Continuing Development of Alternative High-Throughput Screens to Determine Endocrine Disruption, Focusing on Androgen Receptor, Steroidogenesis, and Thyroid Pathways

U.S. Environmental Protection Agency
Endocrine Disruptor Screening Program

Jointly developed by:
U.S. EPA Office of Chemical Safety and Pollution Prevention (OCSPP)
U.S. EPA Office of Research and Development (ORD)
U.S. EPA Office of Water (OW)

NIH National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM)

FIFRA SAP, November 28-30, 2017

Table of Contents

List of Figures

List of Tables

Preface

The U.S. Environmental Protection Agency (EPA) is continuing a series of scientific peer reviews focused on evaluation and validation of high-throughput (HT) and computational approaches for prioritization and screening of chemicals in the Endocrine Disruptor Screening Program (EDSP). The Agency's white paper discusses three topics: an androgen receptor (AR) computation model, a steroidogenesis pathway model, and a proposed thyroid framework.

The AR pathway model is the updated approach (building on material presented at the December 2014 FIFRA SAP meeting (U.S. EPA, 2014a) for determining androgen bioactivity based on a computational model integrating data from 11 HT screening assays. The methods used for this approach are described and results presented. The Agency's work to address the issues mentioned in the December 2014 FIFRA SAP is clearly described. The validation of the computational model is discussed, and the scientific support to accept the HT androgen receptor (AR) model as an alternative for the low-throughput (LT) AR assay currently used in the Tier 1 screening battery is presented. The purpose of this chapter is to illustrate that the previous issues in 2014 have now been resolved, making this model an acceptable alternative to the LT assay. The HT AR model is considered useful by the Agency for prioritizing chemicals for further screening and testing, and for contributing to weight of evidence conclusions for Tier 1 evaluations of a chemical's potential endocrine bioactivity.

This document also describes the development of a HT H295R steroidogenesis model and novel statistical analysis approach for this model. Two variations in the analysis of the HT H295R assay results are being presented for the SAP's consideration. The first variation focuses only on changes in estrogen and testosterone concentrations following treatment with a series of reference chemicals. Data analysis is performed in a similar manner to the current LT model in order to compare the performance of the HT and LT H295R assays. The second variation is considered an improvement by the Agency to identify substances with the potential to disrupt steroidogenesis. The second variation uses a novel statistical approach to integrate the measurements of 9 additional steroid hormones from the HT H295R assay. The integrated statistical measure quantifies the overall impact of the substance on the steroidogenesis pathway. Methods and results are detailed for each variation of the analysis, and the potential advantages of the pathway model approach measuring 11 hormones in the steroidogenic pathway is described. The validation efforts are presented.

Finally, a description is provided of the Agency's initial work in establishing a framework utilizing a network of adverse outcome pathways (AOP) for the evaluation of

perturbation of thyroid function. The conceptual framework will be used by the Agency to inform future prioritization and screening of substances for further testing. The Agency seeks advice on the framework refinement and future development of assays and AOPs to detect disruption of thyroid hormone homeostasis.

EPA is requesting the FIFRA SAP to comment on charge questions related to these three topics. The Agency will bring additional HT assays and computational tools to the Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA) Scientific Advisory Panel (SAP) for peer review as they become available.

The Agency's goals for this SAP include:

- Confirmation that the suggested improvements in the AR pathway model (from the December 2014 FIFRA SAP) have been incorporated, allowing the HT AR pathway model to serve as alternative for the LT AR assay (OSCPP 890.1150) in the form of other scientifically relevant information.
- Confirmation that the newly developed HT H295R steroidogenesis assay (either the version measuring only estrogen and testosterone, the version measuring 11 hormones, or both) can serve as an alternative to the LT H295R assay (OSCPP 890.1550/ OECD TG 456) in the form of other scientifically relevant information.
- Comments and advice concerning the Agency's ongoing development of an approach to detect substances that can perturb thyroid function.

List of Acronyms and Abbreviations

AC50	Concentration required to elicit a 50% response in an *in vitro* assay
AMA	Amphibian Metamorphosis Assay
AOP	Adverse Outcome Pathway
AR	Androgen Receptor
AUC	Area Under the Curve
BW	Body Weight
CASRN	Chemical Abstracts Service Registry Number (chemical identification number)
DIO	Iodothyronine Deiodinase
DSSTox	Distributed Structure-Searchable Toxicity Database
DUOX	Dual Oxidase
EDSP	Endocrine Disrupter Screening Program
EDSTAC	Endocrine Disruptors Screening and Testing Advisory Committee
EOGRT	Extended One Generation Reproduction Test
EPA	United States Environmental Protection Agency
ER	Estrogen Receptor
ERES	Estrogen Receptor Expert System
ERTA	Estrogen Receptor Transcriptional Activation *in vitro* assay
ExpoCast	EPA's Exposure foreCast prioritization research program
FDA	United States Food and Drug Administration
FFDCA	Federal Food, Drug, and Cosmetic Act
FIFRA	Federal Insecticide, Fungicide, and Rodenticide Act
FQPA	Food Quality Protection Act
FRN	Federal Registrar Notice
HDT	Highest Dose Tested
HPV	High Production Volume
HT	High-Throughput
HTE	High-Throughput Exposure
HTTK	High-Throughput Toxicokinetics
IBER	Integrated (RTK) Bioactivity Exposure Ranking
IC50	Half-Maximal Activity. The Concentration of an Inhibitor Where the Response (or Binding) Is Reduced by Half.
IVIVE	*In Vitro-In Vivo* Extrapolation
IYD	Iodotyrosine Deiodinase
KE	Key Event
LAGDA	Larval Amphibian Growth and Development Assay
LT	Low-Throughput
MAD	Mutual Acceptance of Data
MEOGRT	Medaka Extended One Generation Reproduction Test
MIE	Molecular Initiating Event
MySQL	an open-source relational database management system that can be interfaced via a structure query language or SQL
NCGC	NIH National Chemical Genomics Center
NHANES	National Health and Nutrition Examination Survey
NICEATM	NIH National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods
NIH	United States National Institutes of Health
NIS	Sodium-Iodide Symporter
NTP	NIH National Toxicology Program
OCSPP	U.S. EPA Office of Chemical Safety and Pollution Prevention

OPPT	EPA Office of Pollution Prevention and Toxics
ORD	EPA Office of Research and Development
OSRI	Other Scientifically Relevant Information
OW	U.S. EPA Office of Water
QSAR	Quantitative Structure Activity Relationship
RBA	Relative Binding Affinity
RTK	Reverse Toxicokinetics
SAP	Scientific Advisory Panel
SDWA	Safe Drinking Water Act
SEEM	Systematic Empirical Evaluation of Models framework
T3	3,3′,5-Triiodothyronine
T4	Thyroxine
TDCs	Thyroid Disrupting Chemicals
TH	Thyroid Hormone
TK	Toxicokinetic
ToxCast	EPA's Toxicity foreCast prioritization research program
Tox21	Toxicology in the 21st Century – the NTP/NCGC/EPA/FDA consortium for chemical hazard HT
TPO	Thyroperoxidase
TR	Thyroid Hormone Receptor
TRHR	Thyrotropin Releasing Hormone Receptor
TSCA	Toxic Substance Control Act
TSH	Thyroid Stimulating Hormone
TSHR	Thyroid Stimulating Hormone Receptor
WOE	Weight of Evidence

1. Background

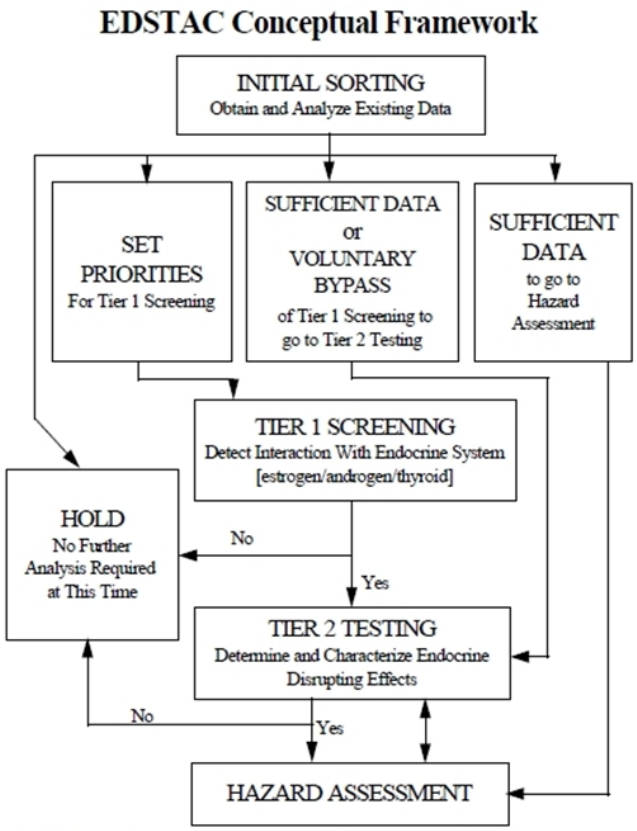
1.1 Statutory Mandate and History of the Endocrine Disruptor Screening Program (EDSP)

Congress passed the Food Quality Protection Act (FQPA) in 1996, which amended the Federal Food, Drug, and Cosmetic Act (FFDCA), and the Safe Drinking Water Act (SDWA) Amendments. FQPA requires that EPA screen pesticide chemicals for their potential to produce effects similar to those produced by the female hormones (estrogen) in humans and gives EPA the authority to screen certain other chemicals and to include other endocrine effects. EPA developed the EDSP in response to the statutory mandate in the Federal Food, Drug, and Cosmetic Act (FFDCA), Section 408(p), to "develop a screening program…to determine whether certain substances may have an effect in humans that is similar to an effect produced by a naturally occurring estrogen, or such other endocrine effects as the Administrator may designate". As part of the EDSP, the statute also gives EPA the authority to "provide for the testing of all pesticide chemicals (active and inert ingredients), as well as any other substance that may have an effect that is cumulative to an effect of a pesticide chemical if the Administrator determines that a substantial population may be exposed to such a substance". In addition to FFDCA, SDWA gives EPA authority to provide for testing "of any other substances that may be found in sources of drinking water if the Administrator determines that a substantial population may be exposed to such substance". Also, provision was made to allow the Agency to "issue an order to a registrant of a substance for which testing is required, or to a person who manufactures or imports a substance for which testing is required, to conduct testing in accordance with the screening program…, and submit information obtained". The mandate also authorizes the Agency and provides the procedure to suspend the sale or distribution of a substance, as necessary "to ensure protection of public health".

1.2 Testing Under the EDSP and the Importance of High-Throughput Screening

To begin meeting these statutory mandates, in 1996 the EPA chartered a Federal Advisory Committee, the Endocrine Disruptor Screening and Testing Advisory Committee (EDSTAC), to address endocrine disruption. The EPA largely adopted the EDSTAC recommendations and established the EDSP in an August 1998 Federal Register notice (U.S. EPA, 1998a). The EDSP was designed with a two-tiered structure where Tier 1 assays are intended to function as a battery to inform on the potential for a chemical to interact with the endocrine system (**Figure 1-1**).

Figure 1-1 EDSTAC Conceptual Framework



(U.S. EPA, 1998a)

The Tier 1 battery of assays was peer reviewed by the FIFRA SAP in 2008 (U.S. EPA, 2008) and were published in 2009 (**Table 1-1**). Tier 1 battery data, and any Other Scientifically Relevant Information (OSRI), associated with a screened chemical is reviewed using a Weight of Evidence (WoE) evaluation to determine if Tier 2 data are needed. EPA published WoE guidance in 2011 (U.S. EPA, 2011b). Tier 2 tests are intended to establish the potential for endocrine disruption and provide dose-response information for chemicals that demonstrated evidence of potential endocrine activity in Tier 1 assays. Tier 2 ecotoxicity tests were peer reviewed by the FIFRA SAP in 2013 (U.S. EPA, 2013d) and were published in 2015.

Table 1-1 Table Listing EDSP Test Guidelines

Listings are grouped by Tier and type (*in vitro* vs. *in vivo*) and each includes: assay/test name (test species); guideline(s); whether the data is informative of potential interaction with the estrogen receptor (E), androgen receptor (A), steroidogenesis (S), hypothalamic-pituitary-gonadal axis (HPG) or hypothalamic-pituitary-thyroid axis (HPT); and, HT alternative status. All Tier 2 tests are *in vivo* assays. (U.S. EPA, 2017b)

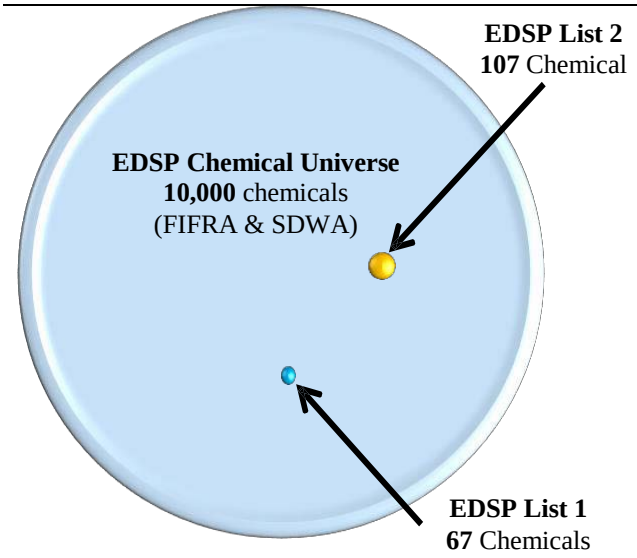
Assay/Test Name (Test Species)	Guideline	E	A	S	HPG	HPT	HT Status
Tier 1 *In Vitro* Test Guidelines							
Androgen Receptor Binding (Rat Prostate Cytosol)	(OCSPP 890.1150 U.S. EPA, 2009c)		■				Proposed[2]
Aromatase (Human Recombinant)	(OCSPP 890.1200 U.S. EPA, 2009d)	■					
Estrogen Receptor Binding Assay Using Rat Uterine Cytosol (ER-RUC)	(OCSPP 890.1250 U.S. EPA, 2009e)	■					Yes[3]
Estrogen Receptor Transcriptional Activation (Human Cell Line HeLa-9903)	(OCSPP 890.1300 U.S. EPA, 2009f) (OECD 455 OECD, 2009d)	■					Yes[3]
Steroidogenesis (Human Cell Line-H295R)	(OCSPP 890.1550 U.S. EPA, 2009a) (OECD TG 456 OECD, 2011)			■			Proposed[2]
Tier 1 *In Vivo* Test Guidelines							
Amphibian Metamorphosis (Frog)	(OCSPP 890.1100 U.S. EPA, 2009b) (OECD 231 OECD, 2009a)					■	
Fish Short-Term Reproduction Assay	(OCSPP 890.1350 U.S. EPA, 2009g) (OECD 229 OECD, 2012a)	■	■		■		
Hershberger Bioassay (Rat)	(OCSPP 890.1400 U.S. EPA, 2011a) (OECD 441 OECD, 2009c)		■	■[1]			
Pubertal Development and Thyroid Function in Intact Juvenile/ Peripubertal Female Rats	(OCSPP 890.1450 U.S. EPA, 2009i)	■			■	■	
Pubertal Development and Thyroid Function in Intact Juvenile/ Peripubertal Male Rats	(OCSPP 890.1500 U.S. EPA, 2009j)		■		■	■	
Uterotrophic Assay (Rat)	(OCSPP 890.1600 U.S. EPA, 2011a) (OECD 440 OECD, 2009b)	■					Yes[2]
Tier 2 Test Guideline							
Avian Two-Generation Toxicity Test in the Japanese Quail	(OCSPP 890.2100 U.S. EPA, 2015a)	■	■	■	■	■	

Assay/Test Name (Test Species)	Guideline	E	A	S	HPG	HPT	HT Status
Medaka Extended One Generation Reproduction Test (MEOGRT) (Fish)	(OCSPP 890.2200 U.S. EPA, 2015b)	■	■	■	■		
Larval Amphibian Growth and Development Assay (LAGDA) (Frog)	(OCSPP 890.2300 U.S. EPA, 2015c)					■	
Reproduction and Fertility Effects (Rat)	(OCSPP 870.3800 U.S. EPA, 1998b)	■	■	■	■		
Extended One-Generation Reproductive Toxicity Study (Rat)	(OECD 443 OECD, 2012b)	■	■	■	■	■	

[1] 5α-reductase inhibition only
[2] EPA, 2017, This SAP white paper
[3] EPA, 2015, FR Notice

The EDSP selected chemicals to be reviewed for endocrine disruption potential. The first list of chemicals to be tested were selected primarily based on exposure and were active pesticide ingredients and high production volume (HPV) pesticidal inert ingredients. The second list of chemicals included these substances and contaminants that may be found in sources of drinking water. In November 2012, EPA published the document entitled "Endocrine Disruptor Screening Program Universe of Chemicals and General Validation Principles" identifying a universe of approximately 10,000 chemicals for EDSP prioritization and screening (U.S. EPA, 2012), based on the statutory authorities of the FFDCA and SDWA. This historical approach to listing for screening chemical substances based primarily on exposure considerations and has only addressed a small fraction of the chemical universe for screening (**Figure 1-2**). Even for the 174 chemicals on Lists 1 (U.S. EPA, 2009k, 2007) and 2 (U.S. EPA, 2013a), EPA has been careful to note that the public should not presume that EPA currently suspects such chemicals interfere with the endocrine systems of humans or other species just by designating a list for screening.

Figure 1-2 EDSP Chemicals



Of the approximately 10,000 EDSP chemicals, including pesticide chemicals (covered under the Federal Insecticide, Fungicide, and Rodenticide Act; FIFRA) and chemicals that may be found in sources of drinking water (covered under the Safe Drinking Water Act, SDWA), only 67 List 1 (U.S. EPA, 2009k, 2007) and 107 List 2 (U.S. EPA, 2013a) chemicals have been identified for screening and testing to date.

Based on the current pace of the Tier 1 screening assays, it could take decades to screen all 10,000 chemicals in the EDSP Universe, millions of dollars, and the sacrifice of many test animals to screen substances for further testing and evaluation. Therefore, EPA's EDSP is actively pursuing the application of computational toxicology and exposure assessments to create a more efficient and robust high-throughput (HT) screening program.

1.3 Summary of Previous EDSP21 SAPs

Incorporating innovative computational toxicological tools allows the Agency to integrate bioactivity and exposure to prioritize and screen chemicals. This approach is consistent with what EDSTAC originally recommended in 1998. The recent advances in computational toxicology have brought EPA to an "evolutionary turning point" for EDSP prioritization, screening, and testing. In addition to rapidly screening thousands of chemicals and overcoming throughput limitations of traditional chemical toxicity testing, EPA expects HT computational and pathway models can serve as alternative tests that will also reduce the need of animals.

In order to use HT methods to prioritize and screen chemicals, EPA must demonstrate that these methods are scientifically sound and can effectively detect endocrine bioactivity via receptor-mediated interactions and broader endocrine pathway effects. In fiscal year 2012, EPA began a multi-year transition for the EDSP to validate and incorporate computational toxicology methods and HT screens. The Agency has requested the Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA) Scientific Advisory Panel (SAP) to provide comment on EPA's scientific framework for the application of HT assays and computational tools as alternative

screening methods in the EDSP (**Table 1-2**). The Agency has carefully considered all feedback that the SAP provided and responded appropriately (**Table 1-3**).

Table 1-2 A summary of the Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA) Scientific Advisory Panel Reviews addressing computational tools for use in the Endocrine Disruptor Screening Program (EDSP)

FIFRA Scientific Advisory Panel Reviews (External Independent Scientific Peer Review)	**Topics for Review**
Scientific issues associated with prioritizing the universe of EDSP chemicals using computational toxicology tools – January 2013	• 8 ER assays • Physical-chemical properties • ERES QSAR
Scientific issues associated with new HT methods to estimate chemical exposure – July 2014	• ExpoCast • SEEM framework • HTTK/RTK
Integrated endocrine bioactivity and exposure-based prioritization and screening for the EDSP – December 2014	• 18 ER assays presented as an alternative model for ER binding, ERTA, and uterotrophic assays • Preliminary AR work discussed • Prioritization using IBER described. • Scientific evaluation for partial Tier 1 battery substitution
HT screening assays and models as alternative methods to Tier 1 battery – November 2017	• 11 AR assays presented as an alternative model for the AR binding assay • HT H295R steroidogenesis assay presented as an alternative to the LT H295R assay • Thyroid pathway framework detailed
FUTURE: Use of computational toxicology for predicting endocrine bioactivity in an adverse outcome pathway framework	• Thyroid-related outcomes • Metabolism • Prioritization • *In vivo* to *in vitro* dose extrapolation • Tier 1 *in vivo* assay alternatives

Abbreviations: ER = estrogen receptor, ERES = ER Expert System, QSAR = Quantitative Structure Activity Relationship, ExpoCast = Exposure Forecaster, SEEM = Systematic Empirical Evaluation of Models, HT = High-throughput, LT – Low-throughput, HTTK = High-throughput ToxicoKinetic, RTK = Reverse ToxicoKinetic, IBER = Integrated Bioactivity Exposure Ranking, AR = Androgen Receptor, E= Estrogen, A = Androgen, T = Thyroid, CompTox = Computational Toxicology.

Table 1-3 A summary of the Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA) Scientific Advisory Panel Comments and Agency Responses.

FIFRA SAP Comments	Agency Responses
January 2013 (U.S. EPA, 2013c): (1) Inclusion of exposure information should be considered earlier in the process. (2) The inclusion of HT assays to measure estrogen, androgen, or thyroid hormonal bioactivity could significantly contribute to prioritization. (3) The eight assays provided for review need refinement and the full suite of assays targeting different endpoints along the estrogen receptor pathway is needed for a full evaluation. (4) HT assays measuring androgen bioactivity should be focused on androgen receptor antagonism. (5) HT assays for the thyroid pathway will involve multiple modes of action that are not receptor based.	Revised the full suite of ER HT assays, targeting different endpoints along the estrogen receptor pathway, and developed an ER pathway model, which was presented to the FIFRA SAP in December 2014 (U.S. EPA, 2014a). The androgen receptor will be addressed in the November 2017 SAP. The thyroid framework presented in the 2017 SAP demonstrates the Agency's efforts to determine multiple modes of action. Exposure and prioritization was addressed in the July 2014 (U.S. EPA, 2014c) and further work will be presented in the future.
July 2014 (U.S. EPA, 2014c): The FIFRA SAP indicated EPA's SEEM framework is scientifically sound and suitable for using HT exposure methods to estimate relative levels of chemical exposures; however, they recommended further ER pathway model evaluation with more chemical data to reduce uncertainty.	EPA adapted and refined the models in the SEEM framework based on the comments of the FIFRA SAP. The adapted framework was presented to the FIFRA SAP (December 2-4, 2014 meeting) (U.S. EPA, 2014a) as part of a proposed integrated bioactivity and exposure approach for prioritization.

FIFRA SAP Comments	Agency Responses
December 2014 (U.S. EPA, 2014a): **ER bioactivity** In general, the Panel agreed that the ER pathway model for assessing estrogen bioactivity had several strengths, *e.g.*, the ER AUC approach was a computationally, resourceful, and insightful approach to determine the estrogenic bioactivity of a chemical. The Panel pointed out the models did not incorporate uncertainty or sensitivity analyses and recommended the Agency explore the inclusion of such analyses. The Panel also noted that the Agency should provide more transparency in describing details about the underpinnings of the model.	EPA scientists published the following scientific papers to support the scientific underpinning of the performance based validation approach of the ER HT assays and the ER pathway model with further refinements in progress. (Kleinstreuer et al., 2016; Browne et al., 2015; Judson et al., 2015) EPA responded to the need for transparency in communication. On June 19, 2015, EPA Published an FRN for Public Comment and Review (U.S. EPA, 2015f). The approach in the FRN ties together the scientific elements that were reviewed by the FIFRA SAP (U.S. EPA, 2013c) and (U.S. EPA, 2014a)\(U.S. EPA, 2014a) to support the use of HT assays and computational tools. In this notice, EPA indicated that the suite of 18 ER HT assays and ER pathway model will be accepted as alternatives to the following three Tier 1 assays: • Estrogen receptor (ER) binding *in vitro* assay (OCSPP 890.1250) • Estrogen receptor transcriptional activation *in vitro* assay (ERTA; OCSPP 890.1300) • Uterotrophic *in vivo* assay (OCSPP 890.1600; OECD No.440) In the FRN, EPA provided information to access all of the *in vitro* HT assay data and computational model bioactivity scores for chemicals run through EPA's ToxCast program to support the EDSP through a publicly-accessible web portal (U.S. EPA, 2017a). EPA is reviewing the public comments and intends to publish the response when completed.

