chemicals with the AR pathway [23] (Mansouri et al. 2018).

2.3. In vitro to in vivo extrapolation

The outputs from the AR model were compared to the results of Hershberger and other studies described in [1] Browne et al. (in prep). IVIVE was used to facilitate comparison between the results of the AR pathway model based on in vitro HTS assays and chemicals identified in the Hershberger database with consistent (anti)androgenic effects in vivo. The equivalent administered dose (EAD) in vivo was extrapolated from the in vitro activity concentrations at statistically significant cutoff (ACCs) for chemicals that were active in any of the 11 AR assays and the pseudo-ACC for chemicals that were predicted active in the AR model (equivalent to an in vitro lowest effect level), or from the highest concentrations tested for in vitro inactive chemicals. This IVIVE approach has been described previously [14–16] (Wetmore et al. 2015, Pearce et al. 2017, Chang et al. 2018). Briefly, experimentally derived in vitro measures or structure-based predictions of hepatic clearance and plasma protein binding were generated for each chemical and used to parameterize one-compartment population-based toxicokinetic rat models built with a custom R script (v. 3.1.2) [16] (Chang et al. 2018). In the model, Monte Carlo simulations were performed to simulate variability across a population of 10,000 individuals, resulting in a distribution of plasma concentrations at steady state (C_{ss}) at a daily dose of 1 mg/kg/day. The 95th percentile of this distribution was then used as a conversion factor to calculate the daily EAD, assuming 100% absorption, that would yield steady-state plasma concentrations equivalent to ACC values from the ToxCast AR assays or the pseudo-ACC for pathway-level activity from the AR model. These estimated doses were compared to LOELs from Hershberger studies with significant (androgenic or anti-androgenic) effects and the NOELs for Hershberger studies with no significant effects.

3. Results

3.1. ToxCast AR model comparison

From the systematic literature review of ~3,200 chemicals, we identified 602 chemical/protocol/study combinations for 216 unique chemicals with Hershberger assay data, and after a full-text review of the study design and data extraction, 373 chemical/study combinations for 134 chemicals remained [1] (Browne et al. in prep). We identified 73 unique chemicals with either 2 Hershberger studies (25 chemicals) or a Hershberger study and another in vivo study with androgen-responsive endpoints (65 chemicals, including 17 with 2 Hershberger studies). Out of these 73 chemicals, 49 chemicals satisfied our proposed criteria for consistent outcomes (positives: 2 in vivo studies with androgenic/anti-androgenic results; negatives: 2 in vivo studies with no effects observed in any in vivo study) and were considered to have reproducible in vivo effects on the androgen pathway, where 24 chemicals did not meet these reproducibility criteria (Figure 1 in [1]; Browne et al. in prep). The 39 chemicals with reproducible in vivo effects that were also tested in ToxCast and had resulting AR model scores (including AUC and confidence scores) were used in our comparison; 10 chemicals were not in the ToxCast library. These 39 “in vivo reference chemicals” include three androgenic chemicals, 20 anti-androgenic chemicals, and 16 negative chemicals (Table 1).

Based on similar analyses of the ER model, an AR model AUC of 0.1 is considered a potential positive (for agonist or antagonist activity) and corresponds to in vitro activity of test chemical concentrations < 100 μ M, model AUC of 0 are considered negative, and AUC scores $0 < \text{AUC} < 0.1$ are inconclusive [11] (Kleinstreuer et al. 2016b). However, the model AUC for AR antagonist activity are further contextualized by confidence scores, which incorporate the confirmation antagonist ARTA data and the information on generalized cell stress and cytotoxicity across multiple cell types [11] (Kleinstreuer et al. 2016b). Confidence scores ranged from 0 to 6 (Table 1), where higher confidence scores indicate greater weight-of-evidence for antagonist activity [11] (Kleinstreuer et al. 2016b). Chemicals with true AR-mediated activity result in a left shift in the concentration-response curves of antagonist confirmation assays (compared to the same assay run with a higher concentration of activating agonist) and chemicals without a shifted response in the confirmation assay were flagged as potentially acting via cell stress or cytotoxicity, regardless of the model score. Similarly, chemicals that may have been incorrectly predicted by the model as acting via interference pathways, but exhibiting the expected shift in the confirmation assay were also flagged as potential antagonists. Chemicals with antagonist activity at concentrations below the cytotoxic region had higher confidence scores compared to those whose predicted antagonist activity overlapped with the cytotoxic concentration range [11] (Kleinstreuer et al. 2016b). Therefore, for antagonist activity, any chemical with a model AUC 0.1 and a confidence score >1 is considered a positive, any chemical with a model AUC of 0 and a confidence score 1 is considered a negative, and all others are considered inconclusive (see Methods for detailed criteria). The AR pathway model agonist/antagonist AUC, confirmation data flags, and confidence scores for the 39 chemicals with reproducible in vivo effects that were also in ToxCast are shown in Table 1. Of the 134 unique chemicals identified in the Hershberger literature review that met the inclusion criteria [1] (Browne et

al. in prep), 94 were run in the suite of ToxCast and Tox21 AR assays, and the AR pathway model outputs are included in Supplemental Table 2.

The three potent steroid androgens (17-methyl testosterone, testosterone propionate, and trenbolone) were reproducibly androgenic in vivo, and they were all strong positives in the AR agonist model (AUC scores >1.5; Table 1). The in vivo anti-androgenic reference chemicals ranged from strong pharmaceuticals (e.g. flutamide, finasteride) to weaker pesticide and industrial chemicals (e.g. benfluralin, dibutyl phthalate). Of the 20 reproducibly in vivo anti-androgenic chemicals also tested in ToxCast, six (30%) were positive, four (20%) were inconclusive, and the remaining 10 chemicals (50%) were negative in the AR pathway model (Table 1). Two of the inconclusive chemicals, finasteride and fenarimol, had confidence scores of 3 and 4 respectively but were predicted by the model as potential technology interference. One chemical, propargite, had a positive AUC>0.1 but was flagged due to an aberrant potency shift in the confirmation data, and the final inconclusive chemical from the positive in vivo reference set, befluralin, was negative in all but two of the cofactor recruitment assays. There were 16 reproducibly in vivo negative chemicals tested in ToxCast, and of these 10 (62.5%) were also negative in the AR pathway model; the remaining six chemicals were inconclusive (Table 1). Three of the inconclusive chemicals, abamectin, folpet and chlorothalonil, had positive AR pathway antagonist model AUC scores but were flagged as potential false positives based on results of the confirmation assay (exhibited a shift in the wrong direction). Another two chemicals, mgk-264 and tetrachlorvinfos, had inconclusive AR pathway antagonist model AUC scores between 0 and 0.01, but had confidence scores >1. Both of these chemicals exhibited the expected shift in concentration-response curves for the confirmation assays, but the shift was not statistically significant based on overlapping confidence intervals [11] (Kleinstreuer et al. 2016b). The final inconclusive chemical from the in vivo negative reference set, methomyl, had a model AUC of 0, but a confidence score of 3 due to activity in the Tox21 confirmation assay well below the region of cytotoxicity.