FIFRA SAP Comments	Agency Responses
December (U.S. EPA, 2014a): **AR bioactivity** The Panel recommended that the HT AR bioactivity assay battery include careful assessments and attention to the potential cytotoxic effects of chemicals that may otherwise appear as false positives due to assay interference. The Panel also recommended that the range of chemical structures tested in the assay battery should be expanded to maximize the screening potential. Furthermore, the AR bioactivity battery should include methods to assess the potential effects of non-classical/ non-genomic mechanisms that mimic or inhibit androgen bioactivity.	In the June 19, 2015 FRN (U.S. EPA, 2015g), EPA stated that it will continue the development of HT assays and computational tools for screening for bioactivity in the estrogen, androgen, steroidogenesis, and thyroid pathways and will use a performance-based approach to validate their use as alternatives to assays in the Tier 1 battery. The Agency's responses to Panel's recommendations on AR will be presented at the current FIFRA SAP.
December 2014 (U.S. EPA, 2014a): **IBER approach** The Panel noted that the Agency's proposed IBER approach was rationally developed and laid a foundation for a theoretical basis with the potential to prioritize further EDSP screening of compounds with estrogenic activity. The Panel highlighted the practicality of the statistical modeling of the IBER approach. The Panel cautioned that the approach needed further refinement, before it is employed by the Agency's Endocrine Disruption Screening Program.	EPA is continuing to work on the development and implementation of HT exposure modeling approaches for use in conjunction with the EDSP. These developments will be presented at a future SAP.

Abbreviations: ER = estrogen receptor, ExpoCast = Exposure Forecaster, SEEM = Systematic Empirical Evaluation of Models, HT = High-Throughput, HTTK = High-Throughput ToxicoKinetic, RTK = Reverse ToxicoKinetic, IBER = Integrated Bioactivity Exposure Ranking, AR = Androgen Receptor, E= Estrogen, A = Androgen, T = Thyroid, CompTox = Computational Toxicology.

In January 2013 (U.S. EPA, 2013c), the SAP reviewed an EPA white paper describing eight HT estrogen receptor (ER) agonist assays and the ER Expert System quantitative structure-activity relationship (QSAR) model as potential tools to be used for prioritizing EDSP chemicals. The SAP found steps in the prioritization scheme to be organized well and clearly described. Panel members encouraged expanding the use of HT and highlighted the need to consider exposure in the process. The SAP also noted that the androgen pathway may be similar to the ER pathway, but efforts should focus on androgen antagonism. The SAP remarked that endocrine effects in the thyroid pathway involve multiple modes of action that are not receptor-based.

For the July 2014 SAP (U.S. EPA, 2014c), EPA presented scientific issues associated with new HT methods to estimate chemical exposure. ExpoCast is an initiative of the EPA Office of Research and Development (ORD) to develop the necessary approaches and tools for rapidly prioritizing and screening thousands of chemicals, based on the potential for human and ecological exposure. Many existing traditional, lower throughput exposure tools require considerable data to generate screening level assessments. ExpoCast efforts have focused on empirically assessing the uncertainty in high-throughput exposure (HTE) predictions by evaluation and calibration with available monitoring data (*e.g.*, biomonitoring or water concentration). By using rigorous statistical techniques to reconcile HTE estimates based on limited available data, the uncertainty in the HTE predictions is quantified. EPA refers to this framework as the "Systematic Empirical Evaluation of Models" (SEEM).

In addition to predictions of exposure, the July 2014 SAP reviewed high-throughput toxicokinetic (HTTK) methods for extrapolating *in vitro* doses to *in vivo* concentration (*in vitro*-to-*in vivo* extrapolation or IVIVE) for chemicals that have been run in a battery of high-throughput endocrine screening assays (*e.g.*, ToxCast). HTTK provides a bridge between bioactivity measured in the high-throughput screening assays and exposure by either estimating tissue concentrations from an administered dose (*i.e.*, what has been called forward toxicokinetics) or inferring administered doses that would be needed to cause tissue bioactive concentrations *in vivo* (*i.e.*, reverse toxicokinetics or RTK).

The December 2014 FIFRA SAP (U.S. EPA, 2014a) review focused on additional HT bioactivity tools for the estrogen and androgen pathways, their use in chemical prioritization, and the initial evaluation of their potential to substitute for specific endpoints in lower throughput EDSP Tier 1 screening assays. In addition, the SAP reviewed an approach that utilized both bioactivity and exposure to prioritize chemicals for further consideration. In the Integrated Bioactivity-Exposure Ranking (IBER), RTK can be used to estimate the daily administered dose (mg/kg BW/day) necessary to produce steady-state *in vivo* blood concentrations equivalent to concentrations showing biological activity in the *in vitro* HT assays (μM). The putative bioactive administered doses can then be directly compared with predicted exposures (mg/kg BW/day).

The ER computational model (Judson et al., 2015) combines data from 18 different *in vitro* assays in the ER pathway assays in order to generate a consensus call of ER bioactivity. The need for this approach is driven by the presence of false positive and false negative results in individual *in vitro* assays. The performance validation of this model was presented at the FIFRA SAP meeting in 2014 (U.S. EPA, 2014a) and further detailed by (Browne et al., 2015).

For technical and cost-related reasons, it will be difficult to procure all EDSP Universe chemicals and process them through all of the relevant *in vitro* assays. Thus, the ongoing development of structure-based models (QSARs and molecular docking) was discussed at the FIFRA SAP meeting in 2014. The Agency envisions use of these models to extrapolate from the current *in vitro* data set and to guide the acquisition of further *in vitro* and *in vivo* data in a targeted fashion. Following the FIFRA SAP meeting in 2014, a consensus QSAR model was published that integrated a total of 40 categorical and 8 continuous models for binding, agonist, and antagonist ER activity (Mansouri et al., 2016). The development of the consensus QSAR model was led by the EPA and involved 17 groups in the United States and Europe and included a total of 40 categorical and 8 continuous models for binding, agonist, and antagonist ER activity.

The first-generation AR pathway model discussed in the FIFRA SAP meeting in 2014 combined data from 9 different HT *in vitro* assays in the AR pathway assays in order to generate a consensus call on AR bioactivity. The need for this approach is driven by the presence of false positive and false negative results in individual *in vitro* assays. The AR pathway model was built from assays measuring multiple endpoints in the AR signaling pathway using different technologies. AR agonist and antagonist AUC scores compare very well with reported bioactivity of reference chemicals across a range of structures (85% concordance). At the time of the review, the AR potency was not yet defined for the reference chemicals used in this previous analysis, many of which are pharmaceuticals designed to target the androgen receptor and are therefore expected to be strongly bioactive.

Though using computational estimates of AR bioactivity to contribute to weight of evidence determinations or substitute for all AR specific Tier 1 battery endpoints is premature at this time (2014), AR bioactivity predicted by the ToxCast AR assays may support prioritizing chemicals with positive model scores for additional screening and testing. A description was also provided of an effort to compare computational model scores from the AR pathway model with results of the Hershberger assay to determine model performance for predicting *in vivo* bioactivity.

At the FIFRA SAP meeting in 2014 (U.S. EPA, 2014a), descriptions were provided of methods to quantitatively predict concentrations at which endocrine-pathway activity can occur; how to convert these concentrations to oral equivalent doses using HTTK methods; and how to estimate exposures for chemicals. IBER is used to determine if a chemical has the potential to activate an endocrine-related pathway at exposures / doses that are likely to occur.

The basic approach for calculating the pathway-activity dose distribution was developed by (Judson et al., 2011). In this procedure, the analyzer estimates the concentration at which the pathway becomes active, uses HTTK to convert from concentration to dose, and then compares this dose against exposure. Each of the quantities (activity concentration, concentration-to-dose conversion, and exposure estimate) is subject to uncertainty and population variability. After combining point estimates and uncertainty and variability ranges, the IBER is the ratio between the lowest X-percentile (*e.g.* 5%) of the bioactivity dose range and the upper Y-percentile (*e.g.* 95%) of the expected exposure range. The IBER is one proposed method for prioritization of limited testing resources for chemicals.

1.4 Comments from the SAP of December 2014 and Agency Responses

1.4.1 ER Bioactivity

In general, the Panel agreed that the ER pathway model for assessing estrogen bioactivity was a computationally resourceful and insightful approach to determine the estrogenic bioactivity of a chemical substances. The Panel pointed out that the ER pathway model did not incorporate uncertainty or sensitivity analyses and recommended that the Agency include these analyses. The Panel also noted that the Agency should provide more transparency in describing details of the model.

Following the December 2014 FIFRA SAP meeting (U.S. EPA, 2014a), EPA scientists authored and co-authored three key papers (summarized below) that addressed a number of the FIFRA SAP comments on the HT ER assays, data analysis, ER pathway model, and performance-based validation.

(Judson et al., 2015) published a detailed description of the ER ToxCast data analysis and ER pathway model, including a discussion of uncertainty and sensitivity in the model. The following items are described in the publication:

- the 18 *in vitro* HT assays used to probe the ER pathways and results from 1812 chemicals
- the set of 45 positive and negative reference chemicals used to evaluate the performance of the ER pathway model
- the approach for processing the concentration-response data
- the scheme to account for non-specific, cell-stress/cytotoxicity-related assay interference (“false positives”)
- the structure and mathematical representation of the ER pathway model

(Kleinstreuer et al., 2016) published a comprehensive database of quality-controlled *in vivo* uterotrophic studies based on the systematic review of scientific literature presented to the FIFRA SAP in December 2014 (U.S. EPA, 2014b). This database was used to assess the reproducibility of the uterotrophic bioassay.

Finally, (Browne et al., 2015) detailed the multi-step performance-based validation process for the 18 ER HT assays and ER pathway model that are used as alternatives to the following three Tier 1 assays: ER binding *in vitro* assay (OCSPP 890.1250), ER transcriptional activation *in vitro* assay (ERTA; OCSPP 890.1300), and uterotrophic *in vivo* assay (OCSPP 890.1600; OECD No.440). ER pathway model performance was evaluated for reference chemicals selected from the uterotrophic database (Judson et al., 2015), as well as results of EDSP Tier 1 screening assays in current practice. The ER pathway model accuracy was 86% to 93% when compared to reference chemicals and predicted results of EDSP Tier 1 guideline and other uterotrophic studies with 84% to 100% accuracy. The HT assays and ER pathway model predictions correctly identified active and inactive reference chemicals, provided a measure of relative ER bioactivity, and rapidly identified chemicals with potential endocrine bioactivities for additional screening and testing.

EPA responded to the need for transparency of how the Agency will use HT assays and computational tools in a Federal Register Notice (FRN) published June 19, 2015 (U.S. EPA, 2015f), including a request for public comment. The "Pivot" approach described in the FRN policy notice integrates the Agency's regulatory policy to use HT assays and computational tools as alternative methods for screening chemicals in the EDSP with the scientific elements supporting the policy, *i.e.* external scientific peer review by the FIFRA SAP (January 2013 (U.S. EPA, 2013c) and December 2014 (U.S. EPA, 2014a)) and data published in the scientific literature. In this notice, EPA indicated that the suite of 18 ER HT assays and ER pathway model will be accepted as alternatives to the following three Tier 1 assays:

- Estrogen receptor (ER) binding *in vitro* assay (OCSPP 890.1250)
- Estrogen receptor transcriptional activation *in vitro* assay (ERTA; OCSPP 890.1300)
- Uterotrophic *in vivo* assay (OCSPP 890.1600; OECD No. 440)

EPA indicated that it would continue to develop a set of "non-animal" HT assays and computational bioactivity models as alternatives to all of the assays in the current Tier 1 screening battery. In the FRN, EPA indicated that all of the *in vitro* HT assay data and computational model bioactivity scores for chemicals run through EPA's ToxCast program to support the EDSP were publicly-accessible through a public web portal (U.S. EPA, 2017a).

1.4.2 AR Bioactivity

The EPA asked the FIFRA SAP for comment on the 9 AR HT assays and the first-generation AR pathway model. The Panel indicated that EPA should increase the robustness and reliability of the AR pathway model. The Panel recommended that the HT AR bioactivity assay battery include careful assessments and attention to the potential cytotoxic effects of chemicals that may otherwise appear as false positives, due to assay interference. The Panel also recommended that the range of chemical structures tested in the assay battery should be expanded to maximize the screening potential. Furthermore, the AR bioactivity battery should include methods to assess the effects from non-classical/non-genomic mechanisms that mimic or inhibit androgen bioactivity.

Subsequent to the FIFRA SAP meeting, EPA has almost tripled the number of reference chemicals (23 chemicals in 2014 vs 37 agonists and 28 antagonists in 2016). A second-generation AR pathway model was developed that integrates data from 11 *in vitro* HT assays. The sensitivity and specificity are now >95% for these reference chemicals, and the model adequately explains those reference chemicals missed, as will be discussed later (Kleinstreuer et al., 2017a). A description of a systematic literature search is presented that identified 158 putative androgen-active and inactive chemicals, 57 agonist and antagonists (positive), and 101 negative reference chemicals.

The publication describing the second-generation AR pathway model also details the approach for processing the concentration-response data, the scheme to account for non-specific, cell-stress/cytotoxicity-related assay interference ("false positives"), and the structure and mathematical representation of the AR pathway model. The AR ToxCast bioactivity data are accessible through a public web portal (U.S. EPA, 2017a). EPA is currently investigating non-classical/non-genomic mechanisms that mimic or inhibit androgen bioactivity.

1.4.3 IBER Approach

The Panel (U.S. EPA, 2014a) noted that the Agency's proposed IBER approach was rationally developed and laid a foundation for a theoretical basis with the potential to prioritize further EDSP screening of compounds with estrogenic activity. The Panel highlighted the practicality of the statistical modeling of the IBER approach. The Panel cautioned that the approach needed further refinement, before it is employed by the Agency's Endocrine Disruption Screening Program. EPA is continuing to work on the refinement and implementation of HT exposure modeling approaches for use in conjunction with the EDSP. These developments will be presented at a future FIFRA SAP meeting for peer review and/or additional input.

1.5 Performance-Based Approach to Establishing Confidence

The Agency continues to develop HT alternative methods to the established EDSP Tier 1 assays, based upon performance-based acceptance criteria, and confirms the validity of the model and acceptance criteria through peer review of the FIFRA SAP and public comment. The use of performance-based approach offers several advantages. The performance-based approach includes identifying toxicity endpoints and/or key events, well-characterized performance standards, and clearly-defined acceptance criteria. The performance-based approach will streamline the validation of both individual assays and defined approaches (testing strategies that may combine *in vitro* and/or *in silico* methods), facilitate mutual acceptance of data (MAD). The performance-based approach will also facilitate innovation and implementation of novel approach methodologies that reduce the need for animal testing and provide potentially better protection of human health and the environment. Moreover, this approach will allow for harmonization with OECD to more effectively use resources and limit duplicative and unnecessary activities by technical workgroups.

New Approach Methodologies (NAMs) refers to any non-animal technology, methodology, approach, or combination thereof that can be used to provide information on chemical hazard assessment. Performance-based approaches for establishing confidence in NAMs are intended to provide a framework that facilitate more efficient development and faster implementation of new test methods. Performance-based approaches are developed in accordance with the principles set forth in the OECD *Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment (GD 34).* Specifically, OECD GD 34 states that "new test methods undergo validation to assure that they employ sound science and meet regulatory needs" (fit-for-purpose), "the validation process should be flexible and adaptable", and that performance must be "demonstrated using a series of reference chemicals" and "evaluated in relation to existing relevant toxicity data."

OECD GD 34 further defines ***relevance*** of a test method as encompassing the regulatory need, usefulness of the alternative method(s) and associated limitations of the test method. ***Reliability*** is defined in GD34 as the extent of reproducibility of results from a test within and among laboratories over time, when performed using the same standardized protocol. However, demonstrating between-lab reproducibility is applicable only when developing a standardized protocol intended to be used by any naïve laboratory in any world region. Many 21st century

technologies are not amenable to transfer into naive laboratories. Cross-laboratory validation may not be necessary when performance-based criteria are fully evaluated and apparent.

In this approach, data from multiple assays, each assessing a separate key event in an Adverse Outcome Pathway (AOP), are integrated in a prediction model of hazard. In keeping with the AOP paradigm, it is the resulting prediction that is used to validate the model and not the performance of the individual assays, since the results from each assay (key event) may play only a limited role on its own. This framework was initially proposed by members of International Cooperation on Alternative Test Methods (ICATM) and provides the foundation for work being done internationally to develop consensus on non-animal approaches towards the complete replacement of the laboratory animal studies commonly used to evaluate skin sensitization. The considerations in this framework are relevant to evaluating a performance-based approach for the EDSP and are adapted accordingly below as the organizing principles used below in the AR and steroidogenesis chapters.

- The alternative *in vitro* assays or computational models should be mechanistically and/or biologically relevant to the hazard being assessed.
- The reliability of the alternative approach should be considered within the context of intended use and accepted best practices within the given field.
- The alternative *in vitro* assays or computational models should be transparently described and all data made available to the public (e.g., any datasets and modelling code are publicly available and its known limitations are clearly described).
- Criteria for selecting reference chemicals should be defined and supporting data should be adequately referenced.
- Evaluation and implementation by third parties must be possible (i.e., the alternative approach must be readily accessible commercially and/or the relevant protocols must be available).
- Uncertainty limitations should be described to the fullest extent possible.
- The alternative *in vitro* assays or computational models should undergo an independent scientific review in order to raise confidence in the approach.

Satisfying the regulatory need is part of establishing relevance. The decision context for a model's or assay's fit-for-purpose need means that the model/assay must be able to determine potential endocrine bioactivity of chemicals (for Tier 1 assays). Measures used to

establish a model/assay fit-for-purpose performance include specificity, sensitivity, and accuracy. These measures can be made by evaluating the model/assay performance using reference chemicals and by comparing the performance to the results of existing guideline-like tests which evaluate similar perturbations. Specificity, sensitivity, and accuracy measures will be presented later in the AR and steroidogenesis chapters but are defined here. Sensitivity and specificity are statistical measures of the performance of a binary classification test. Sensitivity measures the proportion of positives that are correctly identified as such, while specificity measures the proportion of negatives that are correctly identified as such. So, sensitivity quantifies the avoidance of false negatives, and specificity quantifies the avoidance of false positives. Mathematically, sensitivity equals the number of true positives ÷ (number of true positives + number of false negatives). Specificity = number of true negatives ÷ (number of true negatives + number of false positives). Accuracy is the proportion of correct outcomes of a test method.

1.6 Computational and Pathway Models to Integrate High-Throughput Assay Data

Every assay has inherent limitations driven by technological specifications and applicability domains. A biological pathway-based approach that integrates multiple assays mapping to key upstream and downstream events provides a weight of evidence for the true potential of a chemical to activate or repress signaling. This type of additive model helps compensate for the individual shortcomings of any one assay.

A computational model based on the pathway approach has been used previously to predict ER activity (Browne et al., 2015; Judson et al., 2015). A similar approach has also been used to predict AR activity (Kleinstreuer et al., 2017a). Importantly, such a systems biology approach overcomes key technical challenges in integration of HT screening assays.

Performance characteristics of each assay technology will have specific weaknesses that can be mitigated by considering the results from all of the available assay technologies. For example, one assay may be extremely sensitive, but not very specific, leading to a higher incidence of false positives. Conversely, a specific assay may demonstrate increased specificity, leading to a reduced capacity to identify weak positives.

Apart from the pathway itself, cytotoxicity must also be evaluated for some assays. Cytotoxicity in cell-based assays may confound receptor antagonism, *i.e.* it may be difficult to distinguish the source of a decreased assay signal, resulting in higher false positive rate in that particular assay.

The pathway model aims to generate a prediction of bioactivity of greater confidence by understanding where a chemical may act in the biological pathway, minimizing the incidence of false negatives, using a consensus result. This consensus approach enables greater confidence in the assessment of *in vitro* bioactivity for a chemical. As the AR model also includes assays to detect changes at several points in the biological pathway (introduced in Section 2.1), the AR model can also be referred to as an AR pathway model.

In this document, the AR model will be referred to as the "AR pathway model". Maintaining a consistent term of reference should alleviate any confusion, as there is only one AR model referred to in this document.

1.7 Scientific Issues for this FIFRA SAP Meeting

As stated in the June 19, 2015 FRN (U.S. EPA, 2015f), EPA is continuing the development of HT assays and computational tools for screening for bioactivity in the estrogen, androgen, steroidogenesis, and thyroid pathways and will use a performance-based approach to validate their use as alternatives to assays in the Tier 1 battery (**Table 1-1**). HT assays, such as the ER pathway model, serve as a form of other scientifically relevant information (OSRI) that can function both in the prioritization of substances for future evaluation, as well as in the weight of evidence consideration of test results to determine the need for further evaluation (Tier 2). The Tier 1 battery serves to screen chemicals, identifying chemicals that require further testing (in Tier 2) to determine if these chemicals are endocrine system disruptors and to further detail a dose-response relationship for these chemicals.

The current FIFRA SAP review focuses on three sets of scientific issues: (1) the scientific support for use of the 11 AR HT assays and AR pathway model as an alternative to the EDSP Tier 1 *in vitro* AR binding assay (building on material presented at the December 2014 FIFRA SAP meeting (U.S. EPA, 2014a)); (2) the development of a HT H295R steroidogenesis assay as a potential alternative to the current LT H295R steroidogenesis assay in the Tier 1 battery; and (3) the development of an initial thyroid adverse outcome pathway (AOP) framework.

The Agency's goals for this SAP include:

- Confirmation that the suggested improvements in the AR pathway model (from the December 2014 FIFRA SAP) have been incorporated, allowing the HT AR pathway model to serve as alternative for the low-throughput (LT) AR assay (OSCPP 890.1150) in the form of other scientifically relevant information.
- Confirmation that the newly developed HT H295R steroidogenesis assay (either the version measuring only estrogen and testosterone, the version measuring 11

hormones, or both) can serve as an alternative to the LT H295R assay (OSCPP 890.1550/ OECD TG 456) in the form of other scientifically relevant information.

- Comments and advice concerning the Agency's ongoing development of an approach to detect substances that can perturb thyroid function.

2. Androgen Receptor (AR) Pathway Activity

2.1 Introduction

The AR pathway model described herein is intended as a potential alternative for the existing EDSP Tier 1 AR binding assay. This AR pathway model is a computational approach that integrates activity from multiple *in vitro* assays indicative of AR activity in order to make a prediction of "true" receptor activity. "True" receptor activity is defined as AR binding and/or transactivation by reference chemicals that consistently demonstrates this activity in guideline studies and published reports. Critical to an understanding of the predictivity of this model is the performance of these reference chemicals, and systematic curation of sources to define this set of reference chemicals.

The mammalian androgen receptor (AR) signaling pathway was probed using a set of 11 biochemical and cell-based *in vitro*, HT screening assays. These assays indicate perturbation of key events including receptor binding, receptor dimerization, chromatin binding of the transcription factor complex, and gene transcription. A library of 1855 chemicals (including ToxCast Phases I and II and Tox21 results) was screened using this set of assays. AR agonists and antagonists, as well as selective androgen receptor modulators (SARMs), were included in this chemical library. A pathway model was built using these data to generate AR agonist and antagonist scores. Expected patterns of assay activity include: no assays activated (negative); all agonist or all antagonist assays activated; specific subsets of assays across technologies activated; and technology-specific assay activation. The AR pathway model attempts to identify chemicals that may be more or less likely to be AR agonists or antagonists, and clarify signals that may be more likely due to specific types of assay interference, including cytotoxicity and cell stress.

In addition to the AR pathway model presented here for prediction of AR agonist and antagonist activity, a quantitative structure activity (QSAR) is in development. The details of the QSAR model will undergo further evaluation, and the methodology will parallel the approach taken for the ER QSAR model, in which multiple structure-based models were used in a consensus approach (Mansouri et al., 2016).

The computational approach to combine information from multiple AR assays is very similar to the approach previously used to predict estrogen receptor (ER) activity (Browne et al., 2015; Judson et al., 2015). This pathway approach minimizes the incidence of false negatives by using a consensus result based on the understanding where a chemical may act in the AR pathway. Computational and pathway models were discussed in the introduction (see Section 1.6).

The work presented here represents an update to the first generation AR pathway model described in December 2014 (U.S. EPA, 2014a) for a FIFRA SAP. Since that time, the pathway model has been improved in a number of ways, all of which are contained in the following text and (Kleinstreuer et al., 2017a):

- The assay set was expanded from 9 assays to 11 assays, including an antagonist assay run at two different ligand concentrations as confirmation (for a total of 12 datasets);
- The antagonist scores from the AR pathway model were further qualified using a semi-quantitative system for confidence scoring that includes contributions from the strength of the model score, a flag for cell stress/generalized cytotoxicity (see below), and confirmation assay data;
- A cell stress flag (Judson et al., 2016) incorporating cytotoxicity and proliferation inhibition data from 35 assays across a wide variety of cell types has been generated as a caution on potentially confounded results;
- A systematic literature search to identify reference chemicals for AR agonist and antagonist interactions was completed; and,
- A comparison of the AR pathway model results and literature results has been completed for 29 reference chemicals for AR agonism and 28 reference chemicals for AR antagonism.

The current AR pathway model has been published, and the following text, tables, and figures are reproduced or adapted from that publication (Kleinstreuer et al., 2017a, b).

2.1.1 Assay Data

The HT data used as the basis for this AR pathway activity model are publicly available in a variety of formats (U.S. EPA, 2015e). Assay descriptions are also available from these sources. Further, OECD-compliant summaries of the operating procedure, assay annotations, and performance characteristics for each of the assays used in the AR pathway model are provided in the AR Supplemental File.

2.1.2 Overall Approach

The input data for the AR pathway model includes concentration-response data for 11 AR-related *in vitro* HT assays. Five of the assays measure AR interaction irrespective of agonism or antagonism; 4 assays measure events downstream of AR agonism; and, 2 assays specifically measure AR antagonism (see **Table 2-1**). A cytotoxicity filter for the two assays that measured suppression of protein production, *i.e.* AR antagonism, was applied using data from two parallel cytotoxicity assays. A cell stress flag was also developed to flag potential nonselective assay hits attributable to cell stress using additional data from available assays measuring cytotoxicity or reduced cell proliferation. These additional cytotoxicity-related data were used in calculation of a "Z-score" (see Section 2.2.5 for details) for each AR pathway assay where a chemical was active, which provides a means of understanding the likelihood that activity in the AR pathway represents a true AR pathway interaction rather than a nonselective interaction mediated by cell stress of cytotoxicity-related processes. Finally, a semi-quantitative scoring approach that integrates the score in the AR pathway model, the cell stress flag, and activity pattern in a set of confirmation assays, was used to indicate confidence in antagonist pathway activity.

Evaluating and validating the AR pathway model requires high-quality reference data for AR agonist and antagonist activity. Unlike the ER pathway, which has a well characterized set of *in vitro* and *in vivo* reference chemicals (Kleinstreuer et al., 2017a; Browne et al., 2015; Judson et al., 2015), the reference chemical set for the AR pathway is much less developed. Previous work focused on identifying chemicals that were positive or negative for (anti-)androgenicity, without a specific emphasis on potency, and often included compounds that were "presumed" active or inactive (ICCVAM, 2003). Using a comprehensive list of putative AR-active or inactive chemicals from past and present international validation studies, a systematic literature review was performed to compile high-quality *in vitro* AR binding and transactivation (TA) assay data. To facilitate external validation of the AR pathway model results, no ToxCast or Tox21 assay data were included in the literature search. Chemicals with reliable and reproducible *in vitro* results from the literature (see Section 2.3.2) were identified, and chemicals were binned into defined potency categories. The list of proposed reference chemicals and the supporting data are provided, and were used to evaluate the current computational model of AR pathway activity based on the Tox21 and ToxCast assays.

2.2 Methods

2.2.1 High-Throughput Screening Data

Data on 1855 chemicals were generated during ToxCast Phases I and II and Tox21 screening using 11 AR-related HT *in vitro* assays (**Table 2-1**). These include three biochemical radioligand AR binding assays (Novascreen: (Sipes et al., 2013; Knudsen et al., 2011; Hartig et al., 2008)), a coactivator recruitment assay measuring protein:protein interaction between AR and SRC1 at two different time points (Odyssey Thera), one transactivation assay measuring reporter RNA transcript levels (Attagene: (Martin et al., 2010)), three transactivation assays measuring reporter protein level readouts (Odyssey Thera, Tox21:(Lynch et al., 2016)), and two transactivation antagonist assays (Tox21: (Huang et al., 2011; Hartig et al., 2002; Wilson et al., 2002)). One of the transactivation antagonist assays, the Tox21 antagonist luciferase assay in the MDAKB2 cell line (A11), was run as a confirmation assay with a lower concentration of the synthetic ligand R1881 to verify chemical activity specific to the AR pathway. Higher concentration of the ligand should cause a shift in the concentration-response curves of true competitive antagonists to being active at higher concentrations. The chemicals were tested in concentration–response format in all assays except for the cell-free binding assays. The latter assays were initially tested at a single concentration (25 µM), and if significant activity was seen, the chemical was then tested in concentration–response mode. All concentration–response assay data (U.S. EPA, 2017a) were analyzed using the ToxCast data analysis pipeline (Filer et al., 2016; Filer, 2015), which automates the processes of baseline correction, normalization, curve-fitting, hit-calling, and AC50 (half-maximal activity) determination (Filer, 2015). All *in vitro* assays except the RNA transcript reporter assays (Attagene) were normalized to the range of 0-100%, using the positive control response. RNA transcript reporter data were normalized as a fold-change over the solvent control (0.5-1% DMSO, which has been determined to have no effect on assay performance) and then multiplied by a factor of 25 to yield a range of approximately 0-100. The data from each chemical-assay pair was fit to three models: a constant model, a Hill model, and a Gain-Loss model, and the model with the lowest Akaike Information Criterion (Akaike, 1998) was selected. The pipeline also detects a variety of potential confounders, which are annotated as "caution flags." To facilitate computational synthesis across different *in vitro* assays with different numbers of tested concentrations, a set of synthetic concentration-response activities was generated through interpolation for each chemical-assay pair at standardized concentrations using a Hill equation based on the experimentally derived AC50, Hill slope and Top parameters (Judson et al., 2015). All AC50 values were in µM, and the

synthetic concentrations were a 1.5-fold dilutions series of 45 concentrations from 1 pM to 100 µM. The pipeline and all raw and processed data and annotations are publicly available (U.S. EPA, 2015e) and the data processing is described in detail elsewhere (Filer, 2015; Judson et al., 2015).

Table 2-1 Tox21/ToxCast In Vitro Assays Used in AR Pathway Model

ID	Node	Assay Name	Source	Gene	Species	Type	Associated Pathways ǂ
A1	N1	NVS_NR_hAR	Novascreen	AR	Homo sapiens	Receptor Binding	R1; R2; R3
A2	N1	NVS_NR_cAR	Novascreen	AR	P. troglodytes	Receptor Binding	R1; R2; R3
A3	N1	NVS_NR_rAR	Novascreen	AR	Rattus norvegicus	Receptor Binding	R1; R2; R3
A4	N2	OT_AR_ARSRC1_0480	Odyssey Thera	AR; SRC	Homo sapiens	Coregulator Recruitment	R1; R2; R4
A5	N2	OT_AR_ARSRC1_0960	Odyssey Thera	AR; SRC	Homo sapiens	Coregulator Recruitment	R1; R2; R4
A6	N3	ATG_AR_TRANS	Attagene	AR	Homo sapiens	RNA Reporter Gene	R1; R5
A7	N4	OT_AR_ARELUC_AG_1440	Odyssey Thera	AR; ARE	Homo sapiens	Reporter Gene	R1; R6
A8	N4	Tox21_AR_BLA_Agonist_ratio	NCATS/ NCGC	AR	Homo sapiens	Reporter Gene	R1; R6
A9	N4	Tox21_AR_LUC_MDAKB2_Agonist	NCATS/ NCGC	AR	Homo sapiens	Reporter Gene	R1; R6
A10	N5	Tox21_AR_BLA_Antagonist_ratio	NCATS/ NCGC	AR	Homo sapiens	Reporter Gene	R2; R7
A11	N5	Tox21_AR_LUC_MDAKB2_Antagonist	NCATS/ NCGC	AR	Homo sapiens	Reporter Gene	R2; R7
A11*	N5	Tox21_AR_LUC_MDAKB2_Antagonist-confirmation	NCATS/ NCGC	AR	Homo sapiens	Reporter Gene	NA

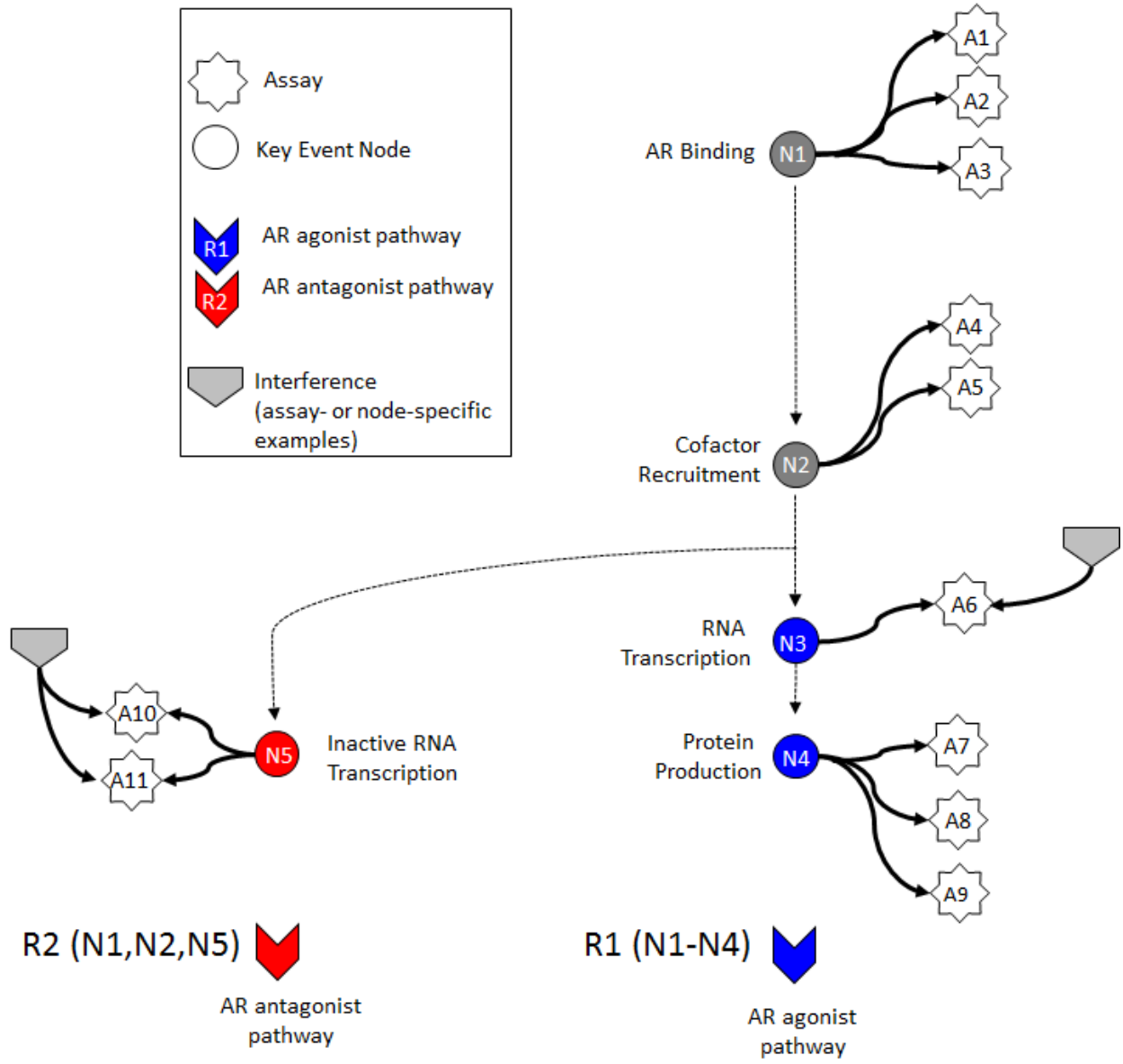
Abbreviations: AR = androgen receptor; ARE = androgen response element; NCGC = NIH Chemical Genomics Center, now part of National Center for Advancing Translational Sciences (NCATS); SRC = c-Src tyrosine kinase.
* Confirmation assay data (overly high concentration of R1881), not used in AR pathway model scores
ǂ Activity in these assays/nodes could be associated with one or more of the following pathways: AR agonist (R1), AR antagonist (R2), or interference (R3-R7). Activity in individual assays could also be associated with assay-specific interference (A1-A11).

2.2.2 AR Pathway Model

A computational network model for AR pathway activity (**Figure 2-1**) was built using 11 ToxCast and Tox21 HT *in vitro* assays (**Table 2-1**) that map to key events in the biological pathway. The R-code for this model is reported here (Watt, 2016). **Figure 2-1** depicts the AR pathway model used to evaluate the integrated *in vitro* assay responses that mirrors previously published work on the ER pathway (Judson et al., 2015), and is based on the series of molecular events that typically occur in a nuclear receptor-mediated response (Gronemeyer et al., 2004). An AR agonist will bind to the receptor monomer (node N1), cause the receptors to dimerize and translocate to the nucleus and recruit coregulators (node N2) to form the complete, active

transcription factor complex. The transcription factor complex binds to the chromatin DNA at specific response element sequences, initiates transcription of mRNA (node N3), and subsequent translation to protein (node N4). An AR antagonist acting through the receptor will bind to the receptor monomer (node N1), cause the receptors to dimerize and translocate to the nucleus and recruit coregulators (node N2), forming a transcription factor complex that binds to the chromatin DNA at specific response element sequences, but is transcriptionally inactive and results in a lack of downstream protein production (node N5). Each of these key event nodes was assessed by one or more of the eleven *in vitro* assays listed in **Table 2-1** (represented in **Figure 2-1** as white stars). **Figure 2-1** shows the two modes of the AR pathway: agonist (nodes associated with R1) and antagonist (nodes associated with R2). The model assumes that a chemical that interacts with the AR will bind and result in either or both of the agonist or antagonist conformations, triggering activity in the appropriate pathway. Each of the *in vitro* assays (A1-A11) is subject to processes that can lead to nonspecific activity independent of the AR pathway event that it is supposed to measure. These may be due to technological interference, artifacts, or other sources of experimental noise. Further, each group of assays that map to a key event node could be affected by non-AR mediated activity specific to that key biological event (such as blocking transcription). Interference pathways R3-R7 correspond to nodes N1-N5. Two examples of interference pathways, one that is assay-specific (A6) and one that is node-specific (R7), are shown in **Figure 2-1** as light gray arrow heads.

Figure 2-1 AR Pathway Model Based on Tox21/ToxCast Assays



Abbreviations: AR = androgen receptor. Graphical representation of the AR pathway model: Circular nodes (N1-N5) represent key biological events along the pathway, where dark grey coloring indicates key events common to agonism and antagonism, and blue and red coloring indicates key events specific to agonism or antagonism, respectively. White stars (A1-A11) represent the *in vitro* assays that measure activity at the biological nodes. Colored arrow heads (R1/R2) represent true AR agonism/antagonism, respectively, and are comprised of the nodes listed in the diagram and their associated assays. Light grey arrow heads demonstrate examples of technology-specific interference or biological interference pathways, where individual assays or specific groups of assays are positive due to non-AR mediated activity. Each *in vitro* assay and each key event node have an assay-specific or biology-specific interference pathway (defined in **Table 2-1**).

2.2.3 Mathematical Representation of the Pathway Model

Following the ER pathway example presented in (Judson et al., 2015), a simple linear additive model is used to predict the relative AR agonist or antagonist activity of a test chemical based on data from the *in vitro* assays that map to the AR pathway in **Figure 2-1**. In the mathematical representation, the term "receptor" can refer to AR-mediated agonism, AR-mediated antagonism, or an interference pathway (mediated via biological activity or nontarget activity associated with a specific technology). The "receptors" R1-R7 associated with each assay or key event node are listed in **Table 2-1**. The model assumes that the value (the efficacy, *A*) returned by an assay at a given concentration is the sum of the contributions from the "receptors" that it measures:

Equation 1

$$A_i = \sum_j F_{ij} R_j$$

Here, the index *i* ranges over the number of assays and the index *j* over the number of "receptors" (where *j*=1 for agonism, *j*=2 for antagonism, and *j*>2 for interference). The elements of the F matrix are 1 if there is a connection between a "receptor" *j* and an assay *i* and 0 otherwise. The model seeks a set of R_j values that minimize the difference between the predicted assay values (A_i^{pred}) and the measured ones (A_i^{meas}) for each chemical–concentration pair. A constrained least-squares minimization approach is used, where the function being minimized is:

Equation 2

$$\varepsilon^2 = \sum_i \left(A_i^{pred} - A_i^{meas}\right)^2 + penalty(R)$$

The term *penalty(R)* penalizes solutions that predict that many "receptors" are being simultaneously activated by the chemical:

Equation 3

$$penalty(R) = \alpha \frac{(x^{10})}{(x^{10} + 0.5^{10})}$$

Equation 4

$$\text{where } x = \sum_{j=1}^{N_{receptor}} R_j$$

The penalty term helps stabilize the solutions and is based on the assumption that it is unlikely that most chemicals will strongly and specifically interact with many dissimilar molecular targets (Judson et al., 2015). The model produces a response value (between 0 and 1)

for each "receptor" at each concentration. These results are summarized as the integral across the concentration range, expressed as area under the curve (AUC):

Equation 5

$$AUC_j = \frac{1}{N_{conc}} \sum_{i=1}^{N_{conc}} sign(slope) \times R_j(conc_i)$$

The biological response of greatest environmental concern is via antagonism of the AR pathway, which is also where most chemical activity is observed. Therefore, the AUC values were normalized to yield a value of 1 for the antagonist positive control. Hydroxyflutamide was used as the antagonist positive control, as recommended by the OECD (OECD, 2010a). The calibration curve plotting the relationship between AUC and activity concentration is given in Figure S1 (see tx6b00347_si_001.pdf in Kleinstreuer et al., 2017b). An AUC value of 0.1 corresponds to activity at ~100 µM; because this was the top tested concentration of most assays (except the Attagene assays), an AUC of ≥0.1 was considered to be positive. AUC values between 0.001 and 0.1 indicate very weak potential activity and were considered inconclusive. AUC values were rounded to 3 significant digits, and values below 0.001 were truncated and set to zero.

2.2.4 Cytotoxicity Filter

Each antagonist assay that measured suppression of protein production (Tox21_AR_BLA_Antagonist_ratio and Tox21_AR_LUC_MDAKB2_Antagonist) also produced viability readouts measuring cell death. These cytotoxicity assays were analyzed using the ToxCast data analysis pipeline, as described above, and the cytotoxicity AC50 was used as a threshold filter for antagonist activity in a pair-wise fashion. Any antagonist response with an AC50 greater than the cytotoxicity AC50 for that chemical-assay combination was discarded. Additional filtering approaches that were both more permissive (no exclusion) and more restrictive (exclusion of AC50 within 20% of the cytotoxicity AC50) were investigated, and the corresponding results for the AR pathway model (as well as the paired cytotoxicity data) are included in Supplemental Material (see tx6b00347_si_001.pdf in Kleinstreuer et al., 2017b). To ensure removal of overtly cytotoxic compounds while still permitting analysis of chemicals that may show antagonist behavior at test concentrations immediately preceding cytotoxicity, and to maintain consistency with the criteria for the reference chemical data extracted from the literature, the threshold approach was chosen for this analysis (Results for the AR Pathway Model) (see tx6b00347_si_005.pdf in Kleinstreuer et al., 2017b). To ensure removal of overtly

cytotoxic compounds while still permitting analysis of chemicals that may show antagonist behavior at test concentrations immediately preceding cytotoxicity, and to maintain consistency with the criteria for the reference chemical data extracted from the literature, the threshold approach was chosen for this analysis.

2.2.5 Cell Stress Flags

In a global analysis of the ToxCast dataset, it was observed that many different types of assays, both cell-based and cell-free, showed a rapid increase in the frequency of responses at concentrations corresponding to regions of cell stress/cytotoxicity (Judson et al., 2016). Potential nonselective assay hits attributed to cell stress were flagged using the distance between the logAC50(assay) and the median logAC50(cytotox), with respect to the global cytotoxicity median of the MAD (median absolute deviation) of the logAC50(cytotox) distributions across all chemicals. Details are given in (Judson et al., 2016). Briefly, for chemicals with two or more positive responses in assays measuring cytotoxicity or inhibition of proliferation, a "Z-score" was calculated for each AR pathway assay hit as:

Equation 6

$$Z(chemical, assay) = \frac{\text{logAC50}(chemical, assay) - median\left[\text{logAC50}(chemical, cytotox)\right]}{globalcytotoxicityMAD}$$

A large Z-score indicates an *in vitro* assay logAC50 at concentrations significantly below those causing cytotoxicity or inhibiting proliferation. Thus, a hit associated with this Z-score is unlikely to be caused by either cell-stress or cytotoxicity-related processes and is more likely to be associated with a target-selective mechanism, *e.g.*, interaction with the AR pathway.

2.2.6 Confirmation Flags

One of the *in vitro* transactivation antagonist assays, the Tox21 antagonist luciferase assay in the MDAKB2 cell line (**Figure 2-1, A11**) was run twice, with two different concentrations of the stimulatory ligand R1881. These data were used to help confirm whether chemical activity was specific (or not) to the AR pathway. The first time the assay was run, the concentration of the ligand R1881 was 10 nM (20x the EC50 of R1881), which resulted in saturation of the assay and a lack of activity for most chemicals, including known weak antagonists, based on the inability to displace the ligand, except for potent steroid antagonists (*e.g.*, flutamide-like compounds). The second time the assay was run with 0.5 nM R1881 and was sensitive to a wider range of chemicals. This second run, with the appropriate R1881 concentration, was included in the AR pathway model, and the data from the first run, with the

high R1881 concentration, were used in a paired fashion to examine compound specificity. A system of flags was applied to identify chemicals that may be activating the pathway through a non-receptor mediated mechanism. For true positives, it was expected that they would either be a hit in both runs, with a shift in the AC50 (from less to more potent), or they would be negative in the first run (when the assay was saturated with R1881) and a hit in the second run (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. Significance of the shift between AC50 values was determined using a bootstrapping approach across chemical replicates to define 95% confidence intervals as outlined below (see Watt, 2017), where overlapping confidence intervals were deemed a non-significant shift.

2.2.7 Uncertainty Quantification

All concentration-response curves used in the AR Pathway Model were analyzed using the R package toxboot v.0.1.0 (Watt, 2017). One thousand bootstrap replicates were generated for each curve using smooth non-parametric bootstrap resampling to get a distribution of fit parameters, model selections, and activity calls. Each bootstrap sample was grouped by chemical and analyzed using the AR Pathway Model with the same workflow as used to generate the point estimates, resulting in a distribution of 1000 AUC values per chemical. The inner 95% confidence interval for each chemical AUC value was calculated on this distribution using the quantile function from the R stats package (R Core Team, 2015) with probabilities 0.025 and 0.975 for the lower and upper threshold of the confidence interval, respectively.

2.2.8 Systematic Literature Review for Reference Chemical Identification

A targeted literature search was performed for quantitative data to refine previously published reference chemical lists and provide potency characterization for AR agonism/antagonism. This effort identified 158 potential reference chemicals with AR agonist or antagonist activity (or lack of activity) from the following international assay validation efforts run by:

- Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM, 2003)
- Organization for Economic Cooperation and Development (OECD, 2010a)
- U.S. EPA Endocrine Disruptor Screening Program (EDSP, (U.S. EPA, 2015d))

- European Union Reference Laboratory for Alternatives to Animal Testing (EURL ECVAM, ongoing)
- Korean Center for Validation of Alternative Methods (KoCVAM, ongoing)

Semi-automated literature searches were conducted for *in vitro* androgen activity data on the superset of chemicals (n=158) using PubMatrix (https://pubmatrix.irp.nia.nih.gov) and Scopus (http://www.scopus.com). Data from *in vitro* AR binding and TA assays were extracted from identified references and compiled into a single database (Supplemental File 1). Search keywords included “androgen”, “androgenic”, “anti-androgen”, and related terms, and are listed in Supporting Information: AR reference literature database and associated literature search keywords(see tx6b00347_si_004.xlsx in Kleinstreuer et al., 2017b).

Using a standardized ontology, the following information was recorded for each chemical-study combination:

- PubMed Identifier, author, year
- Chemical tested, Chemical Abstracts Service Registry Number (CASRN)
- Table or figure where results were reported
- Hit, response, response notes
- Half-maximal activity concentration (AC50 or IC50), standard error measurement, units
- Assay type (tissue, cell culture, or cell-free), tissue of origin (for cell culture), species of origin
- Receptor information, species source
- Reference androgen or anti-androgen
- Number of concentrations tested, highest concentration tested, units, incubation time
- Binding assays only: binding affinity, dissociation constant, relative binding affinity (RBA)
- TA assays only: agonist or antagonist mode, whether cytotoxicity was evaluated, extent of cytotoxicity observed (*i.e.* at IC50)
- TA assays only: reporter type, reporter construct, whether construct was native, transient, or stable

2.2.9 Reference Chemical Criteria

To establish reference chemical lists, high quality AR binding and transactivation (TA) data from the literature search were examined, filtered by conditions such as use of the full-

length receptor and concurrent measurement of cytotoxicity for antagonist-mode data (detailed in **Section 2.3.7**). To determine potency categories, all quantitative AR TA assay data identified were reported as AC50 or IC50 that could be converted to µM units and calculated mean, standard deviation, 95% confidence interval, and number of observations for each chemical. Binding data were used in a confirmatory fashion, where chemicals had to have positive binding results in the literature to be included as candidate positive agonist and antagonist reference chemicals. Based on the distribution of the results, agonist and antagonist reference chemical lists and potency categories were defined according to the following criteria.