The overall concordance between the ToxCast AR pathway model calls and the in vivo reference chemical classifications was 66% (19/29), after excluding ten chemicals with inconclusive calls (Table 1). In general, agreement between chemicals with reproducible in vivo effects and AR model scores was similar to or better than agreement between the in vivo data itself (i.e. 72% reproducibility of the Hershberger or 57% agreement between results from >1 studies, Hershberger or other in vivo, on the same chemical [1] (Browne et al.; Supplemental Table 2).

3.2 Incorporation of dose

To further facilitate comparisons between in vitro and in vivo chemical activity, we used reverse toxicokinetics to extrapolate estimated administered doses (EADs) using the range of ACCs from of the 11 in vitro AR assays, the pseudo-ACCs from the AR pathway model, and the highest in vitro concentration tested (100 μ M) for negative results. The EADs were compared to LOELs and NOELs from Hershberger studies. For the three (reproducibly) androgenic chemicals run in ToxCast, the EADs were less than or similar to the androgenic in vivo LOELs (Figure 2).

For the 20 chemicals with consistent in vivo anti-androgenic effects, nine chemicals had AR antagonist AUC scores > 0 (i.e. seven with AUC > 0.1 and two with AUC between 0 and 0.1) and EADs less than or similar to anti-androgenic in vivo LOELs (Table 1, Figure 3, left panel). Of these nine chemicals, six (chemical names in bold in Figure 3) were considered true positives by the overall AR model call, incorporating both the AUC and the confidence scores. Five of the six true positives chemicals (except for prochloraz) had EAD ranges overlapping LOEL ranges. For prochloraz, the discrepancy between the EAD and the LOEL was similar to the range of doses covered by the LOELs from multiple studies. Of the remaining 11 in vivo anti-androgenic chemicals that had AR antagonist AUC scores of 0 (Table 1), four chemicals had some activity in at least one AR in vitro assay, and the corresponding EADs were below or overlapped the LOELs for three chemicals (DEHP, metalochlor, and pronamide), but were above the LOELs for finasteride (Figure 3, middle panel). Finasteride, a 5 α -reductase inhibitor, had in vivo LOELs that ranged over several orders of magnitude, but were still below the EADs from in vitro assays (Table 1, Figure 3, middle panel). In fact, finasteride was active in both of the Tox21 antagonist mode assays at concentrations below the cytotoxic threshold (confidence score = 3), but had a model score of 0 (predicted as pathway interference due to lack of upstream activity such as binding). The remaining seven in vivo anti-androgens were negative in all in vitro assays (Figure 3, right panel), and with the exception of ethoprop, the EADs calculated from the highest in vitro concentration tested (100 μ M) were below the median LOELs observed in vivo, indicating that the top testing concentration in the in vitro assays may not have been high enough to elicit an AR pathway response for these chemicals.

Looking at the entire Hershberger database from [1] Browne et al. (in prep), there were 94 chemicals with Hershberger results that were also run in ToxCast (including the 39 in vivo reference chemicals used in the analyses described above). Of the 94 total, 16 chemicals were negative across all in vitro assays and in vivo Hershberger studies. For the remaining 78 chemicals with indications of AR pathway activity in either in vitro or in vivo studies, the EADs calculated from a) in vitro activity in any AR assay or b) the highest in vitro concentration tested (for chemicals with no in vitro activity) are presented in Supplemental Tables 8 and 9, in addition to all pharmacokinetic parameters and summary data for comparison. Of the 78 chemicals, 53 had Hershberger NOELs (AR pathway model scores shown in Supplemental Figure 1A), and 35 of those chemicals had activity in at least one AR pathway assay. We compared the EADs derived from in vitro AR assays to in vivo NOELs for these chemicals, and for the majority of chemicals (33/35), the EADs were less than or similar to the in vivo NOELs (Supplemental Figure 1B).

4. Discussion

Alternatives to animal testing are becoming increasingly available and accepted by regulators for characterizing chemical hazards. Greater uptake of alternative methods has resulted from technological advances and coinciding changes to legislation reducing or prohibiting the use of animals in chemical safety determinations [24,25] (European Union 2003, US EPA 2016). As results from animal studies are compiled as benchmarks for comparison to alternative methods, there is growing recognition that guideline animal studies performed under relatively rigid experimental conditions nonetheless have a

unexpectedly high degree of uncertainty, with regard to both predicting their own results (i.e. repeatable results for chemicals tested in more than one study) and predicting effects in humans [1,8,10] (Browne et al. in prep, Kleinstreuer et al. 2018, Kleinstreuer et al. 2016a). Examples illustrating uncertainty in animal data rely on systematic review and curation of existing data, and as such, have been limited to assay study designs with adequate data to support meta-analyses and/or for very few circumstances for which both animal and human data exist for a toxicological endpoint. However, based on these undertakings, the high degree of uncertainty in animal study results may be the rule rather than the exception. Alternative methods are needed to increase the efficiency of chemical safety screening, the relevance to humans, and perhaps, the reliability and reproducibility of results.

Here, we compared the results from the ToxCast AR model, based on 11 in vitro assays, to “in vivo reference chemicals” including three androgenic chemicals, 20 anti-androgenic chemicals, and 16 negative chemicals. Three of three androgenic chemicals (100%) with reproducible in vivo effects had positive AR agonist model calls, and 100% of chemicals that were negative in vivo lacked positive AR model calls (i.e. either no AR model activity (AR AUC scores = 0; 10/16) or inconclusive results (6/16)). The agreement was less strong for antagonist activity; 30% (6/20) of the in vivo anti-androgens had positive AR model AUC scores and high confidence scores; 20% (4/20) were inconclusive and the remaining 50% (10/20) of chemicals had AR model AUC scores = 0. In the discordant cases, the lack of agreement can be better understood by considering the dose predicted from the in vitro concentrations. The anti-androgenic reference chemicals that were false negatives in all the in vitro assays had a median Hershberger LOEL of 200 mg/kg/day, with several LOEL values of 1000 mg/kg/day. In contrast, the median EAD estimated from the top in vitro concentration for those chemicals was 20 mg/kg/day, an order of magnitude lower than the doses at which in vivo responses were observed. In the case of dibutyl phthalate, for example, the Hershberger LOELs ranged from 500 – 1000 mg/kg/day, while the estimated administered dose extrapolated from the top in vitro concentration was 67 mg/kg/day, and similar results were observed for the other phthalates from the entire Hershberger database which were also tested in ToxCast (Supplemental Table 9). Phthalates are hypothesized to act through an unknown molecular initiating event other than the AR [26] (Howdeshell et al. 2017), and the monoester metabolite is usually the anti-androgen active component, which may provide further explanations beyond the obvious contributor of dose for the discordance between the in vitro and in vivo results. Another consideration in this comparison is the impact of internal dose and the limitations of the modeling approach. The one-compartment steady state IVIVE model used here does not account for tissue-partitioning of chemicals nor does it differentiate route of exposure, but assumes a fixed dose rate and 100% absorption. It is possible that for poorly-absorbed compounds, the internal bioavailable in vivo equivalent may be substantially lower than the external in vivo LOEL dose, and similarly the EAD would need to be higher to reach target tissue concentrations where activity is observed in vitro.