Agonist:

- **Positives**: at least three TA experiments, of which at least 70% yielded positive results and at least one positive binding result
    - Strong: mean AC50 less than or equal to 0.1 µM
    - Moderate: mean AC50 greater than 0.1 µM and less than or equal to 1 µM
    - Weak: mean AC50 greater than 1 µM
- **Negatives**: at least three TA experiments yielding negative results and no TA experiments yielding positive results

Antagonist:

- **Positives**: at least three TA experiments, of which at least 70% yielded positive results that were not due to cytotoxicity and at least one positive binding result
    - Strong: mean IC50 less than or equal to 0.5 µM
    - Moderate: mean IC50 greater than 0.5 µM and less than or equal to 5 µM
    - Weak: mean IC50 greater than 5 µM and less than or equal to 25 µM
    - Very Weak: mean IC50 greater than 25 µM
- **Negatives**: at least two TA experiments yielding negative results, and no TA experiments yielding positive results

Chemicals with upper 95% confidence intervals that spanned potency categories were given combined category designations such as “Strong/Moderate” or “Moderate/Weak.”

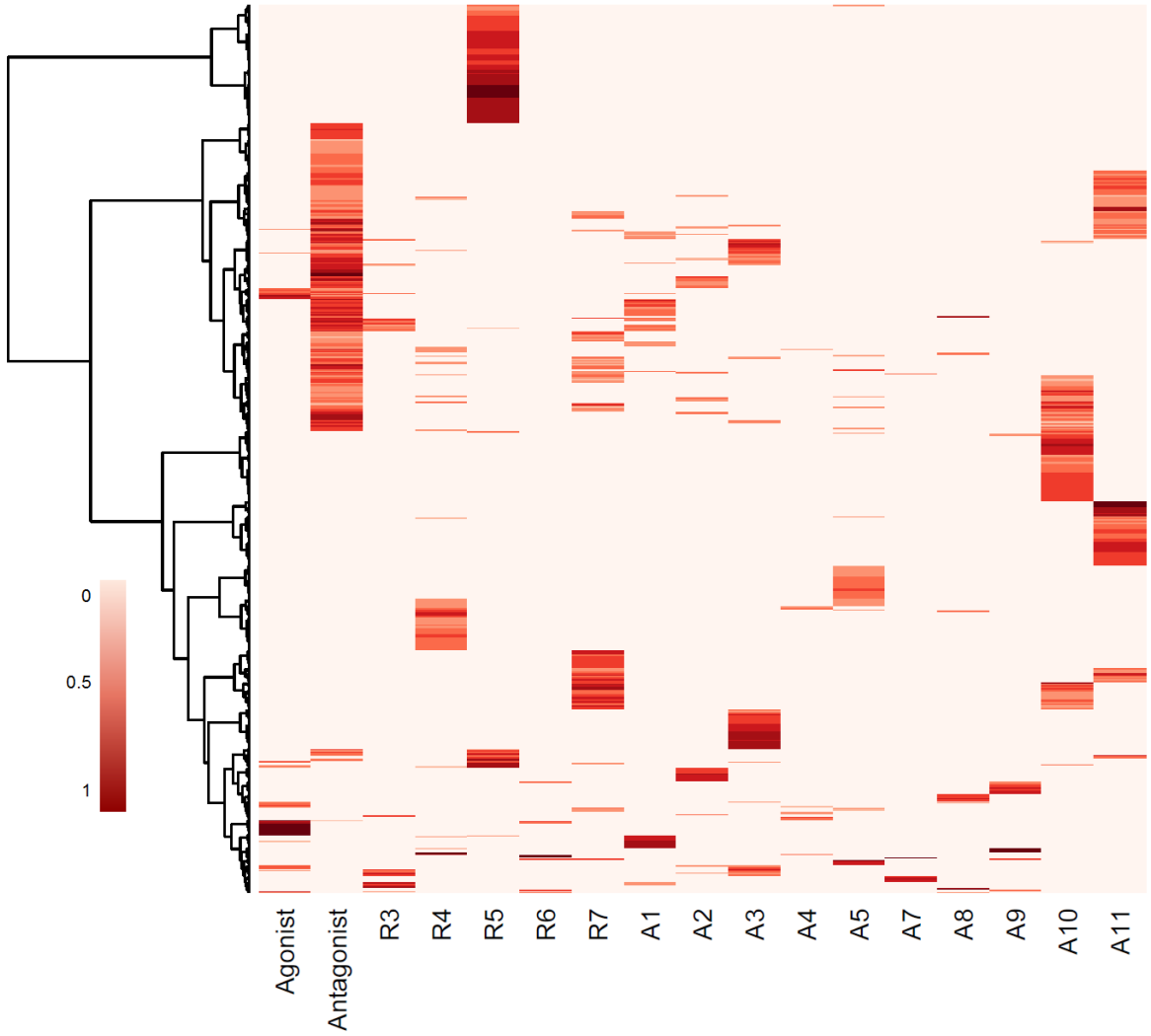
2.3 Results

2.3.1 Activity in the AR Pathway Model Across the ToxCast Library

Of the 1855 chemicals tested in all eleven Tox21/ToxCast AR assays, 1461 (78.8%) were predicted inactive in the AR pathway model, with both agonist (R1) and antagonist (R2) AUC

values below 0.001, while 220 chemicals (11.9%) were predicted to be either androgen agonists (n=33) or antagonists (n=192), with R1 or R2 AUC values > 0.1. Five of the 220 chemicals had significant activity in both agonist (R1) and antagonist (R2) pathways. The remaining 174 chemicals (9.4%) had inconclusive low AR pathway model scores with R1 or R2 AUC values of 0.001 to 0.1. These chemicals were generally weakly active in a small number of assays, and were usually also predicted by the model to be acting through interference pathways. Of the 1461 predicted inactive against the AR pathway, 1092 chemicals were inactive across all the assays, while 369 chemicals demonstrated activity associated with either assay interference or, less likely, weak activity only picked up in one technology type. **Figure 2-2** shows the distribution of AR model pathway scores across the ToxCast chemical library for 763 chemicals that were active in at least one AR pathway assay. Chemicals were either predicted to act via AR agonism (R1), antagonism (R2), biology-specific interference (R3-R7), or assay-specific interference (A1-A11). Figure S1 (see tx6b00347_si_001.pdf in Kleinstreuer et al., 2017b) is a calibration curve to help interpret AUC values in terms of pathway activity concentration, and File S2 (see tx6b00347_si_002.pdf in Kleinstreuer et al., 2017b) contains the results for each assay and the AR pathway model (AUC values and associated confidence intervals for agonism, antagonism, and interference) for all 1855 chemicals. Results of the AR pathway model with uncertainty bounds corresponding to 95% confidence intervals are plotted in Figure S2 (see tx6b00347_si_001.pdf in Kleinstreuer et al., 2017b).

Figure 2-2 Distribution of Model AUC Values Across 763 Chemicals



Heatmap shows the distribution of model area under the curve (AUC) values for 763 chemicals that were active in at least one AR pathway assay. The first two columns represent predictions for Agonist (R1) and Antagonist (R2) activity, and the remaining columns represent predicted assay- (A1-11) or biology- (R3-7) specific interference, corresponding to the pathway diagram in **Figure 2-1** and the interference pathways shown in **Table 2-1**. The darker red indicates higher AUC values, corresponding to more potent activity (scale: 0.001 to 1). Clustering was done using Ward's method.

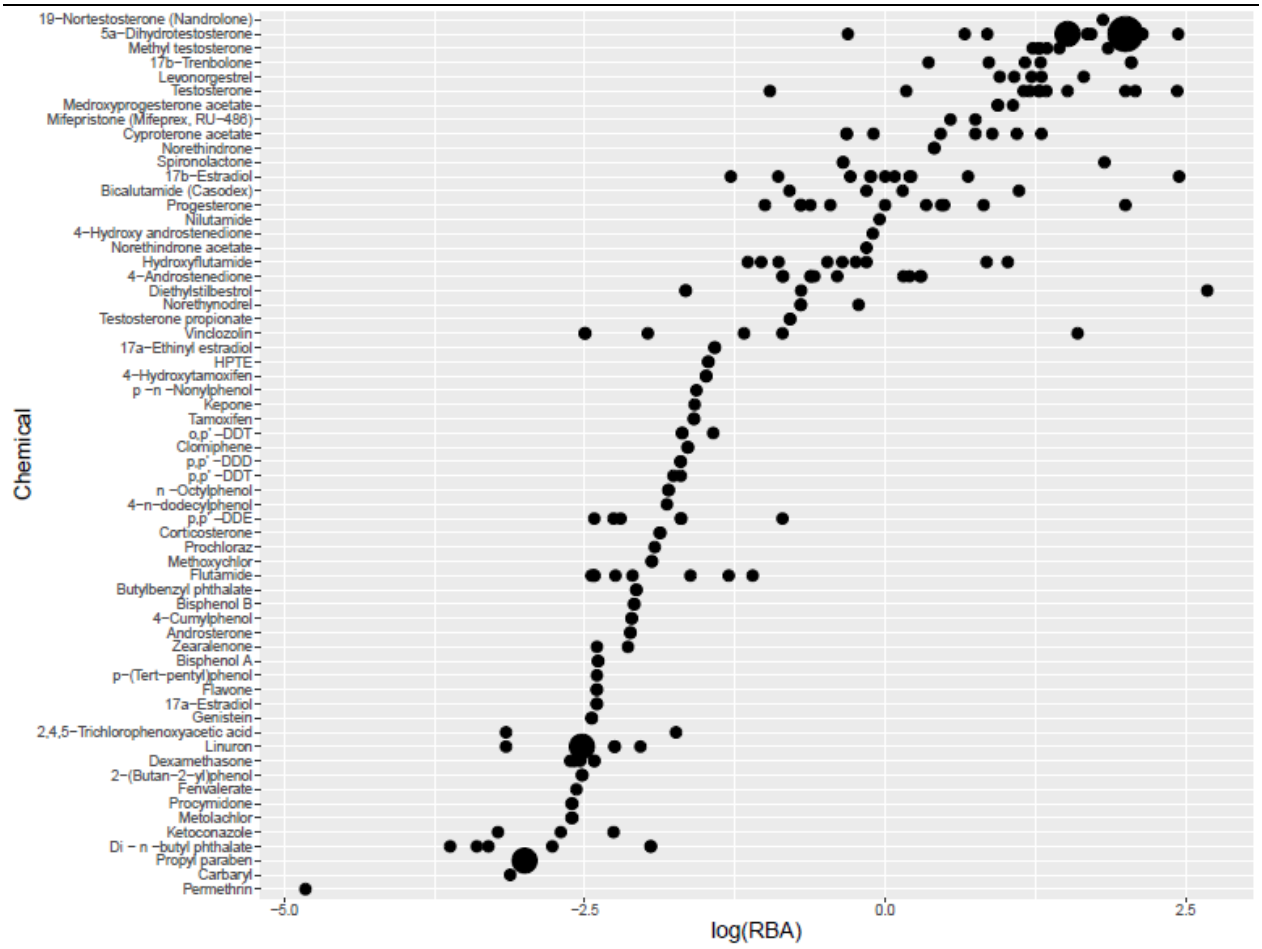
2.3.2 Systematic Literature Review Search Results

The targeted literature search for AR *in vitro* reference data yielded 4,795 chemical-study pairs across 379 publications. Experimental protocol details and chemical effects were recorded in a standardized manner in a structured data table (see tx6b00347_si_003.xlsx in Kleinstreuer et al., 2017b). AR binding data were identified for 111 chemicals, and the data were compiled from 1261 experiments reported in 166 publications. Commonly used assay platforms included cell culture, tissue preparations, and cell-free systems (see Figure S3a in tx6b00347_si_001.pdf as

part of Kleinstreuer et al., 2017b). The majority of the binding assays used full-length receptors (Figure S3b). A total of 26 species were represented among all binding assays, with most using human (39%) or rat (33%) receptors. The four most commonly used reference androgens were methyltrienolone (R1881; 475 assays, 41%), 5α-dihydrotestosterone (DHT; 400 assays, 34%), testosterone (203 assays, 17%), and mibolerone (84 assays, 7%).

Results from experiments using mutant receptors were excluded. Further analyses were conducted on data from binding assays using the full-length receptor and the ligand-binding domain (957 experiments on 95 chemicals). Multiple positive binding results with no negative results were reported for 38 chemicals. Atrazine, cycloheximide, and 2,4-dinitrophenol had multiple negative binding results and no positive results. There were 14 chemicals with only one positive binding result (and no negatives), and six chemicals with only one negative binding result (and no positives). The remaining 34 chemicals had both positive and negative binding results reported, although there was usually a clear majority of positive or negative results for each chemical. Results for binding affinity were reported in many different formats, the most common being relative binding affinity (RBA) or log RBA relative to a positive control. The relative binding data included R1881 (240 results), DHT (168 results), testosterone (97 results), and mibolerone (30 results). As an example, results for log RBA on 61 chemicals relative to the most common positive control compound, R1881, are shown in **Figure 2-3**.

Figure 2-3 AR Binding Affinities Relative to R1881 Reference



Abbreviations: AR = androgen receptor; R1881 = methyltrienolone; RBA = relative binding affinity. Chemicals are listed along the x-axis; y-axis represents the log10 (RBA). The size of the dot increases with the number of observations (range: 1-15). Relative binding affinity decreases moving from top to bottom, with a total of 61 chemicals described.

AR transactivation data were compiled for 160 chemicals (3534 experiments from 287 papers). While six different reporter types were used in the experiments, the majority of experiments used assays with a luciferase reporter (Figure S4a in tx6b00347_si_001.pdf as part of Kleinstreuer et al., 2017b). The use of a full-length receptor was also the most common (Figure S4b in tx6b00347_si_001.pdf as part of Kleinstreuer et al., 2017b). Many assays used a transiently transfected AR (46%) or stably integrated AR (39%), followed by native receptor expression (14%). Most TA assays used the human AR (93%), but receptors from a total of 14 species were represented among all assays in the database. The most common reference androgens were DHT (2262 assays, 64%), R1881 (703 assays, 20%), and testosterone (395 assays, 11%); the most common reference anti-androgens were flutamide (688 assays, 41%),

hydroxyflutamide (487 assays, 30%), bicalutamide (220 assays, 13%), and cyproterone acetate (192 assays, 11%).

Further analyses were conducted on data from the TA assays using the full-length receptor and the ligand-binding domain. Positive and negative TA assay results were reported for 2393 experiments on 133 chemicals. Results were subdivided into modes measuring agonist activity (1447 experiments, 60%) and antagonist activity (946 experiments, 40%). There were 13 chemicals with multiple positive agonist results (*i.e.*, increase in TA) and no antagonist results, all of which also had at least one negative result reported (*i.e.*, no agonist or antagonist activity). However, for most of these chemicals the number of positive agonist results far outnumbered the number of negative results which tended to occur in specific cell or receptor types and/or at low concentrations. There were 32 chemicals with multiple positive antagonist results (*i.e.*, decrease in TA) and no agonist results. All of these chemicals also had at least one negative TA result that tended to occur in specific cell types and/or at low concentrations. There were 17 chemicals with multiple negative (inactive for TA) results and no positive (agonist or antagonist) results. There were 15 chemicals with only one TA result in any category. The remaining 56 chemicals had a mix of positive (agonist and/or antagonist) and negative results. However, for most chemicals there was a clear majority of either agonist or antagonist results.

2.3.3 Potency of Transactivation Agonists

Positive results for TA agonist activity were reported in many different formats and with many different units, the most common being lowest effect level (LEL, 415 results, 49%) and half-maximal activity concentration (AC50, 406 results, 48%). All TA agonist results were converted into log µM units where possible, and the respective agonist potencies based on AC50s for each chemical were compared to negative results in terms of highest dose tested (HDT). The distribution of activity for chemicals with both positive (AC50s, colored dots) and negative (HDTs, black dots) results is shown in **Figure 2-4**.

Figure 2-4 Comparing AR Transactivation Agonist Results

Abbreviations: AC50 = half-maximal activity concentration; AR = androgen receptor; HDT = highest dose tested. Chemicals are listed along the x-axes and the log transformed doses along the y-axis. The colored dots represent positive results in log10 (AC50), and the black dots represent negative results in log10 (HDT). The size of the dot increases with the number of observations (range: 1-79). Agonist potency decreases moving from bottom to top, with a total of 40 chemicals described.

2.3.4 Potency of Transactivation Antagonists

AR TA antagonist potency was evaluated using only data from experiments that concurrently measured cytotoxicity (520 experiments [55%] representing 105 chemicals), with clearly stated acceptance criteria (*e.g.*, <20% loss of viability). Positive results for antagonist activity were reported in many different formats and with many different units, the most common being half-maximal inhibition activity concentration (IC50, 224 results, 64%) and LEL (114 results, 33%). All TA antagonist results were converted to log µM units where possible, and the respective antagonist potencies based on IC50 were compared to the negative results in terms of HDT. The distribution of activity for chemicals with both positive (IC50s, colored dots) and negative (HDTs, black dots) results is shown in **Figure 2-5**.

Figure 2-5 Comparing AR Transactivation Antagonist Results

Abbreviations: AR = androgen receptor; HDT = highest dose tested; IC50 = half-maximal inhibitory concentration. Chemicals are listed along the x-axes and the log transformed doses along the y-axis. The colored dots represent positive results in log10 (IC50), and the black dots represent negative results in log10 (HDT). The size of the dot increases with the number of observations (range: 1-21). Antagonist potency decreases moving from bottom to top, with a total of 54 chemicals described.

2.3.5 AR Pathway *In Vitro* Reference Chemicals

Based on the criteria outlined in the Methods for reproducibility and consistency of response, 37 reference chemicals for AR agonism and 28 reference chemicals for AR antagonism were identified (see **Table 2-2**). Initial reference chemical categorizations included strong, moderate, weak and very weak agonists and antagonists, and negative categorizations, all of which were based exclusively on the curated results from the published literature and did not include any information from the ToxCast or Tox21 assays. There were 11 chemicals that fulfilled reference criteria for both agonism and antagonism, usually as a positive reference in one and a negative reference in the other. Cyproterone acetate was classified as both a weak agonist and a moderate antagonist based on multiple literature results showing selective androgen receptor modulation with agonist and antagonist effects. Of the 54 total reference chemicals, 46 were also tested in ToxCast/Tox21 and could be used for performance-based model validation.

Table 2-2 AR Pathway In Vitro Reference Chemicals

CASRN	Chemical Name	Agonist Potency Category	Antagonist Potency Category	In ToxCast 10/2015 Release
52806-53-8	Hydroxyflutamide	NA	Strong	yes
90357-06-5	Bicalutamide	NA	Strong	yes
122-14-5	Fenitrothion	NA	Strong	yes
84371-65-3	Mifepristone	NA	Strong/Moderate	yes
52-01-7	Spironolactone	NA	Strong/Moderate	yes
63612-50-0	Nilutamide	Negative	Moderate	yes
427-51-0	Cyproterone acetate	Weak	Moderate	yes
80-05-7	Bisphenol A	NA	Moderate/Weak	yes
330-55-2	Linuron	NA	Moderate/Weak	yes
50471-44-8	Vinclozolin	NA	Moderate/Weak	yes
13311-84-7	Flutamide	Negative	Moderate/Weak	yes
67747-09-5	Prochloraz	Negative	Moderate/Weak	yes
140-66-9	4-tert-Octylphenol	NA	Weak	yes
72-43-5	Methoxychlor	NA	Weak	yes
72-55-9	p,p' –DDE	NA	Weak	yes
60207-90-1	Propiconazole	NA	Weak	yes
17924-92-4	Zearalenone	NA	Weak	yes
789-02-6	o,p' –DDT	Negative	Weak	yes

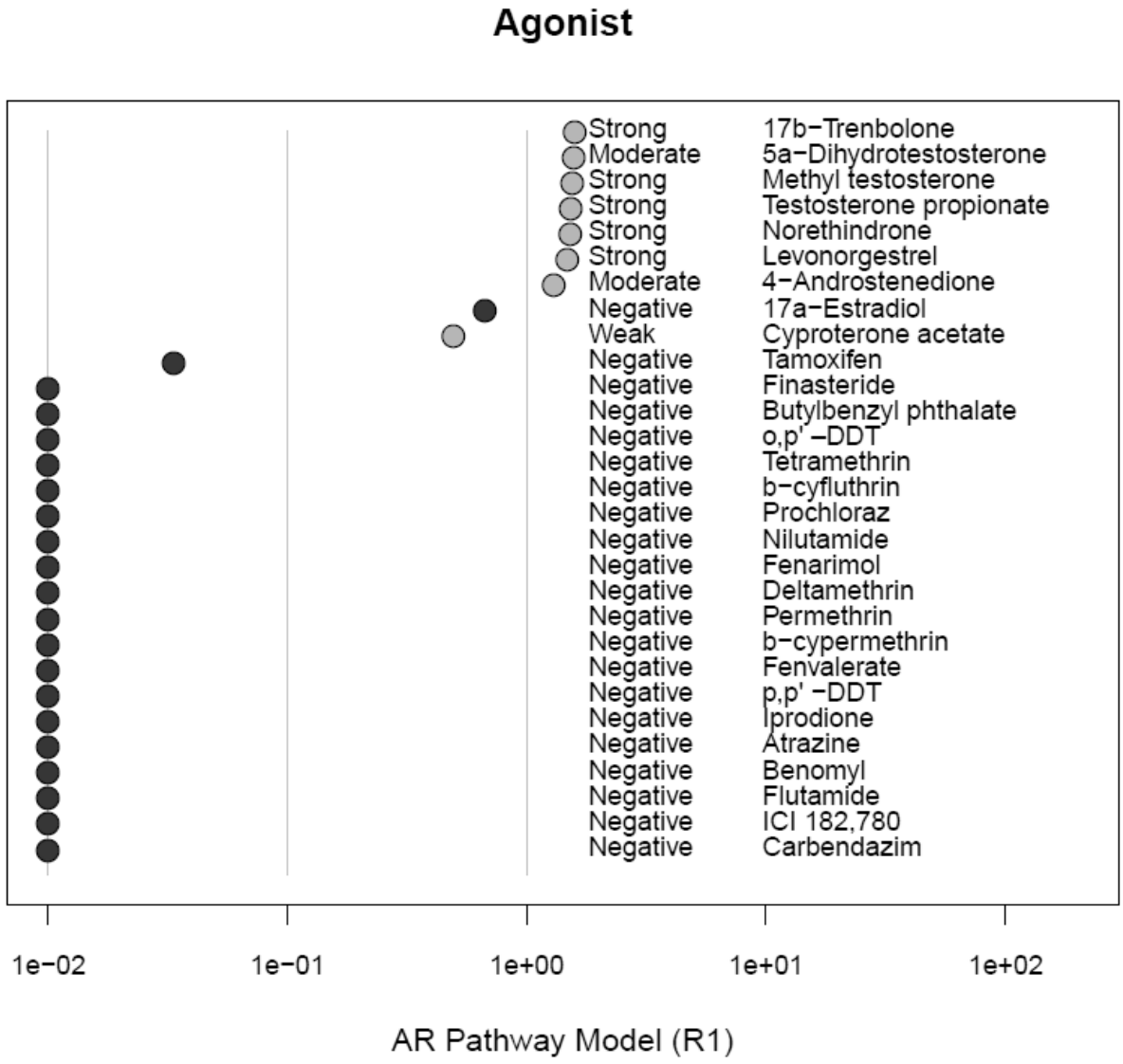
CASRN	Chemical Name	Agonist Potency Category	Antagonist Potency Category	In ToxCast 10/2015 Release
32809-16-8	Procymidone	NA	Very Weak	yes
60168-88-9	Fenarimol	Negative	Very Weak	yes
58-18-4	Methyl testosterone	Strong	Negative	yes
58-22-0	Testosterone	Strong	Negative	Propionate form
63-05-8	4-Androstenedione	Moderate	Negative	yes
1912-24-9	Atrazine	Negative	Negative	yes
52918-63-5	Deltamethrin	Negative	Negative	yes
486-66-8	Daidzein	NA	Negative	yes
16752-77-5	Methomyl	NA	Negative	yes
122-34-9	Simazine	NA	Negative	yes
10161-33-8	17b-Trenbolone	Strong	NA	yes
797-63-7	Levonorgestrel	Strong	NA	yes
965-93-5	Methyltrienolone (R1881)	Strong	NA	no
68-22-4	Norethindrone	Strong	NA	yes
51-98-9	Norethindrone acetate	Strong	NA	no
76-43-7	Fluoxymestrone	Strong/Moderate	NA	no
434-22-0	19-Nortestosterone	Moderate	NA	no
521-18-6	5a-Dihydrotestosterone	Moderate	NA	yes
10418-03-8	Stanozolol	Moderate	NA	no
71-58-9	Medroxyprogesterone acetate	Moderate/Weak	NA	no
68-23-5	Norethynodrel	Moderate/Weak	NA	no
57-91-0	17a-Estradiol	Negative	NA	yes
68359-37-5	b-cyfluthrin	Negative	NA	yes
52315-07-8	b-cypermethrin	Negative	NA	yes
17804-35-2	Benomyl	Negative	NA	yes
85-68-7	Butylbenzyl phthalate	Negative	NA	yes
10605-21-7	Carbendazim	Negative	NA	yes
51630-58-1	Fenvalerate	Negative	NA	yes
98319-26-7	Finasteride	Negative	NA	yes
129453-61-8	ICI 182,780	Negative	NA	yes
36734-19-7	Iprodione	Negative	NA	yes
50-29-3	p,p'-DDT	Negative	NA	yes
52645-53-1	Permethrin	Negative	NA	yes
501-36-0	Resveratrol	Negative	NA	no
10540-29-1	Tamoxifen	Negative	NA	yes
7696-12-0	Tetramethrin	Negative	NA	yes

2.3.6 AR Pathway Model Performance

The predicted activity from the AR pathway model was compared with the curated results for the 46 reference chemicals identified in the literature review and were also tested in ToxCast and Tox21. The results of the model predictions are shown in **Figure 2-6** (29 agonist reference chemicals) and **Figure 2-7** (28 antagonist reference chemicals). An AR pathway model score greater than 0.1 (~activity at concentrations less than 100 µM) was considered positive,

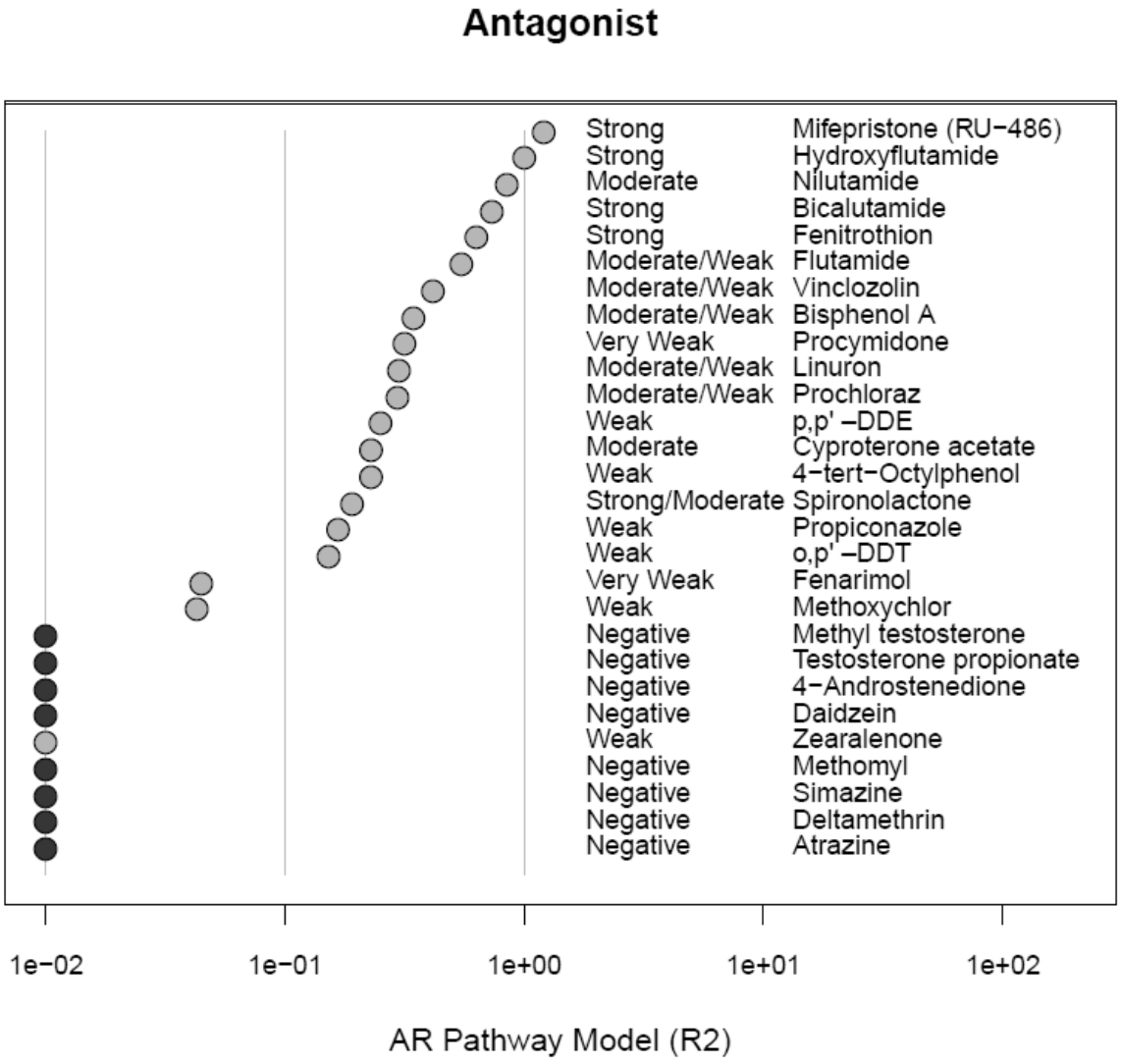
with highER pathway model scores corresponding to stronger potency. With respect to the AR agonist reference chemicals, 17α-estradiol was the only false positive, and there were no false negatives. One negative agonist reference chemical, tamoxifen, had an inconclusive agonist AUC (R1) score of 0.0335. Following the example of (Browne et al., 2015), we evaluated the model performance two ways. If inconclusive scores were considered positive, the AR pathway model had a balanced accuracy of 95.2% (100% sensitivity and 90.5% specificity) against the agonist reference chemicals, and if inconclusive results were excluded, the balanced accuracy was 97.5% (100% sensitivity and 95% specificity). Two of the antagonist reference chemicals, methoxychlor (weak potency) and fenarimol (very weak), had antagonist AUC (R2) scores in the inconclusive range, of 0.0429 and 0.0446, respectively. Zearalenone, categorized in the literature review as a weak antagonist, was a false negative, and there were no false positives. The model predicted that zearalenone was causing assay interference through R7 (corresponding to key event node N5 in **Figure 2-1**), because it hit both Tox21 antagonist assays but none of the upstream assays in the antagonist pathway (binding or coregulator recruitment). The AR pathway model had 97.5% balanced accuracy (95% sensitivity and 100% specificity) when predicting the antagonist reference chemicals and counting the inconclusive results as positive, or 97.2% balanced accuracy (94.4% sensitivity and 100% specificity) if the inconclusive chemicals were excluded. Examples of the concentration-response curves for several reference chemicals are shown in **Figure 2-8**.

Figure 2-6 AR Pathway Model Results for Reference Chemicals – Agonist



Reference chemicals and associated potency categories (from the literature search) are listed along the y-axes and the AR pathway model AUC score for agonism (R1), along the x-axes. Gray dots represent positive reference chemicals and black dots represent negative reference chemicals. AR pathway model scores below 0.01 were truncated at 0.01 for plotting purposes. There was one false positive for agonism (17α-estradiol), and one negative agonist reference chemical with an inconclusive model score (tamoxifen).

Figure 2-7 AR Pathway Model Results for Reference Chemicals – Antagonist



Reference chemicals and associated potency categories (from the literature search) are listed along the y-axes and the AR pathway model AUC score for antagonism (R2), along the x-axes. Gray dots represent positive reference chemicals and black dots represent negative reference chemicals. AR pathway model scores below 0.01 were truncated at 0.01 for plotting purposes. The initial false negative for antagonism (zearalenone) was confirmed as a potential true positive by the antagonist confirmation assay (Tox21_AR_LUC_MDAKB2_Antagonist-confirmation). Two antagonist reference chemicals had AUC scores in the inconclusive region.

Figure 2-8 Concentration Response Curves and AR Pathway Model Results

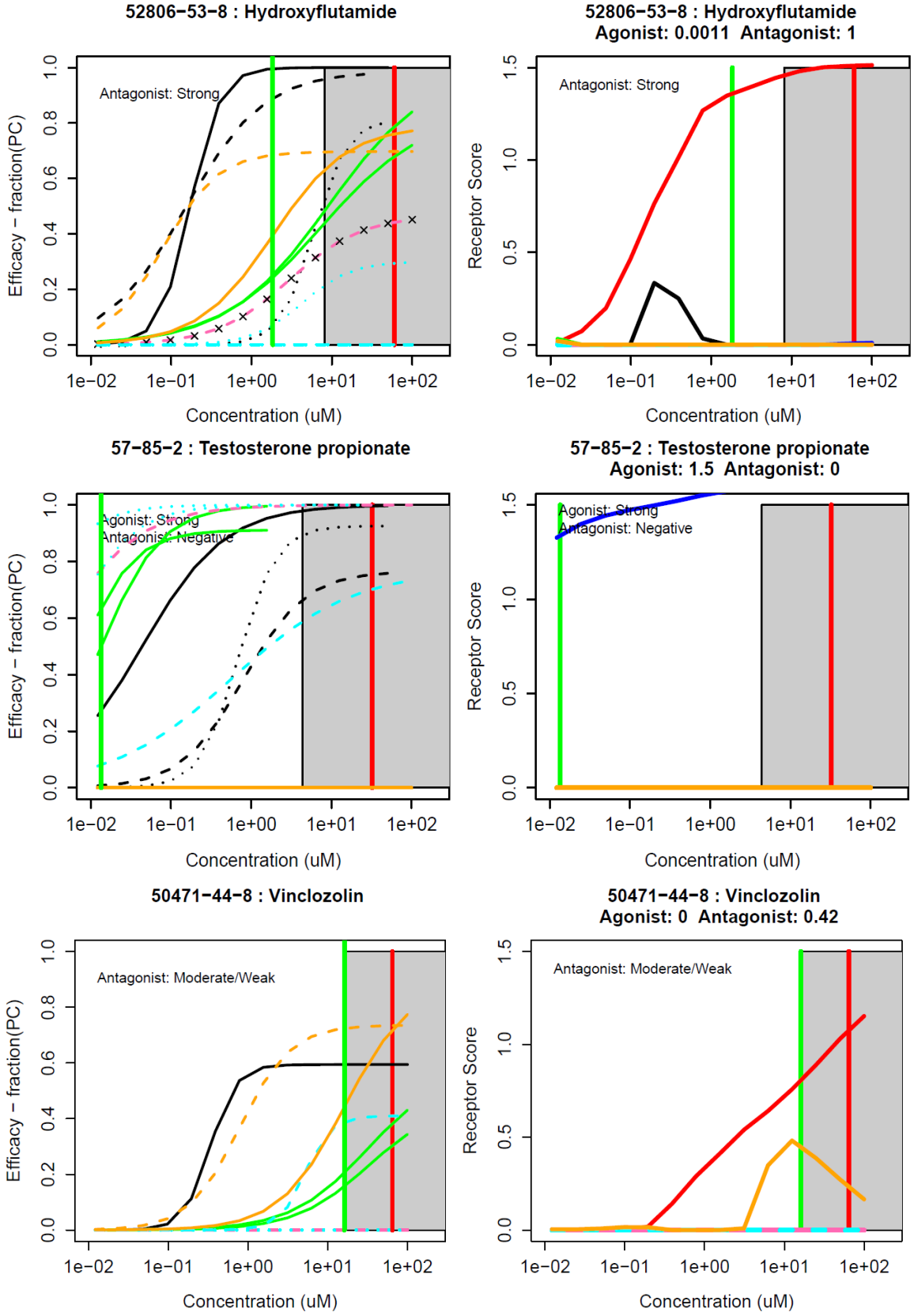


Figure 2-8 Concentration Response Curves and AR Pathway Model Results (Continued)

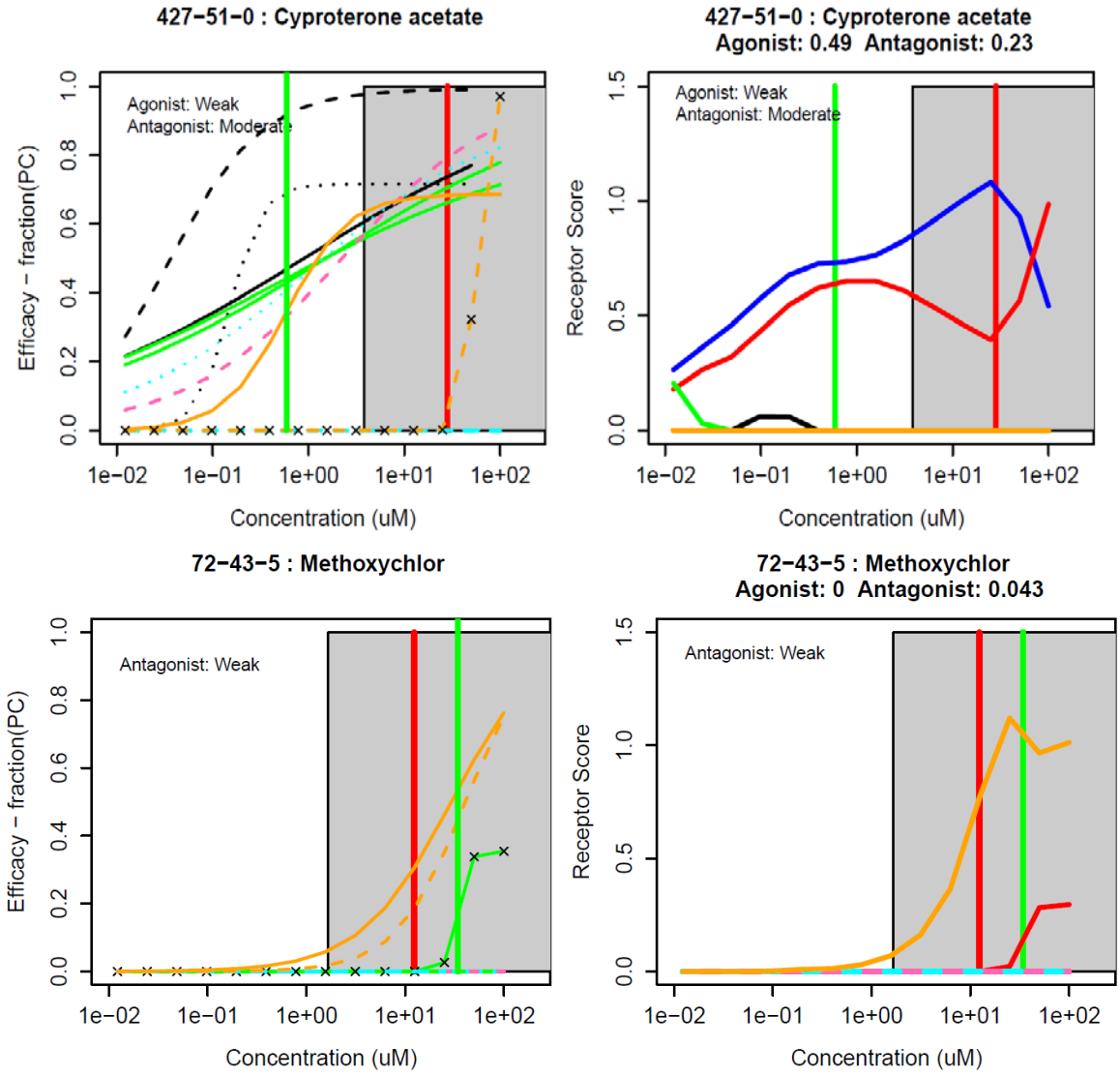
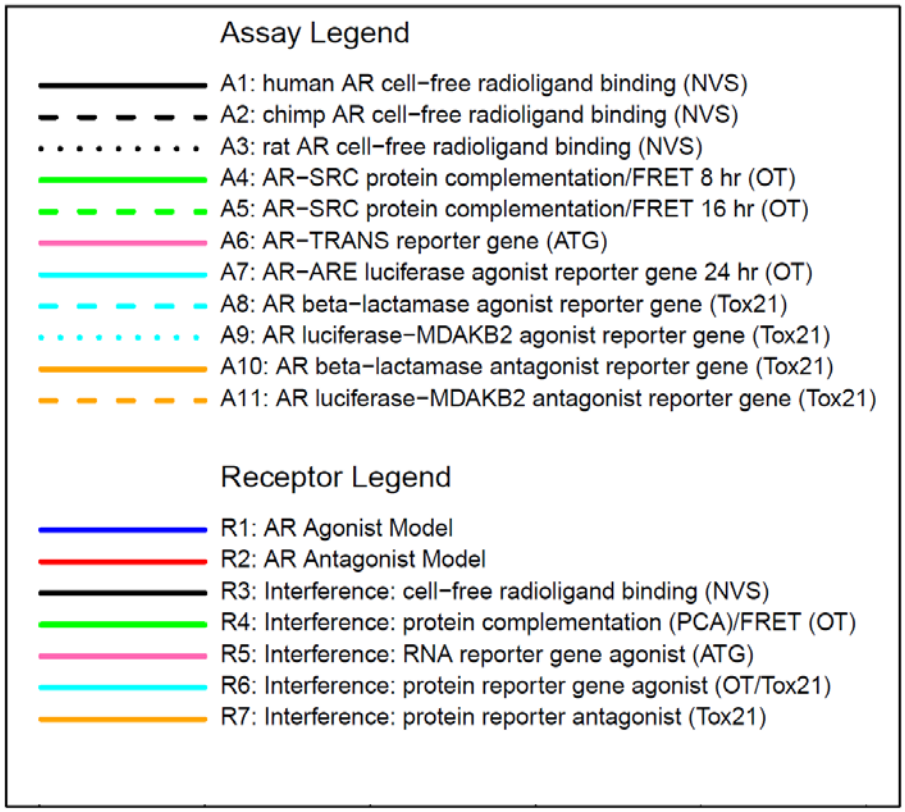


Figure 2-8 Concentration Response Curves and AR Pathway Model Results (Continued)



Results of the AR Pathway model for selected reference chemicals. For each chemical, the left-hand panel shows the concentration response data for the 11 *in vitro* assays, colored by assay group as defined in the legend. The right-hand panel shows the magnitude of the modeled "receptor" responses, where the agonist pathway (R1) is in blue and the antagonist pathway (R2) is in red, and the other interference pathways (R3-R7) are colored as defined in the legend. Model AUC values are displayed below the chemical name and literature-based reference classifications are displayed in the plot. The median cytotoxic concentration for each chemical is indicated by a vertical red line, and the cytotoxicity region (representing 3 median absolute deviations) is indicated by the gray shaded region. A green horizontal bar indicates the median- AC50 of the active assays. Similar plots for all chemicals are given in Supplemental File 3.

2.3.7 Distinguishing Antagonism and Cell Stress

The Z-score provides a measure of proximity (how many median absolute deviations) for the activity of a chemical in a particular assay relative to the median concentration for that chemical across 33 viability and proliferation inhibition assays in the ToxCast library (Judson et al., 2016). Z-scores for every chemical-assay combination in the AR pathway model are reported (see tx6b00347_si_004.xlsx in Kleinstreuer et al., 2017b). A chemical-assay hit with a high Z-score (>3) indicates that AR-related activity occurred at concentrations far below the cytotoxicity threshold and suggests that there was no evidence of cell stress. These hits are more likely to be associated with specific biomolecular interactions with the intended biological target

than the assays are designed to measure. Examples of chemicals with high AUC values for AR antagonism and high average Z-scores include hydroxyflutamide, nilutamide, vinclozolin, linuron, spironolactone, and apigenin. Hits with low Z-scores (activity concentrations in the cell stress/cytotoxicity region) are more likely to be associated with an interference process than hits with high Z-scores. However, due to variable concentration spacing, quantitative uncertainties in AC50 values, and differential sensitivity among cell types, the Z-score cannot be used as a definitive filter and is instead valuable to provide context on the potential specificity of the results.

2.3.8 Antagonism Confirmation Assay Results

The confirmation assay data from the Tox21_MDAKB2_Luc_Antagonist assay with two different concentrations of stimulating ligand (R1881) provided additional insight into chemicals that were potentially acting via a non-receptor mediated mechanism (*e.g.*, generalized cell stress or cytotoxicity) versus chemicals that appeared to be acting via the AR ligand-binding domain. When considering these data, the one "false negative" reference chemical, zearalenone, displayed behavior indicative of true weak antagonist potential, where it was active in both screens and exhibited a potency shift in the expected direction, although the shift was flagged as not significant due to overlapping confidence intervals around the AC50 values. It is worth noting that zearalenone is predicted to be a fairly potent ER agonist (AUC model score of 0.71; (Browne et al., 2015)). There were 128 chemicals that were only active when the assay was stimulated with the lower R1881 concentration; behavior consistent with the potential for weak antagonism. There were 57 chemicals that were active in both runs, and exhibited the expected potency shift with non-overlapping AC50 confidence intervals. Most of these were predicted as true antagonists by the model, including positive antagonist reference chemicals triclosan, and bisphenols A/B/AF. Others (*e.g.* endosulfan sulfate, dinoseb, fenoxycarb) had inconclusive model scores or were predicted to act via interference pathways, such as suppression of protein production (R7, node N5), because they did not hit the binding or coregulator recruitment assays. There were 128 chemicals that were active in both runs and exhibited the expected potency shift, but had overlapping AC50 confidence intervals. There were 65 chemicals that were active in both runs, but exhibited a potency shift in the opposite direction (*i.e.*, more potent in the assay with a higher R1881 concentration), and 22 chemicals that were only active in the assay with a higher R1881 concentration, and inactive in the other run. These included potently cytotoxic compounds (*e.g.* gentian violet), cytostatic compounds (*e.g.* cycloheximide), organometallics, and pesticides. There were 1455 chemicals that were inactive in both runs, most of which were

also inactive against the AR pathway model. Each category of chemical activity is designated by the corresponding "Tox21 Antagonist Confirmation Assay Flag" in File S2 (see tx6b00347_si_004.xlsx in Kleinstreuer et al., 2017b).