Chemicals that act primarily as signal inhibitors or antagonist can be challenging to measure using in vitro systems due to the confounding loss-of-function effects of cytotoxicity and assay interference. For this reason, it is critical to include measures of cytotoxicity and confirmation assays in the interpretation of results. After applying the decision criteria

including confidence scores (based on confirmation flags and cytotoxicity data) in these analyses, the ToxCast AR model had 100% positive predictive value ($PPV = TP/(TP + FP)$) when compared to the in vivo reference chemicals, i.e. there were no false positives. All chemicals that were predicted to be agonists (AUC agonist score >0.1 ; 3 chemicals) or antagonists (AUC antagonist score >0.1 and confidence score >1 ; 6 chemicals) were concordant with the reference classifications derived from reproducible in vivo results. All of the chemicals with AR model AUC scores of >0.1 with negative in vivo results (i.e. potential false positives) were classified as inconclusive based on the confirmation assay data, highlighting the importance of this approach in distinguishing true AR pathway activity mediated through the nuclear receptor. In part due to the analysis of the AR antagonist data set, inclusion of confirmation runs is now standard practice for Tox21 assays.

Based on the outcomes, we hypothesized three primary factors are contributing to the disagreement between in vitro AR model calls and in vivo (anti)androgenic effects. First, the in vitro AR pathway model may have a limited capacity to detect anti-androgens with low potency, as is reflected in the IVIVE analysis. To some extent, this may be complicated by the difficulty in discerning AR antagonist assay activity (i.e. loss-of-function) from cell stress and cytotoxicity. Though the confirmation assay data and associated confidence scores were included to help improve confidence in “true” antagonist activities, low potency AR antagonists may be active at concentrations that are within the same range as concentrations that cause other confounding factors in vitro, or at concentrations that are above the testing range used in the ToxCast assays.

Second, the Hershberger assay was designed to be maximally sensitive to anti-androgens [27] (OECD, 2008a), as the expectation is that most chemicals interacting with the androgen pathway would have a suppressive action. In the validation and optimization of several animal models, the young adult castrated rat, co-treated with a stimulating pharmaceutical androgen and the test chemical, was determined to be the most sensitive method to detect anti-androgen dependent decreases in the weights of accessory sex tissues [27] (OECD 2008a). A variety of chemicals may induce hepatic xenobiotic metabolizing enzymes, resulting in increased clearance of the stimulating androgen and subsequent decrease in accessory sex tissue weight of treated animals when compared with (androgen stimulated) controls in the anti-androgenic mode of the assay [28–30] (Freyberger et al. 2007, Freyberger and Schladt 2009, Marty et al. 2014). This effect could be interpreted as anti-androgenic but may not be reproduced in other studies that do not co-administer a stimulating androgen and may contribute to discordance between the Hershberger and other in vivo rodent studies measuring androgenic endpoints (discussed in detail in [1] Browne et al. in prep), and potentially to the discrepancies observed between the AR pathway model and the Hershberger.

Third, there is a possibility that biotransformation of parent compounds may contribute to the discrepancy between in vitro and in vivo responses. Absence of metabolic capacity is a well-recognized limitation of in vitro assays [31] (OECD 2008b). In contrast to the AR comparison, the ToxCast ER model was an excellent predictor of in vivo uterotrophic activity. The uterotrophic assay is a shorter, simpler assay, and is focused on identifying estrogenic (agonist) chemicals. Additionally, in silico analyses of parent-metabolite ER

agonist activity indicated that while a variety of parent chemicals were predicted to have more estrogenic metabolites, very few of these parent chemicals completely lacked estrogenic activity [32] (Pinto et al. 2016). Similarly, some parent chemicals that are active in vitro appear to have no effect in vivo and may be inactivated. A similar in silico analysis of parent-metabolite predictions will provide additional insight into the impact of metabolism on the predictive capacity of the in vitro AR model.

4.2 Potential use of AR model for endocrine screening

Comparison between Hershberger results and other in vivo rodent studies with androgen responsive endpoints indicate discordance between results of multiple Hershberger assays conducted on the same chemical and between Hershberger results and other in vivo study designs [1] (Browne et al. in prep). While there are a variety of physiological explanations for potential differences between the Hershberger assay and results of chemicals tested in male rodents with intact neuroendocrine axes, this physiology is not likely to entirely explain the discrepancies. In fact, the Hershberger assay is part of the US EPA's EDSP Tier 1 battery and a chemical's potential for activity in the androgen pathway is interpreted along with other mechanistic and in vivo data [33] (US EPA 2011). For the EDSP List 1/Tier 1 chemicals, a positive Hershberger result was not interpreted as a positive screen unless it was accompanied by another positive in vivo assay with androgen pathway endpoints [3] (US EPA 2015b). OECD's Guidance Document for evaluating chemicals for endocrine disruption recommends running the Hershberger assay to confirm effects in vivo following a positive in vitro AR assay, or to confirm an AR mode of action for a response observed in a neuroendocrine intact animal model [34] (OECD 2012). In a limited comparison of model chemicals, positive results in Hershberger assays were also positive in male pubertal assays, and the authors proposed the male pubertal assay is important for confirming effects observed in the Hershberger [35] (Ankley and Gray 2013). The Hershberger assay was intended to be an in vivo bioassay for androgen receptor mediated effects, and it was developed during a time when reliable in vitro AR assays were not available. The assay does not include the full complexity of hypothalamic-pituitary-gonadal axis interactions present in an intact male rodent, but it is certainly more than an in vivo assay that solely queries interaction with the AR pathway.

Given the potential difficulty interpreting results, the variability of the results, and the recommendation of confirming results of the Hershberger in another in vivo assay for regulatory purposes, it is unclear what information the Hershberger assay provides. Negative results in the Hershberger are insufficient to rule out AR-mediated effects, as organ weight changes are highly variable and are not sensitive indicators of AR interactions, weak anti-androgens may not be able to compete with the stimulating androgen, and co-administration of TP, alone, is enough to induce hepatic enzymes thereby potentially increasing metabolism and limiting the effect of the test chemical. Positive anti-androgenic results may be due to test chemical induction of hepatic enzymes and subsequent clearance of the stimulating androgen, or be difficult to interpret due to the use of HPG-interrupted animals and not reproducible in other in vivo assays [1,28–30] (Browne et al. in prep, Freyberger et al. 2007, Freyberger and Schladt 2009, Marty et al. 2014). Further, the new European Union criteria to identify endocrine disruption for plant protection products and biocides are likely to be

adopted in 2018 and require linking an adverse effect observed in an intact animal with an endocrine mode of action (https://ec.europa.eu/health/endocrine_disruptors/policy_en). Under these circumstances, the Hershberger assay would not be sufficient to demonstrate an adverse effect, and due to potential contributions from other toxicity pathways, it is not ideal for demonstrating an endocrine mode of action. A more appropriate approach to screen chemicals for androgenic effects may be an in vitro determination of AR interactions combined with a male pubertal or other in vivo assay with androgen responsive endpoints. The AR model discussed here could certainly be improved by incorporating metabolism in the vitro assays, as well as being coupled with in silico models to predict metabolic activation and inform test substance selection. Further model improvements could be achieved by adjusting the IVIVE approach to account for bioavailability, both in vitro and in vivo.