2.3.9 Antagonist Activity Confidence Scoring

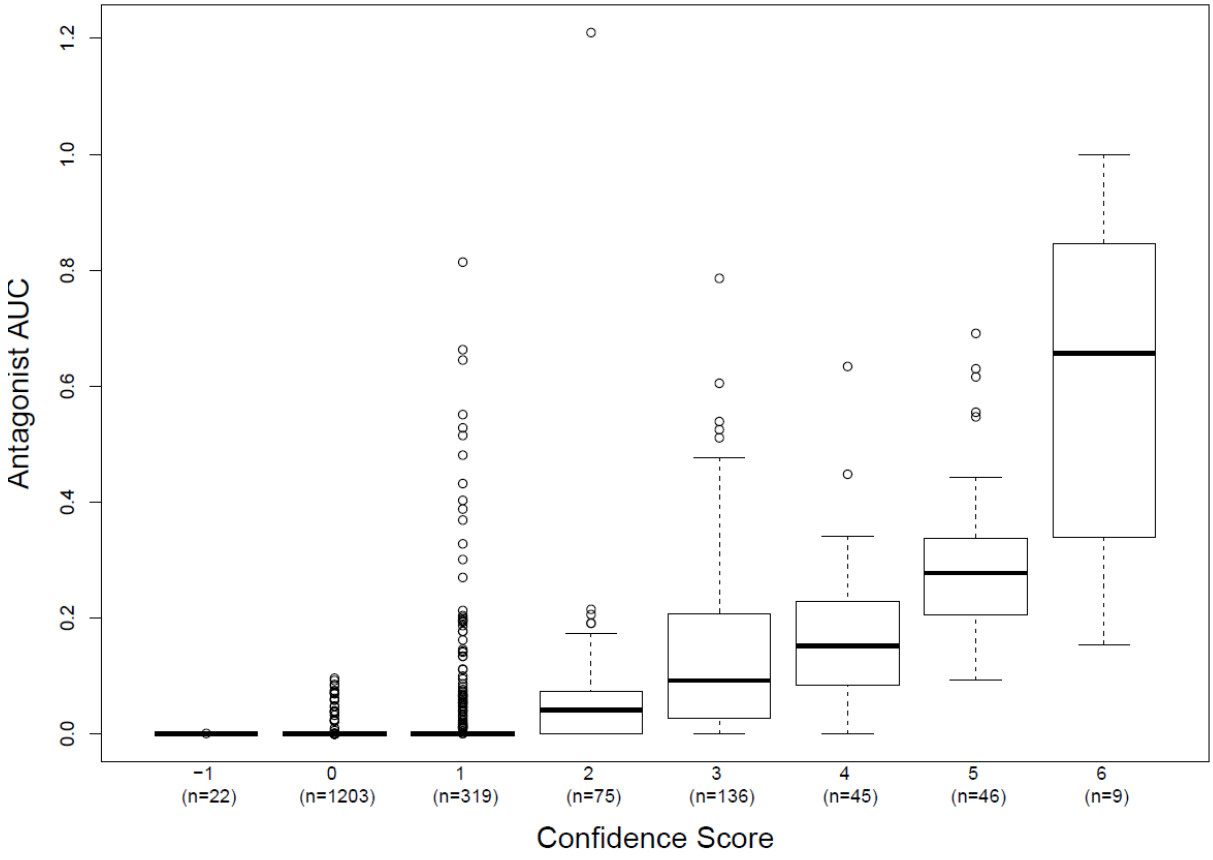
The AR pathway model AUC scores, cytotoxicity information, and confirmation flags were used to inform a simple summary score for each chemical that translates into confidence that the observed activity is via the AR pathway. The schema for assigning confidence scores is shown in **Table 2-3**. The default score for inactive chemicals was set to zero. Chemicals with high antagonist (R2) AR pathway model AUC scores were assigned higher confidence scores, as were those chemicals that were active in the concentration region prior to cytotoxicity (high average Z-scores across the eleven assays). For potential antagonists, those exhibiting the expected potency shift in the confirmation assays were assigned higher confidence scores, while those with data indicating that the chemical was not acting via the receptor were assigned negative confidence scores. The confidence scores from each source were then summed to provide an overall confidence score to facilitate chemical prioritization in a manner that incorporates all the contributing data streams. The positive antagonist reference chemicals all had positive activity confidence scores. All 192 chemicals with R2 AUC values above 0.1 also had positive activity confidence scores, although there were 36 chemicals with low confidence scores (≤2) that were flagged based on the confirmation assay data and may be false positives. Out of the 170 chemicals with inconclusive model antagonist AUC scores (between 0.001 and 0.1), 144 chemicals had positive confidence scores, and 61 of these had high confidence scores (≥3). There were 294 chemicals with positive confidence scores that were negative in the AR pathway model (R2 AUC values of 0), some of which were predicted agonists, and most of which were predicted to act via interference receptors. Of those 294 model negative chemicals, there were 26 chemicals with confidence scores ≥3, which may have been missed by the model and should be examined further for potential antagonist activity. There were 1225 chemicals with activity confidence scores ≤0, meaning that they were either inactive, caused technology-specific interference, or displayed activity indicative of a non-AR mediated response (usually cytotoxicity driven). The distribution of AR pathway model antagonist AUC values across the different confidence scoring bins is shown in **Figure 2-9**.

Table 2-3 Schema for Antagonist Activity Confidence Scoring

Source	Criteria	Confidence Score Contribution ǂ
AR Pathway Model	AUC.R2>0.1	2
	0.1>AUC.R2>0.001	1
Cell Stress/ Cytotoxicity Flag	average Z-score>3	1
Confirmation Assay Data	True antagonist shift (Hit/Hit)	3
	True antagonist shift (No hit/Hit)	2
	FLAG: True antagonist shift, but CI overlap	1
	FLAG: Wrong direction shift (Hit/Hit)	-1
	FLAG: Wrong direction (Hit/No hit)	-1

ǂ Contributions from the three source categories are summed to provide an overall antagonist activity confidence score ranging from -1 to 6.

Figure 2-9 AR Pathway Model Antagonist AUC Distribution by Confidence Score



2.3.10 Comparison with EPA EDSP Tier 1 AR Binding Assay

The current AR pathway model is intended as a potential alternative for the existing EDSP Tier 1 AR binding assay. There are a total of 101 chemicals with data from the EDSP

Tier 1 AR binding assay and data from the current model. Tier 1 AR binding data came from two sources: the ICCVAM assay validation document (ICCVAM, 2003) and results from the first set of test orders issued by the U.S. EPA EDSP, referred to as "List 1" (U.S. EPA, 2016a). The Tier 1 assay measured binding rather than agonism or antagonism, so for comparison, we called a chemical model-positive if the maximum of the agonist or antagonist AUC values was ≥0.1, negative if the maximum was <0.001, and inconclusive if the maximum AUC was between 0.001 and 0.1. For ICCVAM, RBA values were reported ($IC50_{R1881}$ x 100/ $IC50_{test\ chemical}$), while for the List 1 chemicals both RBA and IC50 values were reported. To facilitate comparison, we developed a calibration curve using the List 1 chemicals, based on an observed linear relationship between log(IC50) and log(RBA), which allowed us to estimate IC50 values from RBAs for ICCVAM chemicals. A linear model (see Figure S5 in tx6b00347_si_001.pdf as part of Kleinstreuer et al., 2017b) between the two yielded an RMSE of 0.25 and R2 of 0.84, with both slope and intercept of approximately -1. All data on these comparisons is given in File S5 (see tx6b00347_si_007.xlsx in Kleinstreuer et al., 2017b). Of the 101 chemicals, 7 had equivocal calls in the Tier 1 data and 6 had inconclusive AR pathway model scores (1 chemical overlap), yielding 89 chemicals with comparable data.

Of the 39 List 1 chemicals with both List 1 AR binding assay data and AR model scores, 2 were positive in both, 6 were model positive and Tier 1 negative, 7 were model negative and Tier 1 positive, and 24 were negative in both. The List 1 positive and AR model negative chemicals are 2-phenylphenol, carbaryl, diazinon, dichlobenil, metolachlor, myclobutanil and phosmet. With the exception of phosmet, the IC50 values for these chemicals are well over 100 µM, and so would be expected to be negative in the model, as the top tested concentrations in ToxCast and Tox21 were ≤100 µM. The IC50 for phosmet for binding was 10 µM in Tier 1, in close agreement with the chimp AR binding assay (A2) AC50 of 18 µM in the AR model data; however, the human and rat binding assays did not yield positive hit calls when tested to 40 µM. Phosmet was negative in the AR model data transactivation assays in agreement with a previous published report (Kojima et al., 2004). The model positive / List 1 negative chemicals are abamectin, captan, chlorothalonil, folpet, MGK-264 and propargite. All of these are classified as antagonists in the model, with AUC antagonist values ranging from 0.09 to 0.48. However, all of these chemicals are flagged as potential false positives using the antagonist confirmation assay data, based on either a potency shift in the wrong direction (abamectin, chlorothalonil, folpet, propargite), or no significant shift (captan, MGK-264). The two chemicals called positive in both approaches are propiconazole and tebuconazole. Both of these were classified in the model as

antagonists, and both had significant shifts in the correct direction in the confirmation antagonist assay. In summary, the model positive / List 1 negative chemicals are likely all false positives in the model, but this was detected using the confirmation assay. The model negative, List 1 positive chemicals are all so weak that they would not be detected by the HT assays used in the model because of the upper testing concentration of 100 µM, with the possible exception of phosmet for which no clear call can be made. The model results, including uncertainty bounds, for all the List 1 chemicals are shown in Figure S6 (see tx6b00347_si_001.pdf in Kleinstreuer et al., 2017b).

There were 51 chemicals with data from the model and the ICCVAM validation set for the Tier 1 AR binding assay (atrazine was also on List 1). Of these, 22 were positive in both, 9 were model positive and Tier 1 negative, 1 was model negative and Tier 1 positive, and 19 were negative in both. This yields a sensitivity and specificity of 0.96 and 0.68 respectively. The single ICCVAM chemical that was model negative and Tier 1 positive was atrazine with a RBA of 0.0018 yielding a modeled IC50 of 53 µM, near the upper limit of HT testing. Atrazine was also evaluated in the List 1 process, using literature data which yielded equivocal results, but an ultimate List 1 call of inactive. The 10 model positive, Tier 1 negative chemicals are 17α-estradiol, 4-cumylphenol, apigenin, bisphenol B, clomiphene citrate, cycloheximide, fulvestrant, meso-hexestrol, oxazepam, and reserpine. All of these were classified as antagonists, except for 17α-estradiol and oxazepam, although the former had an agonist AUC (R1) of 0.67 and antagonist AUC (R2) of 0.09. Of these chemicals, 4 had a significant shift in the correct direction in the antagonist confirmation assay (17α-estradiol, 4-cumylphenol, apigenin, bisphenol B), while 3 had the shift in the correct direction, but with overlapping confidence intervals (clomiphene citrate, meso-hexestrol, reserpine). Cycloheximide had a shift in the wrong direction. Fulvestrant and oxazepam also had significant activity in interference channels, so are likely active due to assay interference. In summary, among these model positive / Tier 1 negative chemicals, the model data support true activity for 17α-estradiol (mixed agonist / antagonist) and 4-cumylphenol, apigenin, bisphenol B (antagonists). Note that these are all estrogen receptor agonists. Additionally, in the ICCVAM listing, these are noted as “presumed negative”. The remaining 6 chemicals show evidence for false-positive activity in the model. The model results, including uncertainty bounds, for all the ICCVAM chemicals are shown in Figure S7 (see tx6b00347_si_001.pdf in Kleinstreuer et al., 2017b).

2.4 Discussion

Implementation of HT in ToxCast and Tox21 has generated high-quality quantitative data on thousands of chemicals and potential environmental pollutants. The inclusion of orthogonal assays that query key events along a biological pathway in multiple ways has produced novel and robust hazard screening capabilities. A similar mechanistically-based pathway model to the one presented here for AR is already being used by the U.S. EPA EDSP to identify potential endocrine disruptors acting via estrogen agonism (U.S. EPA, 2015f). The ER pathway model was validated against a well-defined set of reference chemicals (Browne et al., 2015; Judson et al., 2015), which heretofore was not possible for the AR pathway due to the lack of a well-characterized reference chemical set. We now have reported the results from a comprehensive literature review on potential AR reference chemicals, and used the resulting set to evaluate the performance of the AR pathway model based on eleven Tox21/ToxCast assays.

Every assay has inherent limitations driven by technological specifications and applicability domain. A biological pathway-based approach that integrates multiple assays mapping to key upstream and downstream events provides a weight of evidence for the true potential of a chemical to activate or repress signaling, in this case via the AR. This type of approach compensates for the individual shortcomings of any one assay. For example, there were 105 chemicals that were predicted to act through a receptor interference pathway (A7, **Figure 2-1**) because they were only active in the OT_AR_ARELUC_AG_1440 luciferase reporter gene assay measuring downstream transcriptional activation via protein production. None of these chemicals are known to be AR agonists, so it is likely that their activity was correctly flagged as interference and may have been a result of non-specific transcriptional effects. There are also a large number of chemicals that produced hits in one or more of the cell-free receptor binding assays, and were therefore predicted as A1-A3 or R3. Many of these chemicals are surfactants, indicating that these chemicals may have reacted with the proteins or otherwise caused denaturation, leading to displacement of the radioligand and a binding-like signal.

Cytotoxicity and response specificity were further considered and flagged based on chemical patterns across viability assays (i.e., Z-score) and confirmation assay data. An important point regarding application of the Z-score is that, in practice, it is more useful as a flag than an absolute cutoff. In the ToxCast data analysis pipeline, there are additional types of flags, e.g. to indicate noisy data, or hits due to a single point crossing the statistical threshold for

activity. *These flags do not change the hit call, but provide the user a set of cautions or warnings when evaluating data for a particular chemical-assay pair* (U.S. EPA, 2016b). Similarly, the analysis of the confirmation assay data produces a set of flags that instill more or less confidence in true AR antagonist behavior. The initial Tox21_MDAKB2_Luc_Antagonist assay run with a stimulatory R1881 concentration of 10 nM (~20x EC50) identified predominately only the strong antagonists, i.e. steroid pharmaceuticals, that could compete with the high agonist concentration. Many of the weak environmental anti-androgens were inactive against the 10 nM ligand concentration, whereas when the assay was run with 0.5 nM of R1881 (~EC50) many more of the weak antagonists were identified. The shift in potency between the two conditions was useful for identifying indirect inhibitors of the assay signal. Chemicals that had high model scores for antagonism (R2 AUC>0.1) but were flagged for a lack of a potency shift in the confirmation results may not actually be acting through the AR, but rather through generalized cell stress or technology interference. Examples of chemicals in this group include dyes: basic blue 7, rhodamine 6G, and FD&C green No. 3l; organometallics: tributyltin methacrylate and zinc pyrithione; and, pesticides: abamectin and propargite. Conversely, chemicals that were missed by the binding (A1-A3) and coregulator recruitment (A4-A5) assays, but exhibited a potency shift in the confirmation data may have been incorrectly predicted by the model as acting through interference pathways (e.g., R7, corresponding to activity in only A10-11). It is also possible that some antagonists may bind outside the ligand binding domain or otherwise block dimerization, or act on some later step in the pathway. For example, a group of 7 conazoles were classified as antagonists by the AR pathway model, had activity in both runs of the Tox21_MDAKB2_Luc_Antagonist assay, and a corresponding significant potency shift. Another 6 had a shift in the correct direction, but the confidence intervals for the two AC50s overlapped. A clear shift in the confirmation assay data may be sufficient evidence of AR-mediated activity to prioritize the chemical for additional testing, regardless of model score. Chemicals with this type of response that may have been missed by the model were identified and prioritized by the activity confidence scoring system. Any chemical with a non-zero confidence score could be considered to have potential AR pathway activity, with higher scores indicating greater confidence.

Having 11 diverse orthogonal assays along the AR pathway protects against spurious results being driven by one particular technology type. This is evident when considering the

excellent performance of the AR pathway model against the reference chemicals (>95% for both agonism and antagonism). An interesting exception is the putative reference chemical 17α-estradiol, which was classified negative for AR agonism based on multiple literature results; however, the HDTs were ≤10 μM. All 11 Tox21/ToxCast AR assays were activated by 17α-estradiol (AC50/IC50 range 0.1 – 10 μM), resulting in a model prediction of both agonist and antagonist activity. These results could be indicative of true selective AR modulation by this chemical, or heightened sensitivity of the HT assays to strong steroid pharmaceuticals. With the release of these analyses, and the availability of the ToxCast and Tox21 data (https://www.epa.gov/chemical-research/toxicity-forecasting), the reference chemical list can be updated to reflect the contribution of these assays to the body of published literature. Although we refrained from doing so here to provide an external validation for the current AR pathway model, future work could incorporate the ToxCast, Tox21, and other assays into an expanded reference chemicals list. In that case, the contradictory results between the literature analysis and the ToxCast/Tox21 data would suggest removal of 17α-estradiol from future negative reference classifications if the source of crosstalk, whether it is biological or technological, can be determined.

Another potential lesson learned from validating the AR pathway model against the reference chemicals concerns the threshold for positive activity. Two of the weak/very weak antagonist reference chemicals had AUC values in the inconclusive range, around 0.04, due to lack of activity in the binding assays. A limitation of the binding assays specifically is that chemicals were only tested in concentration response if they were active in a single high concentration screen. Both of these chemicals (fenarimol and methoxychlor) had similar profiles, with activity at 30-40 μM in one of the coregulator recruitment assays (A5) and both of the Tox21 antagonist assays (A10-A11). Depending on the application and the desire to minimize false negatives in a regulatory setting, chemicals with both inconclusive and positive AR pathway activity could be prioritized for further testing.

Here, we presented a comparison of the AR pathway model integrating 11 HT assays and the existing *in vitro* AR binding assay in the U.S. EPA EDSP Tier 1 battery. The overall summary of the comparison between the model and the Tier 1 AR binding assay is that the model correctly identifies binders with potency in the tested range (IC50 under 100 μM), but yields a significant number of false positives, especially as putative antagonists. However, most

of these are identified as false positives using a combination of the antagonist confirmation assay, and examination of assay interference channels. Finally, the model provides evidence in contradiction to the ICCVAM designations for at least 4 chemicals (17α-estradiol, 4-cumylphenol, apigenin, and bisphenol B), which should prompt further investigation. Like the ER model (Browne et al., 2015), it appears the AR pathway model is more sensitive and also more quantitative than the EDSP Tier 1 assays, based on the diversity of the 11 HT assays and the computational network that integrates those data. Like the ER model (Browne et al., 2015), it appears the AR pathway model is more sensitive and also more quantitative than the EDSP Tier 1 assays, based on the diversity of the 11 HT assays and the computational network that integrates those data.

2.5 Responses to the Recommendations Noted in the December 2014 FIFRA SAP

2.5.1 Introduction

The first-generation AR model was presented for validation in the December 2014 FIFRA SAP (U.S. EPA, 2014a). Pertinent suggestions noted in that SAP have now been addressed (see Section 2.5.2) Please consider Section 2.8 as *informational*, supporting the conclusion that the AR model can now be used as an alternative to the LT AR assay under OSRI.

The inclusion of orthogonal assays that query key events along a biological pathway in multiple ways has produced novel hazard screening capabilities. A similar mechanistic pathway model to the one presented here is already being used by the EPA EDSP to identify potential endocrine disruptors acting via estrogen agonism (U.S. EPA, 2015f).

The second generation AR pathway model integrates the results of 11 biochemical and cell-based *in vitro*, HT screening assays. The methodology was covered in detail in Section 2.2. For many of the assay endpoints multiple orthogonal assays were also included, increasing confidence in the combined results. For AR antagonism, a confirmatory pair of assays was performed at two different ligand concentrations providing the ability to observe a diagnostic shift in potency indicative of receptor-mediated activity. The assay results were also combined with cytotoxicity information via a confidence scoring system, to contextualize the results and reduce potential false positives. This model allows several advantages over the current AR assay, and these advantages are presented in Section 2.8.2.

2.5.2 Response to Comments

The first-generation AR model (in an earlier stage of development) was previously reviewed during the December 2014 FIFRA SAP (U.S. EPA, 2014a). The following details a

summary of the areas that the previous SAP wanted improved and how the Agency acted to address these issues:

- **Comment**: Particular attention should be given to issues related to the factors and chemicals that contribute to cytotoxicity and cell stress. The majority of chemicals interacting with AR have antagonist activity so assays, and AUC values must be able to distinguish between cell toxicity/cell stress and authentic AR antagonism.

  Agency Response: The use of a Z-score, as a measure of cell stress/cytotoxicity and detailed in Section 2.2.5 was implemented and is considered to be helpful in avoiding misclassification of chemicals due to cell stress in the assays and assay interference, as detailed in Section 2.3.7.
- **Comment**: As presented to the Panel, the AUC value range is narrow and lacks significant magnitude/range for discriminating between AR bioactivity values/scores that assigned to specific chemicals. The Panel encourages the inclusion of a wider range of chemicals among different structural classes to inform the future studies using these methodologies.

  Agency Response: At least 1855 chemicals have been analyzed through this model. Through a systematic literature search, 37 agonist and 28 antagonists were identified as reference chemicals with varying potencies compared to only 23 total reference chemicals in 2014. Thus, the number of reference chemicals were almost tripled. Potency categories included negative, weak, moderate, and strong for agonists; antagonist categories were the same except with the addition of a very weak category. The methodology for the systematic literature search and criteria for the selection of reference chemicals are presented in Sections 2.2.8 and 2.2.9, respectively.
- **Comment**: Optimize the assessment of activities, particularly antagonism. Particular attention should be given to issues related to assay interference.

  Agency Response: Sensitivity and specificity are now >95% for the second-generation AR model (Section 2.3.6). The use of confirmatory assays (Section 2.2.6) has enhanced the accuracy.
- **Comment**: Measures should be taken to demonstrate that results from the model are reproducible.

  Agency Response: Uncertainty analysis was run for the model (see Section 2.2.7), and the results are reported in (Kleinstreuer et al., 2017b):

- Figure S7 (see tx6b00347_si_001.pdf in Kleinstreuer et al., 2017b) in "Results for the AR pathway model on 1855 chemicals" reports all 55 ICCVAM chemicals with the AR AUC score +/- CI).
- "Comparison of the results for the chemical groups" reports all of the AR AUC scores +/- CI (see tx6b00347_si_002.pdf in Kleinstreuer et al., 2017b).
- The "AR pathway model" Excel Supplemental File shows all the scores and the "Detailed Data" tab presents the 95% CI bounds (see tx6b00347_si_004.xlsx in Kleinstreuer et al., 2017b).

  These results demonstrate adequate reproducibility for the model.

- **Comment**: Details of the methods and results must be available to increase transparency.
  Agency Response: The AR Supplemental File shows details of each assay used. Supplemental files are also available that provide a summary of the results (Kleinstreuer et al., 2017b). The R-code for the analysis is supplied here (Watt, 2016). Extensive efforts were made in this document to be comprehensive in supplying information in order to be completely transparent.
- **Comment**: The EPA team was encouraged by the Panel to build on the battery of AR bioactivity assays.
  Agency Response: Two additional assays were added to the battery bringing the total from 9 to 11. Considering the excellent predictive capacity of this model (Section 2.3.6), additional assays may be unnecessary.
- **Comment**: Whereas the current focus is on the AR nuclear receptor genomic activity pathway, attention should also be given to the development of alternative AR-related assays that do not follow the classical genomic/nuclear receptor pathway. Metabolism and *in vivo* conversion of parent chemical compounds to active metabolites remains a concern with the current battery of *in vitro* assays. The SAP also suggested that the Agency address the ability to replicate the multiplicity of biological actions that chemicals produce *in vivo*, such as through bioactivation, non-genomic androgenic effects, and potential off-target effects.
  Agency Response: The Agency is concerned with the ability of *in vitro* models to predict *in vivo* effects, and efforts have been made in that regards. The Agency is considering *in silico* approaches and additional assays with metabolic competency to address these issues. However, the Agency is proposing the HT H295R assay as an alternative for the LT H295R assay. Consequently, the HT H295R assay does not have to have characteristics that the LT H295R assay does not have.

2.6 Limitations of the current model and future refinements

2.6.1 Metabolic Capacity

Limitations of this model, and almost all other HT-based approaches, include the lack of or limited metabolic capacity of the systems. The ATG_AR_TRANS assay (A6) was performed using a HepG2 cell line, which has limited metabolic capacity; the other cell-based assays (A4-A5, A7-A11) were conducted in HEK293T and MDA-Kb2 cells which are generally regarded as metabolically incompetent. It should be further noted that the AR binding assay currently in the EDSP Tier 1 battery also lacks metabolic competency.

The ability to incorporate metabolic competence into high-throughput screening assays is the subject of ongoing research at the US EPA Office of Research and Development. In addition, the US EPA has partnered with the National Toxicology Program (NTP) and the National Center for Advancing Translational Science (NCATS) to engage the broader scientific and commercial communities in tackling the issue of incorporating metabolism into cell-based and cell-free assay systems through the release of a challenge competition (EPA news release). Further, chemical structure-based models are under development to identify chemicals predicted to undergo transformation to more bioactive metabolites. A proof-of-principle for using predicted metabolites in structure-based model predictions is available for the ER pathway model (Pinto et al., 2016).

2.6.2 Chemical Library Restrictions

Currently the AR pathway model and associated high-throughput screening assays are limited by the chemical libraries made available to the ToxCast and Tox21 screening programs. The current libraries are restricted to DMSO-soluble chemicals (Richard et al., 2016). Future plans also include expanding chemical testing to a water-soluble library.

2.7 Future Use of the AR Pathway Model

2.7.1 Alternative to Other EDSP Tier 1 Assays

To ultimately interpret AR pathway activity and other mechanistic events in a biological framework that includes potential adverse *in vivo* outcomes, efforts are underway to establish reference chemicals for additional endpoints and map these to adverse outcome pathways. Following the example of the uterotrophic database (Kleinstreuer et al., 2016), work is ongoing to compile *in vivo* androgen and anti-androgen data from the U.S. EPA EDSP Tier 1 Hershberger assay (U.S. EPA, 2009h). Predicting the results of the Hershberger assay will likely require additional HT assays to be developed in order to build a model that incorporates more molecular-initiating events linked to altered androgen status because the Hershberger assay incorporates additional biology outside of activity in the AR pathway. Chemicals may modulate androgen status in the Hershberger assay via AR pathway activity, inhibition of 5α-reductase, or inhibition of other enzymes critical to steroid hormone synthesis. Thus, additional assay data would be needed to cover the complex biology present in the Hershberger model.

Future incorporation of the AR pathway model into predictions of Hershberger activity will also require application of high-throughput *in vitro* toxicokinetic assays and *in vitro*-to-*in vivo* extrapolation (IVIVE) approaches to provide an *in vivo* dose context (Wambaugh et al., 2015; Wetmore et al., 2015; Chang et al., 2014; Wambaugh et al., 2014; Rotroff et al., 2010b). The high-throughput toxicokinetic assays and IVIVE approaches allow the conversion of steady state blood concentrations (equivalent to *in vitro* potency values (µM) from the AR pathway model) into estimates of *in vivo* administered doses (mg/kg/d). These efforts can be used to validate additional high-throughput *in vitro* assays and computational models to predict more complex developmental and reproductive effects.

2.7.2 Prioritization and Risk Assessment

Although the HT results and AR pathway model predictions have demonstrated ability to effectively prioritize environmental compounds for endocrine disrupting potential on a hazard basis, they should be integrated with exposure estimates for decision making in a risk assessment framework (Paul Friedman et al., 2016; Teeguarden et al., 2016; U.S. EPA, 2015f). The integrated bioactivity and exposure (IBER) methodology was presented to a FIFRA SAP in December 2014 (U.S. EPA, 2014); in this approach, the bioactivity based on the ER or AR

pathway model are put in a dose context and compared with high-throughput exposure estimates (Wambaugh et al., 2014) to prioritize chemicals for further evaluation.

2.8 Performance-Based Approach to Establishing Confidence: Considerations & Conclusions for the AR Pathway Model

2.8.1 The Performance-Based Approach

The Agency has compiled a database of literature results that includes a wide array of AR binding and transactivation data, and used it to characterize a range of potential AR agonist and antagonist reference chemicals. The proposed reference chemical lists and associated potency categories can be used for current and future test method evaluations. The AR pathway model validated against this independently curated set of reference chemicals identified a number of environmental chemicals as potential AR antagonists, with varying degrees of confidence. These prioritized chemicals should be examined in the context of human and environmental exposures, metabolism, and persistence to characterize the risk of endocrine disruption and adverse outcomes in humans or wildlife.

The narrative in Section 1.5 provides an overall framework for establishing confidence for new, alternative approaches. In the case of the AR pathway model, each of the considerations have been met in the regulatory context of screening and prioritization of chemicals for purposes of evaluating potential for disruption of the AR pathway:

- *Mechanistically and/or Biologically Relevant Assays:* The suite of 11 *in vitro* HT assays used as inputs to the AR pathway model covered multiple known key events (e.g., receptor binding, receptor dimerization, chromatin binding of the transcription factor complex, and gene transcription) in the AR pathway. Five of the assays measure AR interaction irrespective of agonism or antagonism; 4 assays measure events downstream of AR agonism; and, 2 assays specifically measure AR antagonism.
- *Reliability considering accepted best practices within the given field:* Reliability of the data and the model presented here must be interpreted within the context of being fit for the purpose of screening for prioritization of large numbers of chemicals. Thus, uncertainties in model outputs may be higher than those used to derive higher tier regulatory values (e.g., RfDs, RfCs), but still useful for prioritization as has been done for computational ER models (Browne et al., 2015; Judson et al., 2015).

The AR pathway model was used to integrate results from 11 Tox21/ToxCast HT assays. For many of the assay endpoints multiple orthogonal assays were also included, increasing confidence in the combined results. For AR antagonism, a confirmatory pair of assays was performed at two different ligand concentrations providing the ability to observe a diagnostic shift in potency indicative of receptor-mediated activity. The assay results were also combined with cytotoxicity information via a confidence scoring system, to contextualize the results and reduce potential false positives. The AR pathway model results were validated against an independently curated set of reference chemicals and shown to be over 95% accurate for predicting both AR agonism and antagonism.

- *Transparency:* Supplemental files and figures listed in Section 2.6 contain the detailed information about the assays and data used, results, and associated analyses. The supplemental files include OECD-compliant summaries of the operating procedure, assay annotations, and performance characteristics for each of the assays used in the AR pathway model. The HT data (U.S. EPA, 2017a, 2016b, 2015e) used as the basis for this AR pathway activity model and code for analysis (Watt, 2016) are also publicly available. Assay descriptions are available from these sources.
- *Uncertainties and Limitations:* Uncertainty in the AR pathway modeling results was characterized using a bootstrapping approach. For each *in vitro* assay and concentration response curve used as input in the model, one thousand bootstrap replicates were generated to obtain a distribution of fit parameters, model selection, potency estimates, and activity calls. The uncertainty associated with the concentration response in each assay was propagated to the AR model score (i.e., AUC value) enabling the calculation of confidence intervals for each chemical. These confidence intervals are provided in the supplemental files. The limitations of the *in vitro* HT assays and AR pathway model are described in Sections 2.4 and 2.6.
- *Reference Chemicals:* A systematic literature review was performed that identified AR binding data for 111 chemicals and AR transactivation data for 160 chemicals. The range of binding affinities and agonist potency estimates spanned over 7 log units, while antagonist estimates spanned 4 log units. Based on rigorous criteria outlined in the Methods for reproducibility and consistency of response, 37 reference chemicals for AR agonism and 28 reference chemicals for AR antagonism were identified. These reference

lists included strong, moderate, weak and very weak agonists and antagonists, and negative categorizations. There were 11 chemicals that fulfilled reference criteria for both agonism and antagonism, usually as a positive reference in one and a negative reference in the other. Of the 54 total reference chemicals, 46 were also tested in ToxCast/Tox21 and were used for AR model validation.

- *Peer Review:* The FIFRA SAP reviewed the initial version of the AR pathway model in 2014. The improvements to this model are summarized in Sections 2.5 and 2.8.2. This work was peer reviewed and published in open literature (Kleinstreuer et al., 2017a). The 2017 FIFRA SAP will provide another external peer review.

2.8.2 Summary and Conclusions

The Agency considers that the 11 HT assays integrated in the AR pathway model, along with cell stress and cytotoxicity measures, as a robust characterization of bioactivity compared to the single validated AR binding assay used to assess the bioactivity of chemicals by (ICCVAM, 2003). Like the ER pathway model (Browne et al., 2015), it appears the AR pathway model is more sensitive and also more quantitative than the EDSP Tier 1 assays, based on the diversity of the 11 HT assays and the computational pathway model that integrates those data. This AR pathway model will serve as a form of other scientifically relevant information that can function both in the prioritization of substances for future evaluation, as well as in the weight of evidence consideration of test results to determine the need for further evaluation.

All pertinent suggestions from the December 2014 FIFRA SAP have been implemented. The Agency proposes that the relevance and fit-for-purpose criteria have been met for this model in prioritization and Tier 1 screening. Also, uncertainty analysis demonstrates that the model can generate reproducible results.

This model offers the following advantages and benefits over the current guideline approach:

- This model, integrating 11 *in vitro* assays, is considered a more robust and informative way to assess androgen receptor binding than the current single *in vitro* assay;
- The model can provide data that may contribute to future weight of evidence evaluations of a chemical's potential of interacting with the androgen receptor;
- Provide data indicative of how (agonist, antagonist, mixed) a chemical is interacting with the AR;

- Rapidly prioritize chemicals in the EDSP universe based on quantified AR pathway bioactivity; and
- Does not require animal use, as the assays employ recombinant AR enzyme (cell-free) or established cell lines. The current LT *in vitro* AR assay does require animal use from which rat prostate cytosol is obtained.

Furthermore, the Agency's prioritization approach will use EPA's ToxCast HT *in vitro* screening assays and computational models. Conducting the LT EDSP Tier 1 screening battery requires a great deal of financial, temporal, animal, and human resources. Since the first test orders were issued in 2009, 52 chemicals have been screened in the battery and List 1 weight of evidence determinations have been finalized by EPA. Incorporating HT assays into the existing EDSP framework can transform an undertaking taking decades to centuries to complete into a project that can likely be completed in years and simultaneously increase the likelihood that the most bioactive chemicals are prioritized. (Exposure will also be a component of prioritization, but the integration of exposure estimates into prioritization will not be discussed at this SAP.) This strategy in the EDSP approach is consistent with the 2007 NAS report, specifically to: (1) provide broad coverage of chemicals examined; (2) reduce the cost and time of toxicity testing; (3) reduce animal use; and (4) develop a robust scientific basis for assessing health effects of environmental agents. This strategy is also supported by the 2017 NAS report (National Academies of Sciences et al., 2017).

Based on characteristics described here, EPA concludes that the AR pathway model is sufficiently robust for purposes of screening and prioritization of chemicals for evaluating potential for disruption of the AR pathway. Furthermore, the AR pathway model can be used as an alternative to the Tier 1 AR assay (OCSPP 890.1150).

2.9 Supplemental File

AR Supplemental File. AR Assay Descriptions:

Includes OECD-compliant assay summaries for all of these assays included in the AR pathway model.

3. Steroidogenesis Pathway Activity

3.1 Introduction

3.1.1 Current EDSP Tier 1 *In Vitro* Approach for Screening for Perturbation of Steroid Biosynthesis

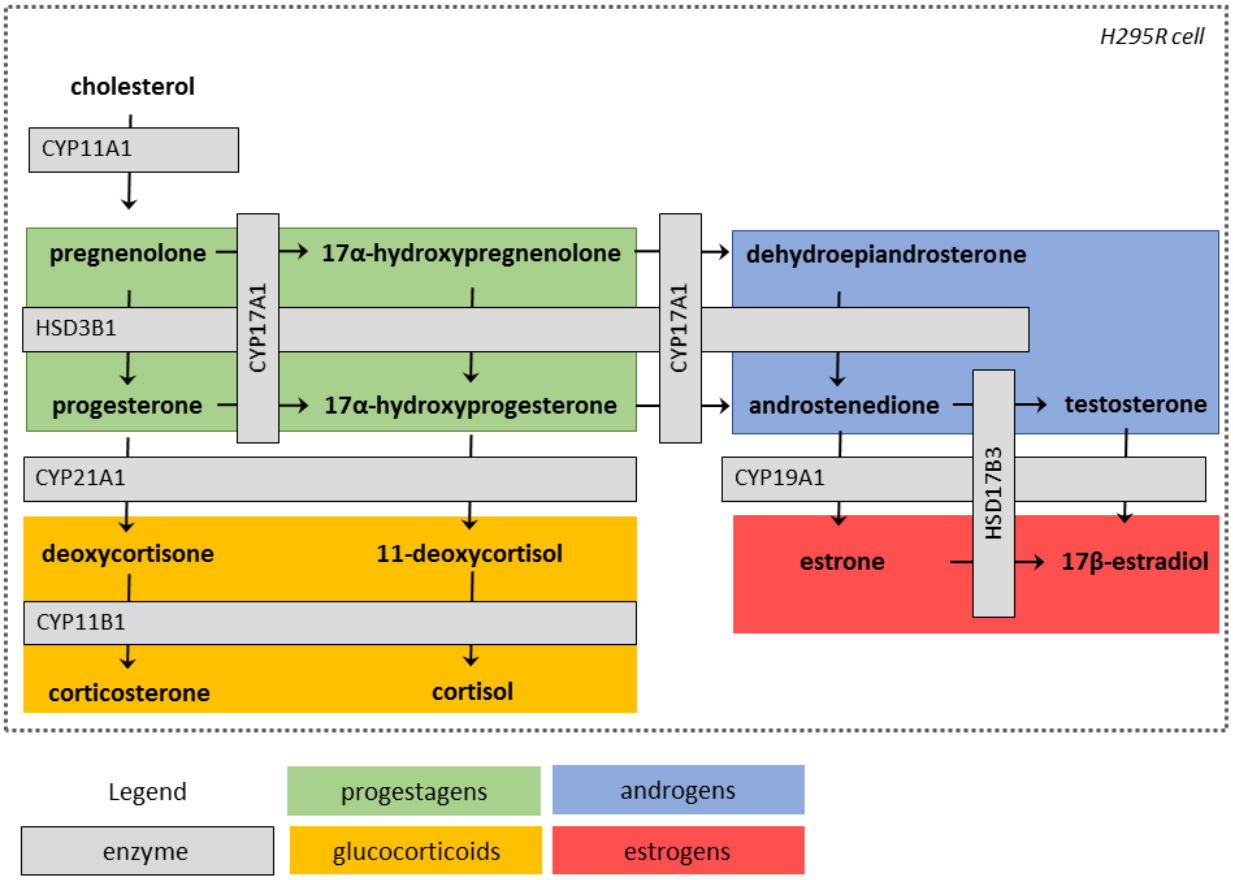
The EPA EDSP Tier 1 battery includes two assays that aim to determine whether or not a chemical may affect *in vitro* steroid hormone biosynthesis: the human recombinant aromatase assay (OCSPP 890.1200) and the steroidogenesis assay (OCSPP 890.1550) using the human H295R adrenocortical carcinoma cell line (U.S. EPA, 2009a, d). The H295R cell line demonstrates the biological characteristics of zonally undifferentiated human fetal adrenal cells and is used to evaluate steroid hormone biosynthesis that would normally occur in adult adrenal cortex, ovaries, and testes (Gracia et al., 2006; Gazdar et al., 1990). The utility of the H295R assay in screening for putative steroidogenesis disruptors has been recognized internationally; both the OECD and the EDSP have developed test guidelines for utilizing the H295R cell line in a cell-based steroidogenesis assay to detect potential chemical perturbation of estradiol (E2) and testosterone (T) production (OECD, 2011; U.S. EPA, 2009a).

Assay data for the H295R assay, as conducted according to OCSPP 890.1550, have been reviewed for 45 of the 52 EDSP List 1 chemicals (due to registrant submission of other-scientifically relevant information in support of waiver requests). Based on the size of the EDSP chemical universe, it would be ideal to develop screening-level information for as many relevant chemicals as possible in order to inform further endocrine prioritization considerations. The purpose of this chapter is to describe progress on a HT approach to prioritization of putative steroid biosynthesis pathway disruptors.

3.1.2 Objectives

The objectives of this chapter are three-fold: (1) compare the performance of the HT H295R assay with the current Tier 1 LT H295R assay focused only on changes in E2 and T concentrations following treatment with a series of reference chemicals; (2) introduce a novel statistical approach that integrates the measurements of E2, T, and 9 additional steroid hormones from the HT 295R assay to quantify the overall impact of the substance on the steroidogenesis pathway; and (3) to provide a regulatory perspective on potential future use of the HT H295R assay.

Figure 3-1 Overview of the Steroidogenesis Pathway



Representation of the steroid biosynthesis pathway expressed in H295R cells. Figure adapted from Karmaus *et al*. and Saito *et al*. (Karmaus et al., 2016; Saito et al., 2016).

3.1.3 Background on the OECD-Validated H295R Assay and the ToxCast HT H295R Assay

Conduct of the H295R assay by the OECD Test Guideline (TG) 456 or the EPA test guideline (OCSPP 890.1550) involves measurement of only E2 and T in the cell culture medium from exposed H295R cells as indicators of steroidogenesis disruption (OECD, 2011; U.S. EPA, 2009a). Briefly, when performed to guideline specifications, H295R cells are acclimated in 24-well plates for 24 hours, exposed for 48 hours to test chemical in triplicate, and then medium is removed for steroid hormone measurement by ELISA or analytical chemistry. The cells are then evaluated for viability. This assay procedure was previously adapted (Karmaus et al., 2016) for HT application in the US EPA ToxCast program, with primary modifications including: the use of a single concentration pre-screen to determine chemicals most likely to perturb

steroidogenesis in multi-concentration screening; the uniform use of a 48-hour pre-stimulation period with forskolin; measurement of 13 steroid hormones using high-performance liquid chromatography followed by tandem mass spectrometry (HPLC-MS/MS); and the use of a 96-well format. These modifications were all intended to increase screening efficiency and fill data gaps related to *in vitro* steroidogenesis for large numbers of chemicals. The HT H295R screening effort demonstrated that the assay performed reproducibly and robustly with positive controls, forskolin and prochloraz, and prototypical modulators including the conazole fungicides (Karmaus et al., 2016). Statistical analysis using the Z'-factor, an indicator of assay robustness was calculated for each hormone to evaluate the signal-to-background difference, inter-sample variability, and the ability to distinguish positive and negative test chemicals from noise (Zhang et al., 1999). As reported previously, ten of the 13 measured steroid hormones in the HT H295R assay demonstrated a median $Z' \geq 0.5$ under stimulation with forskolin and inhibition with prochloraz, indicating a robust assay. Three hormones were excluded when a different data analysis methodology was used previously due to a number of samples with concentrations below the lower limit of quantitation (Karmaus et al., 2016). Strictly standardized median difference (SSMD), a measure of effect size, was also calculated to demonstrate overall assay quality and directionality. Forskolin generally increased hormone quantities with good dynamic range (SSMD values ≥7), whereas prochloraz generally inhibited hormone production with good dynamic range (SSMD values ≤-7) (Karmaus et al., 2016). These assay quality metrics suggest that the HT H295R screening assay may be useful not only for evaluating disruption of estrogen and androgen synthesis specifically, but also in measuring the effects on the synthesis of a broader suite of steroid hormones and the steroid biosynthesis pathway.

3.1.4 Analysis Approach for the HT H295R Data

All of the HT H295R steroid hormone data, including E2 and T, were analyzed per a similar methodology to the one outlined in the OECD inter-laboratory validation study for OECD TG 456 (Hecker et al., 2011), rather than using the ToxCast data pipeline as used previously for a subset of these data (Karmaus et al., 2016). This comparison was performed across a set of 25 reference chemicals and enables a direct evaluation of whether the HT H295R assay may function as a possible alternative to obtain screening level information about chemical effects on *in vitro* production of E2 and T specifically.

To enable prioritization beyond only E2 and T production, a novel statistical approach was developed that integrated the effects across 11 of the hormones measured (OH-pregnenolone, progesterone, OH-progesterone, 11-deoxycortisol, deoxycorticosterone (DOC), cortisol, corticosterone, androstenedione, T, estrone, and E2) for each concentration of a chemical tested. A statistical measure based on Mahalanobis distance (SAS, 2012; De Maesschalck et al., 2000) and the HT H295R data for prediction of effects on E2 and T production for 25 reference chemicals. The integrative statistical ranking metric developed herein was then added to the comparison of the reference chemicals in order to demonstrate the added quantitative value of this metric beyond simple consideration of the number of steroid hormone analytes perturbed.

3.2 Methods

3.2.1 Chemical Library

Previously, data were collected using the HT H295R assay for 1,998 unique test chemicals at a single high concentration, with 514 of these chemicals screened in multi-concentration response (Karmaus et al., 2016; U.S. EPA, 2015e). Including this current study, 2,012 unique test chemicals have been screened at a single high concentration (100 µM, solubility- and viability-permitting), with 656 chemicals assayed in concentration-response ranging from 0.041 nM to 100 µM. One chemical, triadimenol, was assayed in concentration-response with two different concentration ranges, and as such is given two unique chemical identifiers in any analyses (for a total of 657 chemicals, but 656 unique CAS numbers). The chemicals were selected from the ToxCast phase I, II, III, and endocrine 1000 (E1K) libraries, which were compiled based on commercial availability and solubility in dimethyl sulfoxide (DMSO) among other considerations to capture a broad chemical space (Richard et al., 2016). Phase I contained a high percentage of pesticide active ingredients and chemicals for which additional *in vivo* data were available; phases II and III broadened the chemical landscape and included a greater diversity of chemical use types (Richard et al., 2016). The E1K chemical library, a set of roughly 800 chemicals enrichhed for endocrine-active chemicals, was also included. Data on the complete ToxCast chemical library is publicly available for download (U.S. EPA, 2015e). A top nominal stock concentration of 100 mM in DMSO was attempted, solubility-permitting, for the entire library.

The majority (approximately 80%) of the 656 chemicals advanced for concentration-response screening in the HT H295R assay demonstrated changes of 1.5-fold or greater relative to control in single concentration screening for ≥ 3 steroid hormone analytes from the pathway at

the maximum tested concentration that maintained ≥ 70% cell viability. Most of the chemicals advanced to concentration-response screening (approximately 60%) affected ≥ 4 steroid hormones in single concentration screening (Karmaus et al., 2016), with some exceptions to include additional positive and negative chemicals for reference and specific chemical classes of interest. The rationale for this pre-selection screening workflow was three-fold: (1) on a hypothetical basis, modulation of even one enzyme in the pathway would theoretically perturb the concentrations of at least 4 steroid hormones; (2) empirically, the recall sensitivity or percentage of positive responses that repeated between single concentration and concentration-response screening was high (86%) when a cumulative total of ≥ 4 hormones were affected in single concentration screening (Karmaus et al., 2016); and, (3) identification of chemicals with the greatest perturbation of the interrelated steroidogenic pathway responses represented a sensible approach to reducing the resources needed to screen a chemical set in concentration-response.

3.2.2 HT H295R Assay and Quantification of Steroid Hormones

The HT H295R assay (Karmaus et al., 2016) is comprised of four main experimental components: (1) H295R cell culture and treatment; (2) cell viability assay using the MTT (3-[4,5-dimethylthiazol-2-y]2,5-diphenyltetrazoliumbromide) tetrazolium reduction assay; (3) quantification of steroid hormones in the media from exposed H295R cells; and, (4) statistical analysis of steroid hormone concentrations. The HT H295R assay was conducted in accordance with the OECD TG 456 (OECD, 2011), with modification to increase the throughput of the assay. Key aspects of the assay design, conduct, and analysis by the OECD Test No. 456 and the HT H295R assay are summarized and compared in **Table 3-1**.

Table 3-1 H295R Steroidogenesis Assay Methodology Comparison

A summary of the OECD TG 456 requirements versus the performance of the HT H295R assay (as currently implemented).

Design Phase	Aspect	OECD TG 456	HT H295R
Cell culture	Plate format	24-well, but OECD TG specifies other plate formats can be used (*e.g.* 48-well)	96-well
	Experimental timeline	24 hours acclimatization of cells, followed by 48 hours chemical exposure, terminated at sample collection.	Overnight acclimatization of cells, followed by 48 hours pre-stimulation with forskolin, followed by 48 hours chemical exposure, terminated at sample collection.
	Cell passage	5-10	5-10

Design Phase	Aspect	OECD TG 456	HT H295R
	Target cell confluency	50-60%	50-60%
	Replicates	Triplicate technical replicates Triplicate biological replicates	Duplicate technical replicates Most of the library had one biological replicate; ~16% was screened with two-three biological replicates.
Viability testing	Viability measures	Live/Dead® or MTT assay	MTT assay
	Cell viability threshold	≥ 80%	≥ 70%
Hormone detection	Baseline stimulation	None[1]	Cell are pre-stimulated for 48 hours in medium containing 10 µM forskolin.
	Minimum basal production	500 pg/mL or ≥ 5-fold method detection limit (MDL) for T and 40 pg/mL or ≥ 2.5-fold MDL for E2.	Following forskolin stimulation, DMSO-exposed H295R demonstrated 2.19 ± 0.32 ng/mL and 1.57 ± 0.36 ng/mL for T and E2, respectively. This is ≥ 5-fold the LLOQ for both hormones.[2]
	Accuracy	Within 30% of nominal concentrations.	98.1-101.7% recovery for 13 hormones (Karmaus et al., 2016).[2]
	Precision	Variation between replicate samples should be ≤25%.	Percent relative standard deviation for controls ranging from 3.3 – 10.0% during assay optimization for the 13 hormones measured (Karmaus et al., 2016).[2]
Steroid hormone data analysis	Analysis	Normally distributed data: an analysis of variance (ANOVA) with differences from vehicle control evaluated using a Dunnett's test. Non-normally distributed data: Kruskal-Wallis test followed by a Mann-Whitney *U* test.	Initial data analysis used the ToxCast data pipeline (tcpl) (Filer et al., 2016) to enable standardization of the data with other HT data and a first look at the data. Data analysis for comparison to the OECD reference chemicals involved use of an ANOVA with differences from vehicle control evaluated by Dunnett's test (new in this work).
	Criteria for positive	Two consecutive concentrations and/or maximum non-cytotoxic concentration significantly different from control.	See (Karmaus et al., 2016) for a description of the tcpl analysis employed for a first analysis. For the ANOVA approach presented here: two consecutive concentrations and/or maximum non-cytotoxic concentration significantly different from control.

[1]22-R-Hydroxycholesterol has been suggested as a medium supplement (20-40 µM) to increase basal E2 production as needed, but it is not part of the standard protocol. Further, the OECD validation report (2008) noted that, "during the qualifying experiments it was only expected that the laboratory showed conformance with the performance criteria for E2 induction after exposure to the stimulator forskolin."
[2]Note these are reported performance results and not criteria for acceptance of the HT H295R assay data.