Our analyses suggest that many parent chemicals that have no effects in vitro have reproducibly anti-androgenic effects in vivo, either due to the biotransformation of the chemical or altered hepatic clearance. The metabolic capabilities of in vitro assays vary substantially [36] (Castell et al. 2006), but even in vitro test systems with limited innate metabolism can be run with addition of S9 fractions or other co-cultures that improve metabolic capabilities, e.g. [31,37] (OECD 2008b, van Vugt-Lussenburg et al. 2018). Several of the Tox21 partners (EPA, NCATS, and NIEHS/NTP) launched the “Transform Toxicity Testing Challenge” in 2016 to inspire scientists to retrofit existing HTS assays to incorporate metabolism, and the project announced the Stage Two winners in late 2017 (<https://www.epa.gov/innovation/announcing-transform-toxicity-testing-challenge-stage-two-winners>). From this analysis, 11 reproducibly in vivo anti-androgenic chemicals have AR pathway antagonist model AUC scores = 0 (Table 1), though this includes finasteride which exerts an anti-androgenic effect through 5 α -reductase inhibition rather than AR interaction, and was deemed inconclusive based on the overall AR model call. The primary in vivo anti-androgenic effect of finasteride is mediated by 5 α -reductase activity, so it might not be expected to have a positive in vitro AR model score, but it is interesting that it exhibited activity in two of the in vitro AR antagonist assays. Finasteride and dutasteride have been shown to exhibit anti-androgen effects by blocking steroid hormone uptake in prostate cancer lines, and the effect observed in the Tox21 assay is likely due to androgen depletion via transporter inhibition in MDA-kb2 cells [38] (Wu et al. 2013). It is possible that other in vivo anti-androgenic chemicals may act through 5 α -reductase inhibition, and while our literature search to identify 5 α -reductase inhibitors returned relatively few chemicals [1] (Browne in prep; Supplemental Table 5), this is no doubt confounded by the lack of a reliable assay to detect this effect. Further, for the 12 chemicals that were identified in the literature search, 11 of 12 were detected by the AR model, though many of those chemicals (e.g. prochloraz, progesterone, estrone, 17 β -estradiol) have recognized activities at concentrations lower than either AR receptor or potential interaction with 5 α -reductase (Supplemental Table 7). The AR model may have a limited ability to detect weakly anti-androgenic effects and one way to increase the sensitivity of the in vitro model would be to consider all AR antagonist model AUC scores > 0 as potentially positive. In this case, it would be especially important to consider AR antagonist model AUC scores in the context of confirmation flags to reduce potential false positives. While some chemicals that are

positive in the Hershberger assay may be acting through in vivo mechanisms other than the AR (e.g. finasteride), it is likely that some combination of testing higher doses and accounting for potential metabolic activation of chemicals would improve the agreement between in vivo and in vitro results.

5. Conclusions

Compared to the ToxCast ER model performance against the uterotrophic bioassay of estrogenic chemicals (>85% accuracy; [6] Browne et al. 2015), the ToxCast AR model was a poor predictor of the Hershberger outcome (66% accuracy). One possible reason for disagreement between the Hershberger assays and the in vitro model is the relatively longer exposure duration (10 days in the Hershberger assay versus 3 in the uterotrophic) allowing for greater induction of hepatic xenobiotic enzyme activity. In addition, the Hershberger assay is a considerably more complicated assay, including five tissue weights (versus one in the uterotrophic assay) and accounting for both androgenic and anti-androgenic actions. There is a great deal of interest in adding metabolic capabilities to in vitro assays to improve the simulation of test chemical fate in organisms [31] (OECD 2008b), and this would likely improve the predictive capacity and confidence of in vitro replacements for animal tests.

The level of disagreement within the Hershberger itself [1] (Browne et al. in prep) indicates the substantial noise in the assay. Prior to the widespread availability of reliable in vitro assays for the AR pathway, the Hershberger provided valuable information on many chemicals. However, rather than working to improve in vitro AR model agreement with results from a bioassay that may be difficult to interpret and does not strongly influence overall weight of evidence [3] (US EPA 2015b), efforts might be better invested in developing reliable in vitro co-culture techniques that simulate hepatic metabolism, provide reproducible data on the mechanism of action, and may supplant a need for lower tier in vivo studies. Though our preliminary literature investigation identified very few chemicals that inhibited 5 α -reductase activity, this is difficult to assert in the absence of a reliable in vitro assay. The Hershberger assay is one of the few methods that can potentially identify chemicals that interact with 5 α -reductase, though subtle differential changes in specific AST weights may be difficult to discern. Development of a reliable in vitro or in silico method for this mode of action would clearly identify chemicals that inhibit 5 α -reductase and may further reduce the need for expensive, time consuming, and animal-based Hershberger studies.

In vivo methods are not practical approaches for screening thousands of chemicals for potential endocrine activity. Faster, efficient, and accurate methods are needed to prioritize chemicals for additional endocrine screening and testing. Such methods may also contribute to strategic testing that can bypass the need for specific tests. Our analyses highlighted the need for considerations of cytotoxicity and signal interference when evaluating potential antagonists or other signal inhibitors. In part because measures of cytotoxicity, cell stress, and appropriate shifts in concentration-response curves were included, the AR model had 100% positive predictive value, i.e. there were no false positives. To further improve the efficiency of chemical screening, the AR model could be refined to improve the sensitivity for detecting low potency chemicals that interact with the androgen pathway (primarily as

anti-androgens) and chemicals that are bioactivated/inactivated in vivo. AR model results for chemicals considered with results of the male pubertal assay or another in vivo with androgen responsive endpoints assay conducted in an intact animal would help to anchor adverse effect to an endocrine mode of action. In circumstances where knowledge of the specific endocrine mode of action is required, the AR pathway model based on 11 in vitro assays may in fact provide a more accurate measure of chemicals that interact exclusively with the AR than the in vivo Hershberger bioassay, which may be confounded by complex biological interactions and chemicals acting through other toxicity pathways.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The authors gratefully acknowledge the assistance of the ICCVAM reference chemical working group and L. Earl Gray for assistance in identifying criteria for high quality animal studies, Charlene Matten and Karen Hammernik of the EPA Office of Science Coordination and Policy for discussions on the literature curation, and the technical reviewers: Mike DeVito (NTP) and Ron Hines (EPA) for constructive feedback.

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

LITERATURE CITED

1. Browne P, Kleinstreuer NC, Ceger P, Deisenroth C, Baker N, Markey K, Casey W. in prep. Development of a Curated Hershberger Database.
2. US EPA. 2015a Use of High Throughput Assays and Computational Tools: Endocrine Disruptor Screening Program; Notice of Availability and Opportunity for Comment, 80 Fed. Reg. 118 (June 19, 2015). Office of Chemical Safety and Pollution Prevention, U.S. Environmental Protection Agency, Washington DC Available online at: <https://www.federalregister.gov/articles/2015/06/19/2015-15182/use-of-high-throughput-assays-and-computational-tools-endocrine-disruptor-screening-program-notice>
3. US EPA. 2015b Endocrine Disruptor Screening Program Tier 1 Assessments. Office of Chemical Safety and Pollution Prevention, U.S. Environmental Protection Agency, Washington DC Available online at: <https://www.epa.gov/ingredients-used-pesticide-products/endocrine-disruptor-screening-program-tier-1-assessments> [Accessed 05/02/16]
4. US EPA. 2012 Universe of Chemicals and General Validation Principles, U.S. Environmental Protection Agency, Endocrine Disruptor Screening Program, Washington DC Available online at: http://www.epa.gov/sites/production/files/2015-07/documents/edsp_chemical_universe_and_general_validations_white_paper_11_12.pdf [Accessed 03/08/16].
5. NRC. 2007 Toxicity testing in the 21st century: A vision and a strategy. National Academies Press: Washington, DC; Vol. 25, p 136–138.
6. Browne P, Judson RS, Casey WM, Kleinstreuer NC, Thomas RS. 2015 Screening Chemicals for Estrogen Receptor Bioactivity Using a Computational Model. *Environ Sci Technol* 49(14): 8804–8814 [PubMed: 26066997]
7. Ezendam J, Braakhuis HM, Vandbriel RJ. 2016 State of the art in non-animal approaches for skin sensitization testing: from individual test methods towards testing strategies. *Arch Toxicol* 2016 9 14.
8. Kleinstreuer N, Hoffmann S, Alepee N, et al. 2018 Non-Animal Methods to Predict Skin Sensitization (II): an assessment of defined approaches. *Crit Rev Toxicol*; 48(5):359–374 [PubMed: 29474122]

9. Judson RS, Magpantay FM, Chickarmane V, Haskell C, Tania N, Taylor J, Xia M, Huang R, Rotroff DM, Filer DL, Houck KA, Martin MT, Sipes N, Richard AM, Mansouri K, Setzer RW, Knudsen TB, Crofton KM, Thomas RS. 2015 Integrated Model of Chemical Perturbations of a Biological Pathway Using 18 In Vitro High Throughput Screening Assays for the Estrogen Receptor. *Toxicological Sciences*; 148(1):137–54 [PubMed: 26272952]
10. Kleinstreuer NC, Ceger PC, Allen DG, Strickland J, Chang X, Hamm JT, et al. 2016a A Curated Database of Rodent Uterotrophic Bioactivity. *Environ Health Persp* 124(5): 556–562.
11. Kleinstreuer NC, Ceger P, Watt ED, Martin M, Houck K, Browne P, Thomas RS, Casey WM, Dix DJ, Allen D, Sakamuru S, Xia M, Huang R, Judson R. 2016b Development and validation of a computational model for androgen receptor activity. *Chem Res Toxicol* 2017; 30(4):946–964 (Epub ahead of print Dec 9 2016) [PubMed: 27933809]
12. US EPA. 2009a Endocrine Disruptor Screening Program Test Guidelines. OPPTS. 890.1400: Hershberger Bioassay. Available at: <https://www.regulations.gov/document?D=EPA-HQ-OPPT-2009-0576-0008>.
13. OECD 2009 OECD Guidelines for Testing of Chemicals, Section 4. Test No. 441: Hershberger bioassay in rats. A short-term screening assay for (anti)androgenic properties. 10.1787/9789264076334-en
14. Wetmore BA, Wambaugh JF, Allen B, Ferguson SS, Sochaski MA, Setzer RW, Houck KA, Strobe CL, Cantwell K, Judson RS, LeCluyse E, Clewell HJ, Thomas RS, Andersen ME. 2015 Incorporating High-Throughput Exposure Predictions With Dosimetry-Adjusted In Vitro Bioactivity to Inform Chemical Toxicity Testing. *Toxicol Sci*. 148:121–36. [PubMed: 26251325]
15. Pearce R, Setzer R, Strobe C, Sipes N, & Wambaugh J 2017 htk: R Package for High-Throughput Toxicokinetics. *Journal of Statistical Software*, 79(4), 1–26. [PubMed: 30220889]
16. Chang X, Casey WM, Allen DG, Ceger PC, Choksi NY, Hsieh JH, Wetmore BA, Ferguson SS, DeVito MJ, Sprankle CS, Kleinstreuer NC. 2018 In vitro to In vivo Extrapolation for Estrogenic Activity of Environmental Chemicals. *Env Health Persp* In press.
17. US EPA. 2009b U.S. EPA. Series 890-Endocrine Disruptor Screening Test OPPTS 890.1500: pubertal development and thyroid function in intact juvenile/peripubertal male rats. Available at: <http://www.regulations.gov/#!documentDetail;D=EPA-HQ-OPPT-2009-0576-0010>. Accessed January 29, 2016.
18. Inglese J, Johnson RL, Simeonov A, Xia M, Zheng W, Austin CP, Auld DS. 2007 High-throughput screening assays for the identification of chemical probes. *Nat Chem Biol* 3, 466–479. [PubMed: 17637779]
19. Thorne N, Auld DS, Inglese J. 2010 Apparent activity in high-throughput screening: origins of compound-dependent assay interference. *Curr Opin Chem Biol* 14, 315–324. [PubMed: 20417149]
20. Bruns RF, Watson IA. 2012 Rules for identifying potentially reactive or promiscuous compounds. *J Med Chem* 55, 9763–9772. [PubMed: 23061697]
21. Hsieh JH, Sedykh A, Huang R, Xi M, Tice RR. 2015 A Data Analysis Pipeline Accounting for Artifacts in Tox21 Quantitative High-Throughput Screening Assays. *J Biomol Screen*. 20, 887–897. [PubMed: 25904095]
22. Judson R, Houck K, Martin M, Richard AM, Knudsen TB, Shah I, Little S, Wambaugh J, Sezer RW, Kothiya P, Phuong J, Filer D, Smith D, Reif D, Rotroff D, Kleinstreuer N, Snipes N, Xia M, Huang R, Crofton K, Thomas RS. 2016 Analysis of the Effects of Cell Stress and Cytotoxicity on In Vitro Assay Activity Across a Diverse Chemical and Assay Space. *Toxicol. Sci*. 153: 323–339. doi: 10.1093/toxsci/kfw148
23. Mansouri K, Kleinstreuer NC, Grulke C, Richard A, Shah I, Williams AJ, Judson RS. 2018 Virtual Screening of Chemicals for Endocrine Disrupting Activity through CERAPP and CoMPARA Projects. DOI: 10.13140/RG.2.2.19612.80009
24. European Union. 2003 Directive 2003/15/EC of the European Parliament and of the Council of 27 February 2003 amending Council Directive 76/768/EEC on the approximation of the laws of the Member States relating to cosmetic products. *Official Journal of the European Union*.66:26–35.
25. US EPA. 2016 Frank R. Lautenberg Chemical Safety for the 21st Century Act updates to the Toxic Substance Control Act. <http://uscode.house.gov/view.xhtml?path=/prelim@title15/chapter53&edition=prelim>

26. Howdeshell KL, Hotchkiss AK, Gray LE Jr. 2017 Cumulative effects of antiandrogenic chemical mixtures and their relevance to human health risk assessment. *Int J Hyg Environ Health*. 220:179–188 [PubMed: 27923611]
27. OECD 2008a Background Review Document on the Rodent Hershberger Bioassay. Number 90. Series on Testing and Assessment. ENV/JM/MONO(2008)17. [http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono\(2008\)17&doclanguage=en](http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2008)17&doclanguage=en). [Accessed 10 august 2017]
28. Freyberger A, Ellinger-Ziegelbauer H, Krotlinger F. 2007 Evaluation of the rodent Hershberger bioassay: testing coded chemicals and supplementary molecular-biological and biochemical investigations. *Toxicology*. 239:77–88. [PubMed: 17688994]
29. Freyberger A, Schladt L. 2009 Evaluation of the rodent Hershberger bioassay on intact juvenile males-testing of coded chemicals and supplementary biochemical investigations. *Toxicology*. 262:114–120. [PubMed: 19467291]
30. Marty MS, O'Connor JC. 2014 Key learnings from the Endocrine Disruptor Screening Program (EDSP) Tier 1 rodent uterotrophic and Hershberger assays. *Birth Defects Res B Dev Reprod Toxicol*; 101(1):63–79 [PubMed: 24515841]
31. OECD. 2008b Detailed Review Paper on the Use of Metabolising Systems for In Vitro Testing of Endocrine Disruptors. Series on Testing and Assessment. Number 97. [http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono\(2008\)24&doclanguage=en](http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2008)24&doclanguage=en) [Accessed on 11 August, 2017]
32. Pinto CL, Mansouri K, Judson R, Browne P. 2016 Prediction of estrogenic bioactivity of environmental chemical metabolites. *Chem Res Toxicol*. 29:1410–1427. [PubMed: 27509301]
33. US EPA. 2011 Weight-of-Evidence: Evaluating results of EDSP Tier 1 screening to identify the need for Tier 2 testing. U. S. Environmental Protection Agency (2011). Available online at: <http://www.regulations.gov/#!documentDetail;D=EPA-HQ-OPPT-2010-0877-0021> [Accessed 03/08/17].
34. OECD 2012 Guidance Document on Standardised Test Guidelines for Evaluating Chemicals for Endocrine Disruption. Series on Testing and Assessment, No. 150. <http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono%282012%2922&doclanguage=en>. [Accessed on 10 August, 2017].
35. Ankley GT, Gray LE. 2013 Cross-species conservation of endocrine pathways: a critical analysis of tier 1 fish and rat screening assays with 12 model chemicals. *Environ Tox Chem*. 32:1084–1087.
36. Castell JV, Jover R, Martinez-Jimenez CP, Gmez-Lechn MJ. 2006 Hepatocyte cell lines: their use, scope and limitations in drug metabolism studies, *Expert Opinion on Drug Metabolism & Toxicology*, 2:2, 183–212 [PubMed: 16866607]
37. van Vugt-Lussenburg BMA, van der Lee RB, Man HY, Middelhof I, Brouwer A, Besselink H, van der Burg B. 2018 Incorporation of metabolic enzymes to improve predictivity of reporter gene assay results for estrogenic and anti-androgenic activity. 1;75:40–48. doi: 10.1016/j.reprotox.2017.11.005. Epub 2017 Nov 21.
38. Wu Y, Godoy A, Azzouni F, Wilton JH, Ip C, Mohler JL. 2013 Prostate cancer cells differ in testosterone accumulation, dihydrotestosterone conversion, and androgen receptor signaling response to steroid 5 α -reductase inhibitors. *The Prostate*. 73:13 10.1002/pros.22694

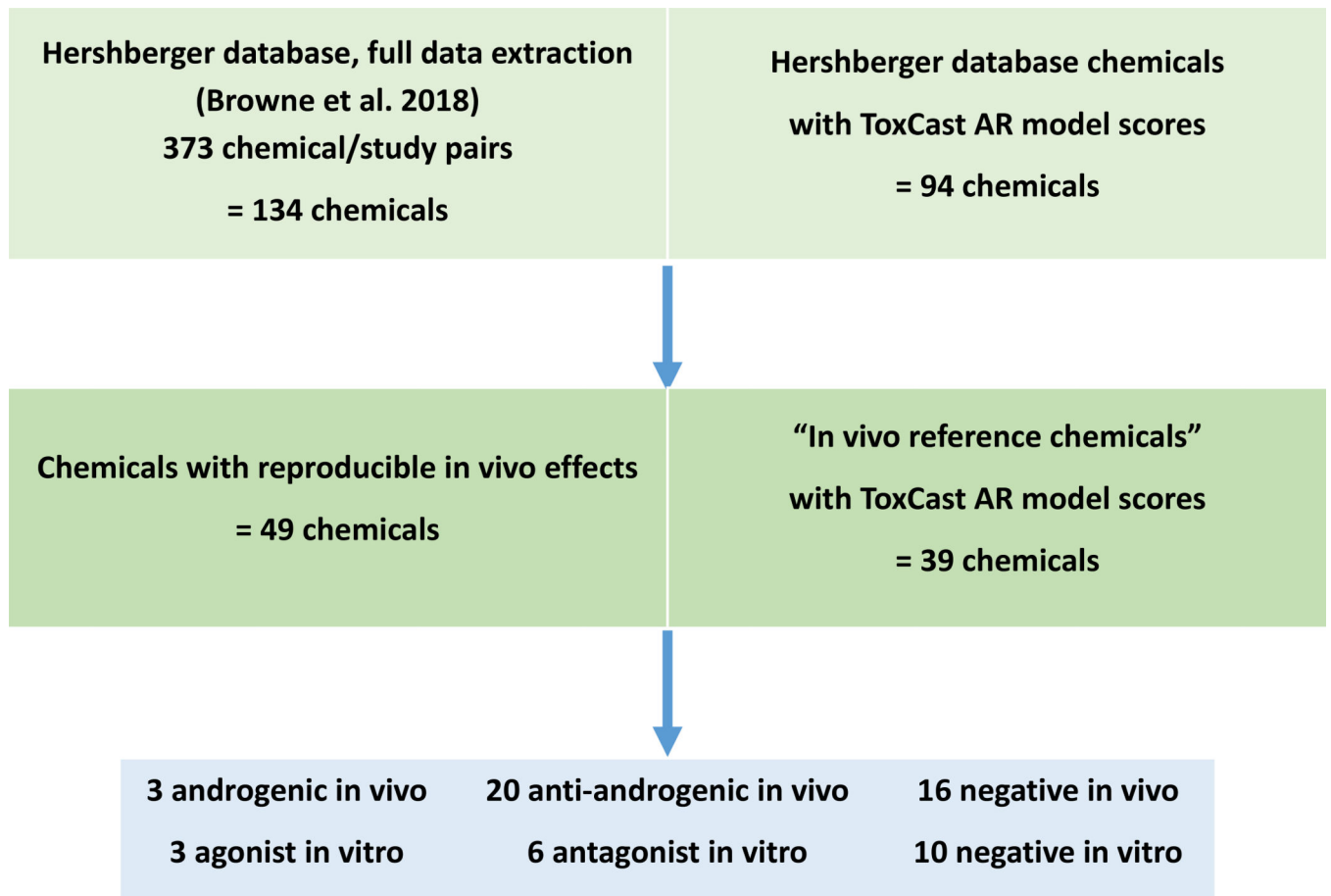


Figure 1.

Workflow shows numbers of chemicals included in the Hershberger database after meeting inclusion criteria (Browne et al. in prep), chemicals with reproducible in vivo effects that were considered “in vivo reference chemicals”, and the final comparison with ToxCast AR model calls. At each step, chemicals that also had data from 11 high throughput ToxCast/Tox 21 AR in vitro assays are indicated.

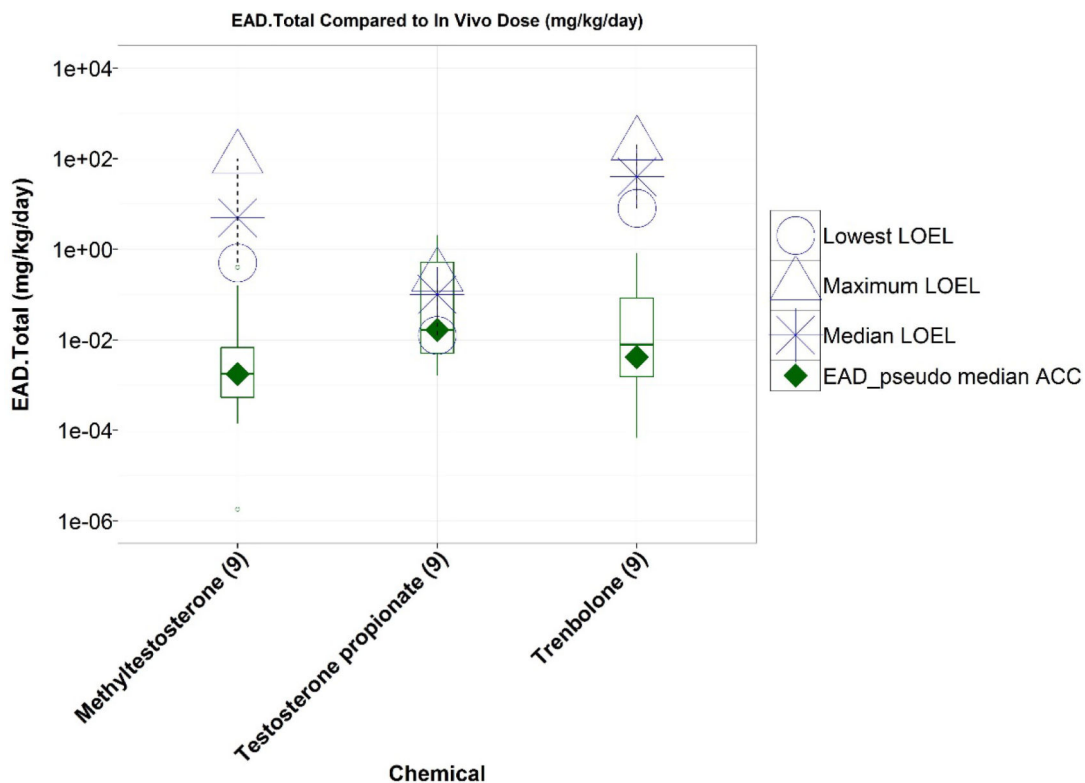


Figure 2. Comparison of lowest observed effect levels (LOELs, open symbols) from Hershberger agonist studies to chemical-specific equivalent administered doses (EADs) derived from nine ToxCast/Tox21 in vitro AR agonist pathway assays. Boxplots represent distribution of EADs from in vitro assay activity concentrations at statistically significant cutoffs, and the EAD from the AR pathway agonist model pseudo-ACC is represented by a green diamond. All three chemicals were active in all nine agonist-mode assays (numbers in brackets).

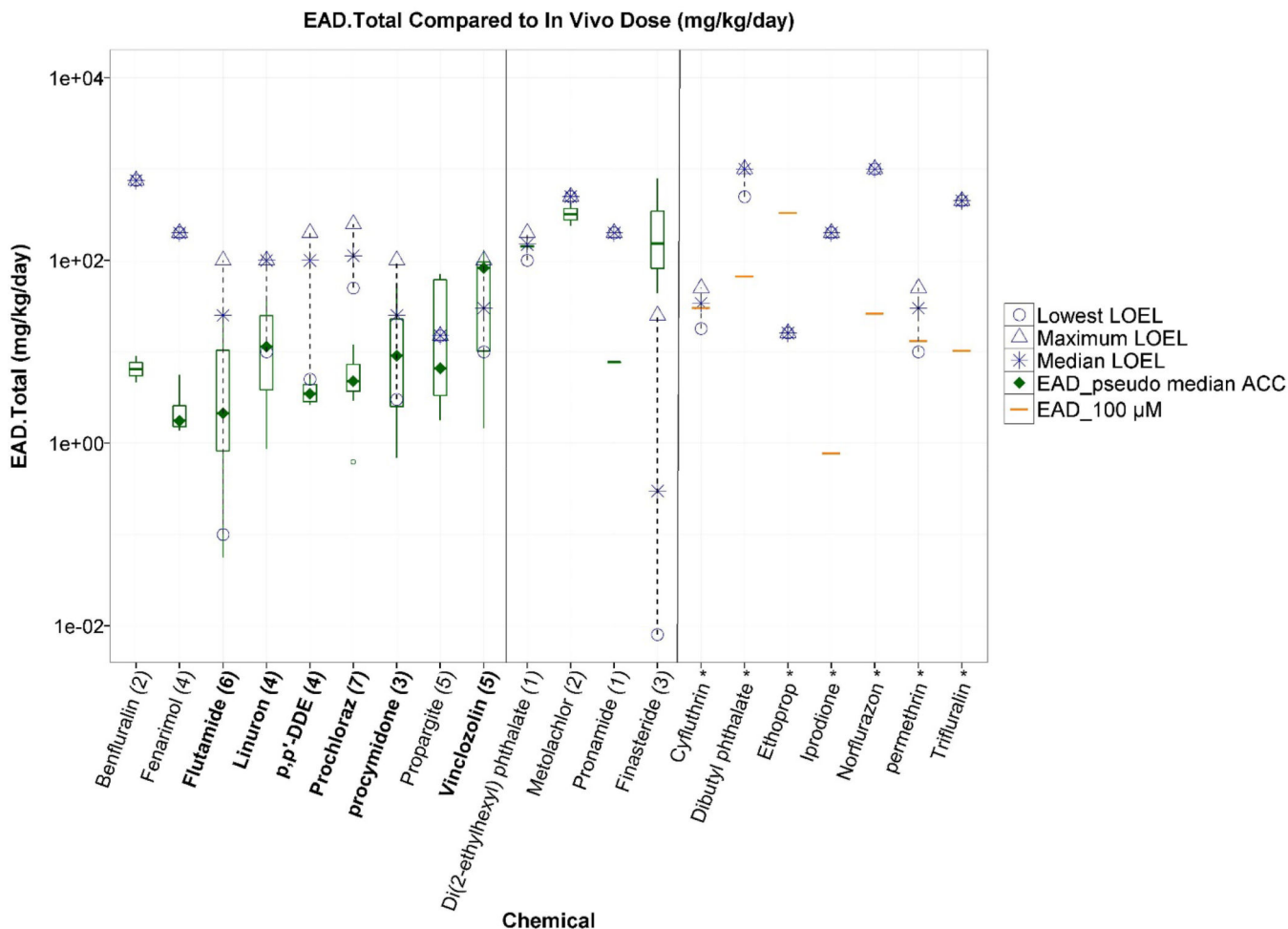


Figure 3. Comparison of lowest observed effect levels (LOELs, open symbols) from Hershberger antagonist studies to chemical-specific equivalent administered doses (EADs) derived from six ToxCast/Tox21 in vitro AR antagonist pathway assays. Boxplots represent distribution of EADs from in vitro assay activity concentrations at statistically significant cutoffs, the EAD from the AR pathway antagonist model pseudo-ACC is represented by a green diamond, and the EAD from the highest in vitro concentration tested (100 μ M) is represented by an orange line. The left panel contains chemicals with an AR model AUC >0, the middle panel has those with a model AUC of 0 that were active in at least one in vitro assay, and the right panel contains chemicals that were inactive in vitro. Numbers in brackets represent the number of antagonist-mode assays a chemical was active in, where chemicals that were inactive across all assays are indicated with an asterisk. Chemicals in bold are those that had an overall AR pathway model call of positive for antagonism.

Table 1.

Summary of comparison between 39 chemicals with reproducible in vivo effects (green=androgenic, red=anti-androgenic, grey=negative) and results of the ToxCast AR pathway model (green=agonist, red=antagonist, yellow = inconclusive, grey=negative). Abbreviations: p,p'-DDE = p,p'-dichlorodiphenyldichloroethylene, AUC = area under the curve (AR pathway model score), CI = confidence intervals, NA = no activity (in antagonist confirmation assays), LOEL = lowest observed effect level, NOEL = no observed effect level, EAD = equivalent administered dose, ACC = activity concentration at cutoff, HCT = highest concentration tested.

Chemicals	CAS	In Vivo Reference Classification	AUC Agonist	AUC Antagonist	Antagonist Confidence Score	Confirmation Assay Flag	Overall AR Model C	Concordance with In Vivo	HB LOEL range or NOEL (mg/kg/day)	EAD from range of In Vitro ACC or HCT (mg/kg/day)
testosterone propionate	57-85-2	Androgenic	1.53	0	1	NA	Agonist	1	0.0125 – 0.25	1.6E-03 – 2
17-methyl testosterone	58-18-4	Androgenic	1.55	0	1	NA	Agonist	1	0.5 – 100	1.8E-06 – 0.4
trenbolone	10161-33-8	Androgenic	1.59	0	1	NA	Agonist	1	8 – 200	1.4E-07 – 0.8
flutamide	13311-84-7	Anti-androgenic	0	0.547	5	True antagonist shift (Hit/Hit)	Antagonist	1	0.1 – 100	0.06 – 26.1
linuron	330-55-2	Anti-androgenic	0	0.3	5	True antagonist shift (No hit/Hit)	Antagonist	1	10 – 100	0.9 – 41.3
vinclozolin	50471-44-8	Anti-androgenic	0	0.416	6	True antagonist shift (Hit/Hit)	Antagonist	1	10 – 100	1.5 – 129.8
procymidone	32809-16-8	Anti-androgenic	0	0.316	6	True antagonist shift (Hit/Hit)	Antagonist	1	3 – 100	0.7 – 57.2
p,p'-DDE	72-55-9	Anti-androgenic	0	0.251	4	True antagonist shift (No hit/Hit)	Antagonist	1	5 – 200	2.6 – 5
prochloraz	67747-09-5	Anti-androgenic	0	0.295	5	True antagonist shift (Hit/Hit)	Antagonist	1	50 – 250	0.6 – 12
finasteride	98319-26-7	Anti-androgenic	0	0	3	True antagonist shift (No hit/Hit)	Inconclusive	NA	0.008 – 25	43.6 – 787.9
propargite	2312-35-8	Anti-androgenic	0	0.187	1	FLAG: Wrong direction shift (Hit/Hit)	Inconclusive	NA	15 – 15	1.8 – 71
fenarimol	60168-88-9	Anti-androgenic	0	0.0446	4	True antagonist shift (Hit/Hit)	Inconclusive	NA	200 – 200	1.4 – 5.6
benfluralin	1861-40-1	Anti-androgenic	0	0.00012	0	NA	Inconclusive	NA	750 – 750	4.6 – 9
permethrin	52645-53-1	Anti-androgenic	0	0	0	NA	Negative	0	10 – 50	13.2
bis(2-ethylhexyl) phthalate (DEHP)	117-81-7	Anti-androgenic	0	0	0	NA	Negative	0	100 – 200	142.7 – 142.7
noflurazone	27314-13-2	Anti-androgenic	0	0	0	NA	Negative	0	1000 – 1000	26.4
ethoprop	13194-48-4	Anti-androgenic	0	0	0	NA	Negative	0	16 – 16	333.3
cyfluthrin	68359-37-5	Anti-androgenic	0	0	0	NA	Negative	0	18 – 50	30.4
iprodione	36734-19-7	Anti-androgenic	0	0	0	NA	Negative	0	200 – 200	0.8
pronamide	23950-58-5	Anti-androgenic	0	0	0	NA	Negative	0	200 – 200	7.7 – 7.7
trifluralin	1582-09-8	Anti-androgenic	0	0	0	NA	Negative	0	450 – 450	10.3
dibutyl phthalate (DBP)	84-74-2	Anti-androgenic	0	0	0	NA	Negative	0	500 – 1000	67.4
metolachlor	51218-45-2	Anti-androgenic	0	0	0	NA	Negative	0	500 – 500	238.6 – 427
carbofuran	1563-66-2	Negative	0	0	0	NA	Negative	1	0.3	441
oxamyl	23135-22-0	Negative	0	0	0	NA	Negative	1	0.5	410.1
esfenvalerate	66230-04-4	Negative	0	0	0	NA	Negative	1	9	26.2
2,4-dinitrophenol	51-28-5	Negative	0	0	1	FLAG: Antagonist shift, but CI overlap	Negative	1	10	1.2 – 1.3
chlorpyrifos	2921-88-2	Negative	0	0	0	NA	Negative	1	12	10.4 – 12.1
metribuzin	21087-64-9	Negative	0	0	0	NA	Negative	1	120	418.5
metalaxyl	57837-19-1	Negative	0	0	1	NA	Negative	1	375	0.05 – 0.05
3-amino-1,2,4-triazole	61-82-5	Negative	0	0	0	NA	Negative	1	1000	641.9
flutolanil	66332-96-5	Negative	0	0	0	NA	Negative	1	1000	21.8 – 21.8
pentachloronitrobenzene (PCNB)	82-68-8	Negative	0	0	1	FLAG: Antagonist shift, but CI overlap	Negative	1	1000	0.3 – 4.1
methomyl	16752-77-5	Negative	0	0	3	True antagonist shift (No hit/Hit)	Inconclusive	NA	1	4.9E-04 – 0.7
abamectin	71751-41-2	Negative	0	0.388	1	FLAG: Wrong direction shift (Hit/Hit)	Inconclusive	NA	5	2.4 – 18.7
tetrachloroinfos	22248-79-9	Negative	0	0.0315	3	FLAG: Antagonist shift, but CI overlap	Inconclusive	NA	350	17.9 – 70.2
folpet	133-07-3	Negative	0	0.141	1	FLAG: Wrong direction shift (Hit/Hit)	Inconclusive	NA	800	7.8 – 65.2
mgk-264	113-48-4	Negative	0	0.0935	2	FLAG: Antagonist shift, but CI overlap	Inconclusive	NA	850	2 – 7.6
chlorothalonil	1897-45-6	Negative	0	0.481	1	FLAG: Wrong direction shift (Hit/Hit)	Inconclusive	NA	1000	3.4 – 311.9