Cell Culture and Treatment

The cell culture, treatment, and assay conditions of the HT H295R assay have been described previously in detail (Karmaus et al., 2016). All cell culture and treatments were conducted by Cyprotex US, LLC (formerly CeeTox, Inc.) (Kalamazoo, MI). Briefly, H295R cells (ATCC CRL-2128) were expanded for five passages and frozen in batches in liquid nitrogen. Prior to experimentation, batches of H295R cells were thawed and passed at least four times, taking care that the maximum passage number used for experimentation was ten. Cells were maintained in a 1:1 mixture of Dulbecco's Modified Eagle's Medium with Ham's F-12 Nutrient mixture (DMEM/F12) supplemented with 5 ml/L ITS+ Premix (BD Bioscience) and 12.5 ml/L Nu-Serum (BD Bioscience). Cells seeded at 50-60% confluency into 96-well plates were acclimated overnight. Culture medium was then replaced with 175 µL of medium containing 10 µM forskolin to stimulate steroidogenesis for 48 hours. The forskolin stimulus medium was replaced with medium supplemented with test chemical or controls (forskolin, prochloraz, or digitonin) added to a final concentration of 0.1% DMSO. On each 96-well plate, duplicate treatment wells were included for all chemical treatments as well as controls (10 µM forskolin and 3 µM prochloraz), in addition to two to four DMSO solvent control wells and six cell viability control wells (250 µM digitonin). The test chemicals were assayed on eight different dates, and each experimental date is used to indicate block throughout the study in order to account for observed block effects. Most test chemicals were assayed in one plate-block combination with technical duplicates only; approximately 16% of the screened library (107 of 656 unique chemicals screened in concentration-response) were assayed on more than one plate-block combination. Following 48 hours of test chemical exposure, medium was removed, split into two vials of ~75 µL media each, and stored at -80°C prior to steroid hormone quantification.

Cell Viability Assay

Cell viability was evaluated by MTT cytotoxicity assay after chemical treatment in all studies, and was previously described in (Karmaus et al., 2016). Briefly, after chemical exposure and removal of media, 100 µL of 0.5 mg/ml 3-[4,5-dimethylthiazol-2-y] 2,5-diphenyltetrazoliumbromide (MTT) solution was added to the cells remaining in the 96-well treatment plates. Following a 3-hour incubation at 37°C and 5% CO_2 to allow formazan-MTT crystal formation, the MTT solution was removed and blue formazan salt crystals were solubilized using 100 µL anhydrous isopropanol with shaking for 20 minutes. Absorbance at 570 and 650 nm were measured using a BioTek Synergy H4 plate reader. Background correction of

absorbance units was used to determine percent change relative to controls. All plates contained multiple control wells including 10 µM forskolin (n=4; control for stimulation of steroidogenesis), 3 µM prochloraz (n=4; control for the inhibition of steroidogenesis) and digitonin (n=6; control for cell death).

For the first 1,998 chemicals screened, cytotoxicity was used to establish a maximum tolerated concentration (MTC) per chemical sample with a target cell viability ≥70%, as reported previously (Karmaus et al., 2016). ToxCast chemicals were evaluated at a maximum nominal concentration of 100 µM, where possible. MTT cytotoxicity evaluation was also conducted for the duplicates of all concentrations for chemicals tested in the concentration-response studies (CR; six-point CR established by 3-fold serial dilutions from the MTC).

For the 85 additional chemicals with multi-concentration data reported herein for the first time, the MTT assay was run for all concentrations attempted in the HT H295R assay, and the MTC was not used to limit the concentration-response curve. If a stock concentration of 100 mM was achieved, then each chemical was tested at 100, 33.33, 11.11, 3.70, 1.23, and 0.41 µM in the MTT assay for these 85 chemicals. Otherwise, the same dilution series was performed using the highest possible stock concentration of test chemical. The purpose of this change in the experimental workflow was to enable full concentration-response curves for the steroid hormone analysis to be visualized without implementing the MTC logic, which may have limited the ability to observe effects on steroid biosynthesis in cases when the difference between cytotoxic concentration and a viable and efficacious concentration may have been small.

Steroid Hormone Quantification

Frozen medium samples from treated HT H295R assays were shipped on dry ice to OpAns, LLC (Durham, NC) for extraction and quantification of steroid hormones. As described previously (Karmaus et al., 2016), samples were thawed to room temperature prior to liquid-liquid extraction. Steroid hormones were extracted from media samples using methyl tert-butyl ether (MTBE). An extra derivatization with dansyl chloride was included for estrogen (estrone and E2) detection only. Steroid hormones were separated and quantified using high-performance liquid chromatography (HPLC) followed by tandem mass spectrometry (MS/MS). Specifically, reverse phase C18 gradient elution with electrospray positive ionization was used followed by MS/MS detection. All data were acquired using MassHunter Workstation Acquisition version B03.01 (Agilent Technologies, Inc.), and processed using MassHunter Quantitative Analysis for QQQ.

The lower limit of quantification (LLOQ) and upper limit of quantification (ULOQ) were reported previously (Karmaus et al., 2016) using a 7-point standard curve. The precision and accuracy of the extraction and quantification methods are briefly reviewed in Table 1; the recovery for all thirteen hormones ranged from 98.1-101.7% recovery, and the percent relative standard deviation (%RSD) of the spiked standards ranged from 3.3-10.0%, as reported by (Karmaus et al., 2016). During the sample analysis process, samples were flagged as "not-detected" or "not-quantifiable" when the sample was available, but the steroid hormone analyte was below the LLOQ; in such cases, a surrogate value of the LLOQ/√2 was substituted for analyses herein. Any sample measurement flagged as "not reportable" was set to "NA" for any subsequent analysis. A comparison of the method detection limits (ng/mL) for OECD TG 456 and the HT H295R assay (Karmaus et al., 2016) is provided in **Table 3-2**.

Table 3-2 Comparison of Method Detection Limit (from OECD TG 456) and Reported LLOQs for HT H295R

The gray cells highlight the comparison of testosterone (T) and estradiol (E2), as these are the only two hormones with minimum method detection limits (MDL) in the performance criteria for the test guidelines. The LLOQs for all of the hormones measured in (Karmaus et al., 2016) are listed.

Hormone Family	Steroid Hormone	OECD TG 456	HT H295R Assay	
		Method Detection Limit (ng/mL)	Lower limit of quantitation (LLOQ) (ng/mL)	Upper limit of quantitation (LLOQ) (ng/mL)
Androgen	**Testosterone (T)**	**0.1**	**0.1**	**20**
	Dehydroepiandrosterone[1]	NA	3	600
	Androstenedione	NA	1	200
Estrogen	**Estradiol (E2)**	**0.01**	**0.03**	**6**
	Estrone	NA	0.03	6
Progestagen	17α-hydroxyprogesterone	NA	0.2	40
	17α-hydroxypregnenolone		5	1000
	Progesterone		0.2	40
	Pregnenolone[1]		2	400
Glucocorticoid	11-Deoxycortisol		5	1000
	Deoxycorticosterone		0.5	100
	Cortisol		0.5	100
	Corticosterone		0.5	100

[1]Dehydroepiandrosterone and pregnenolone were excluded from further analysis in the work herein as 69.9% and 53.4% of the measured values for these two steroid hormones, respectively, were below the LLOQ.

3.2.3 Data analysis

All data were analyzed in R (R Core Team, 2015). The R scripts are available statically as Supplemental File 13. Updated versions (including the version for this Whitepaper) are

available at: ftp://newftp.epa.gov/COMPTOX/NCCT_Publication_Data/Haggard/2017_Prediction_of_H295R_steroidogenesis_Pathway_Perturbation/.

Cell Viability Assay Data Processing

Initially, and as described previously (Karmaus et al., 2016), the MTT assay was used to establish a MTC per chemical sample for the first 514 chemicals that were assayed in concentration-response by evaluation of each chemical sample at a target top concentration of 100 µM, solubility-permitting, and then seeking to find a concentration that would maintain cell viability of ≥ 70%. Chemicals that yielded H295R cell viability of 20-70% were diluted 10-fold, while those with <20% viability were diluted 100-fold and re-evaluated. Dilutions were made until ≥70% viability was achieved for all chemicals establishing the MTC. The MTT method differed from (Karmaus et al., 2016) for the additional 85 chemicals (Supplemental File 3, unique plate IDs for 04112017) reported for the first time in this manuscript in that no MTC was determined. MTT assay data were collected for all six concentrations tested, with a target top concentration of 100 µM and decreasing half-log increments (33.33, 11.11, 3.70, 1.23, 0.41), with adjustments made based on chemical solubility.

The concentration-response MTT data for all 656 chemicals screened were processed using the ToxCast data pipeline (tcpl) (Filer et al., 2016) for comparison with the HT H295R steroid hormone data. The data were analyzed as percent control, where the baseline value was defined as the plate-wise baseline of the DMSO control wells:

Equation 7

$$response = \frac{value - baseline\ value}{baseline\ value}\ x\ 100$$

Consistent with previous estimations of the variability around the baseline response for this assay, a 70% cutoff criterion (Karmaus et al., 2016) was established for the purpose of filtering steroid hormone data. This cutoff criterion (allowing up to 30% cell viability loss) corresponded to approximately 4.4-times the baseline median absolute deviation (6.81). Cell viability filtering was performed by matching the MTT percent control response to the steroid hormone data; if the cell viability decreased by >30%, the steroid hormone data for that concentration of a chemical was excluded from any further analysis.

The normalized data by concentration and the resultant plots of these data for 655 of 656 chemical samples are available in Supplemental Files 1 and 2, respectively. Two chemicals, colchicine (CASRN 64-86-8) and digoxigenin (CASRN 1672-46-4), were included as viability

controls, and as expected resulted in substantial loss of cell viability, leaving only one concentration with viable cells; as such, steroid hormone data were not analyzed for these two chemicals. A third chemical, quizalofop-ethyl (CASRN 76578-14-8) had data quality flags in the source files from the vendor that suggested these data should not be used; as such, these data were excluded from any further analysis of steroid hormones or cytotoxicity. This reduced the set of chemicals with concentration-response hormone data available from 656 to 654 unique chemicals, corresponding to 766 chemical samples. Of these 766 chemicals samples, when a 70% cell viability filter was applied, 715, 36, six, five, and four chemicals retained six, five, four, three, and two concentrations for analysis of the concentration-responses for steroid hormones (see Supplemental File 3 for the master steroid hormone data table).

Analysis of Variance (ANOVA) and Post-Hoc Testing for Significance of Effects on Steroid Hormone Profiles

When concentration-response data were available, the vendor-provided source files with raw steroid hormone data (quantified as ng/ml) were converted to micromolar (µM) units and each steroid hormone assay component was analyzed, per the analysis methodology in the OECD TG 456, by an analysis of variance (ANOVA) followed by a post-hoc Dunnett's test with alpha set to 0.05 (a complete table of these values is available as Supplemental File 3). The DMSO control data originating from the same plate the chemical was tested on were used as the sample for comparison. In most cases, a minimum of two technical replicate samples within one plate were available for each chemical-concentration-hormone test. In some cases, a chemical may have appeared in multiple blocks of the study; in this case, the data for each block were analyzed separately due to the presence of block effects.

Per the OECD TG 456 (OECD, 2011) and the inter-laboratory validation report (Hecker et al., 2011), for a positive result, two consecutive concentrations (not necessarily including the top concentration) had to produce results significantly different from control for a steroid hormone analyte (only 8% of positive responses in the HT H295R assay did not include a significant maximum concentration). A positive result was also counted if the significant effect occurred only at the maximum concentration tested that still maintained ≥ 70% cell viability. A minimum efficacy threshold of a 1.5-fold change from DMSO control was applied for context as suggested by the OECD inter-laboratory analysis, as some results were deemed statistically significant by ANOVA but were still less than 1.5-fold different from DMSO control.

3.2.4 Computation of the Mean Mahalanobis Distance to Derive a Maximum Mean Mahalanobis Distance by Chemical

A statistical approach based on the Mahalanobis distance was employed herein to characterize the magnitude of change for 11 steroid hormones produced by H295R cells. A mean Mahalanobis distance (mMd) was calculated to summarize the 11 steroid hormone responses measured following exposure to each chemical concentration screened in the assay. Then, the maximum mean Mahalanobis distance (maxmMd) was selected from the set of mMd values generated for a chemical. The maxmMd then serves as a single numeric value to characterize the magnitude of effect on synthesis of 11 steroid hormones for a given chemical screened in the HT H295R assay. Below, the computation of the mMd and maxmMd are described, followed by a detailed description of the computation of the covariance matrix used to compute mMd values.

Calculation of the Mahalanobis Distance Metrics

A Mahalanobis distance is a generalization of Euclidian distance that adjusts for the variance and covariance among the hormone measures at each concentration screened (SAS, 2012; De Maesschalck et al., 2000). Although 13 hormones were measured in the HT H295R assay, measurements of two of these hormones frequently indicated a value below the LLOQ; pregnenolone and DHEA were often missing (53.1% and 69.5% of all measurements) and have been excluded from this approach, leaving 11 hormone measures for analysis. Thus, a Mahalanobis distance-based approach was used to indicate the effect of each test chemical concentration in 11-dimensional space.

To calculate the Mahalanobis distance, the response at each concentration of a test chemical was considered as a point in an 11-dimensional space; each axis corresponds to the natural logarithm of the measured concentration of one of the hormones included in this analysis, respectively. In brief, the degree to which variation among replicates is correlated across hormones was estimated, and a covariance matrix that characterizes both the noise variance and correlation among hormone levels across replicates, after taking chemical and concentration into account, was constructed. Conceptually, this is equivalent to rotating and scaling the hormone concentrations to a set of new variables that are uncorrelated with each other and have the same standard deviation, followed by computation of the Euclidean distance in this new space (**Figure 3-2**).

Due to the need to compare distances based on different numbers of hormone analyte data for a given test chemical (*e.g.*, due to missing data), a mean Mahalanobis distance (mMd) statistic was computed), *i.e.* the Mahalanobis distance divided by the square root of the number

of hormones used to compute it (see Eq. 8, below). The mMd for a given test compound between the hormone concentration at the c^{th} concentration relative to that at the DMSO vehicle control concentration was computed as:

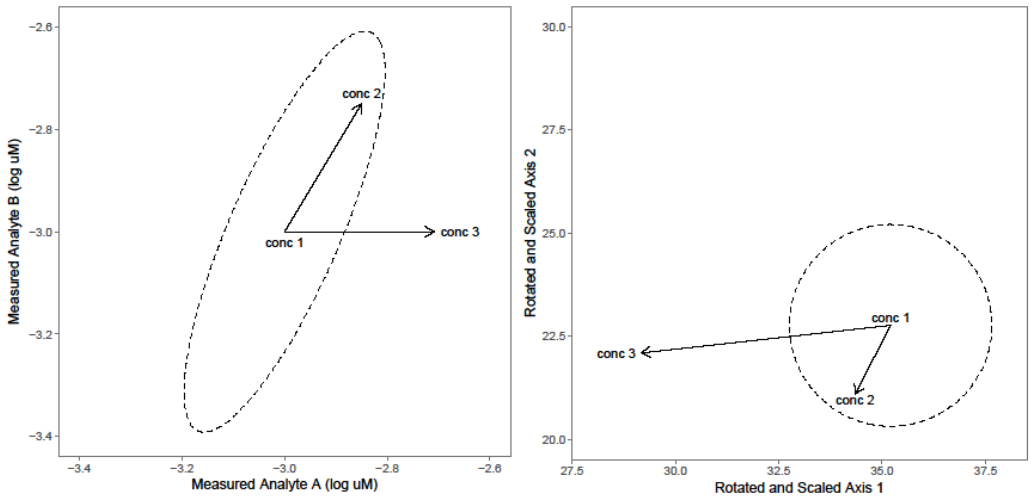
Equation 8

$$mMd = \sqrt{(\boldsymbol{y_c} - \boldsymbol{y_1})^T \boldsymbol{\sum^{-1}} (\boldsymbol{y_c} - \boldsymbol{y_1})/N_h}$$

For this analysis, y is the vector of natural log-transformed steroid hormone concentrations at the c^{th} concentration, y_1 is the vector of natural log-transformed steroid hormone concentrations for the DMSO control, N_h is the number of hormones with measurements for this chemical, and $\sum$ is the estimate of the covariance matrix.

The maximum mMd (maxmMd) is the maximum of the set of mMd values computed for all concentrations of a test chemical.

Figure 3-2 Illustrating the Difference Between Mahalanobis Distance and Euclidean Distance



In the left panel, the three points represent the mean (log10) concentrations of hormones A and B at three concentrations of a test chemical, labeled conc 1, conc 2, and conc 3 in increasing order. In terms of hormone concentrations, the response at conc 3 for hormone A is twice as far from that at conc 1 as is the response at conc 2; however, the Euclidean distances of conc 2 and conc 3 to conc 1 are the same. The ellipse around the response at conc 1 represents the error distribution of both hormones. The variance among measurements (at the same concentration of test compound) of hormone B is greater than that among measurements of hormone A, and the error of measurements of hormones A and B are correlated. In this example, the standard deviation of hormone A around its mean is 0.08, that of hormone B is 0.16, and the correlation is 0.8. Therefore, in this case conc 3 is a greater number of standard deviations away

from conc 1 compared to conc 2. The plot on the right shows all the points transformed so that the new variables are uncorrelated and have the same standard deviation. The error "ellipse" is transformed to a circle. The distances between points in this space are the Mahalanobis distances. In this example, the response at conc 3 is four times as far from that at conc 1 as is the response at conc 2.

Critical value for Positive Steroidogenesis Pathway Results Using the mMd

A critical value to assess significance was derived to distinguish mMd values that are greater than what would likely result from sampling noise. The critical value accounts for the multiple comparisons arising from comparing each concentration group to the control. The critical value reflects the similarity between mMd and the Hotelling T^2 statistic used to compare two groups with multiple measurements (Mardia et al., 1979). Hotelling's T^2 is analogous to the usual t- or F-statistics used for comparisons of single characteristics in that T^2 evaluates the difference between two groups (*i.e.* the response of one concentration compared to that of its plate DMSO control) relative to the variability among measurements within the groups. Instead of simply computing the variance within the groups, as would be performed for a univariate response, a variance-covariance matrix was computed and accounts for the variation and covariation of the multiple steroid hormone measurements in the HT H295R assay (described in the next section). For this analysis, all the test chemical concentrations and control groups were used to determine this within-group variability. This yields an estimate of the within-group variance-covariance matrix which is more precise than the one that would be used for T^2. With the variance-covariance matrix known, we employed the method developed by (Nakamura and Imada, 2005) to adjust for multiple comparisons for multivariate tests. This is analogous to adjusting for multiple comparisons for univariate tests such as Dunnett's procedure. Nakamura and Imada's method requires equal sample sizes across comparison groups, so a critical value for the set of mMd values for a test chemical was derived by assigning the sample size for a concentration group as the largest of the sample sizes across hormones evaluated in that group, and the sample size for all the comparisons for a given test chemical as the median sample size across concentration groups. The critical value was derived for a nominal Type I error of 0.01. Because of the sample size decision just described, and the fact that the covariance matrix is estimated, even though the sample used was large, this approach should only approximate the actual Type I error. The resulting critical value for the mMd varied across the set of chemicals as the critical value is related to the number of hormones with data for each chemical. The critical values ranged from 1.15 to 1.81, with a median of 1.64 and a mean of 1.58, for all of the chemicals with available data for mMd computation.

Any observed mMd value for a chemical exceeding the critical value was considered a positive for potential steroidogenesis pathway disruption. The maxmMd was adjusted for the critical value (maxmMd – critical value = adjusted maxmMd) to ensure that the maxmMd exceeded the critical value; this difference should be greater than zero for a positive pathway result.

3.2.5 MANOVA and Computation of the Covariance Matrix

The steroid hormone responses measured in the HT H295R assay represent a multivariate response, and as such, a variance-covariance matrix was computed to account for the variation and covariation of the multiple steroid hormone measurements. An estimate of the covariance matrix that characterizes both the noise variance and correlation among measured steroid hormone concentrations across replicates, after taking chemical and concentration into account, was needed to compute the mMd as indicated above. Due to the presence of block effects between chemicals assayed on different days, separate covariance matrices were computed for each assay date, resulting in a total of eight individual covariance matrices. The covariance matrix used in the mMd computation was constructed per the following procedure:

- A multivariate linear model of the unique set of chemicals within each block was fit using the natural log-transformed hormone concentrations from the HT H295R assay. The model includes terms for plate-specific values for all DMSO controls, and a separate mean for each test chemical concentration across all the measured steroid hormone analytes. All these entities were replicated on the same plate. Outlier detection was performed by fitting all data to the multivariate linear model and identifying where the standard deviation of the residuals for a chemical-concentration technical replicate pair was greater than one for any steroid hormone analyte measured (indicating a ~2.7 fold-change difference in steroid hormone concentration between technical replicates). This resulted in the removal of 18 of 4655 unique chemical-concentration replicate pairs. The matrix of residuals from the fits of the filtered data across all the plates within each block were used to estimate a variance and covariance matrix.
- To retain estimates for the largest possible number of chemicals and to keep the estimation process simple, if data for a particular hormone were missing for a chemical within a block, the hormone measure was dropped from that block prior to linear model fitting. This affected only one of the eight blocks, which contained some missing data for estrone and E2, representing 81 unique test chemicals. In this case, the computed covariance matrix for this block included only nine of the 11 steroid hormone analytes.

- The full pooled 11 × 11 covariance matrix (omitting DHEA and pregnenolone) used for the mMd calculation was estimated as the unweighted average of the eight block-specific covariance matrices.

The resulting pooled covariance matrix was positive-definite (a requirement for a proper covariance matrix).

3.2.6 Comparison Methodology for HT H295R to OECD Reference Chemicals

Chemicals with comparable data for comparison

Ten of the 12 core reference chemicals included in the OECD H295R inter-laboratory validation study (Hecker et al., 2011) have been screened using the HT H295R assay, including: aminoglutehimide, atrazine, benomyl, butylparaben, ethylene dimethanesulfonate, forskolin, letrozole, molinate, nonoxynol-9 (Polyoxyethylene(10) nonylphenyl ether), and prochloraz (**Table 3-3**). Trilostane and a protein hormone, human chorionic gonadotropin, have not been screened in the HT H295R assay. In addition to the 12 core chemicals for reference, 16 chemicals were included as "supplemental" verification for the inter-laboratory validation, with testing in only 2 laboratories in the OECD inter-laboratory validation instead of 5 laboratories (Hecker et al., 2011). These data have a greater amount of uncertainty than the core reference chemicals due to disagreements reported between the two testing laboratories. Fifteen of these 16 chemicals have HT H295R data for comparison (**Table 3-3**).

One of the ten core reference chemicals with data for comparison, "nonoxynol-9," has presented some uncertainties with respect to the nature and concentration of the substance tested in the OECD inter-laboratory validation. Nonoxynol-9, as defined by CASRN 26027-38-3, is an unknown, variable composition, biological (UVCB) substance. The composition and representative molecular weight used in the underlying Hecker study was not reported. As such, EPA is unsure of the precise molecular weight and composition of the nonoxynol-9 used in the original OECD validation studies. Therefore, it is unclear if the substance, and the nominal concentration tested, are comparable between the OECD inter-laboratory validation study and the HT H295R screening. This uncertainty is further supported by discrepancies between the OECD inter-laboratory validation report and the HT H295R screening for cytotoxicity. Although there was variability among labs, in the OECD inter-laboratory validation study, cell viability appeared to range from 80-100% at 1 µM, and from 25-100% at 10 µM (interpolated from graphs) (Hecker et al., 2008). Due to cytotoxicity, the MTC for nonoxynol-9 in the HT H295R assay was 0.4 µM. It is unknown if these differences in cytotoxicity are due to variability in

testing between the assay systems, or due to differences in the composition and/or computed concentration of the substance.

Table 3-3 DSSTox Reference Information for the Chemicals Used for Comparison of OECD and HT H295R approaches

DTXSID	Preferred Name	HT H295 R Data	CASRN	INCHI Key	Average MW
Core Reference Chemicals					
DTXSID8022589	Aminoglutethimide	Y	125-84-8	ROBVIMPUHSLWNV-UHFFFAOYSA-N	232.28
DTXSID9020112	Atrazine	Y	1912-24-9	MXWJVTOOROXGIU-UHFFFAOYSA-N	215.69
DTXSID5023900	Benomyl	Y	17804-35-2	RIOXQFHNBCKOKP-UHFFFAOYSA-N	290.32
DTXSID3020209	Butylparaben	Y	94-26-8	QFOHBWFCKVYLES-UHFFFAOYSA-N	194.23
DTXSID40196931	Ethylene dimethanesulfonate	Y	4672-49-5	QSQFARNGNIZGAW-UHFFFAOYSA-N	218.24
DTXSID8040484	Forskolin	Y	66575-29-9	OHCQJHSOBUTRHG-KGGHGJDLSA-N	410.51
DTXSID4023202	Letrozole	Y	112809-51-5	HPJKCIUCZWXJDR-UHFFFAOYSA-N	285.31
DTXSID6024206	Molinate	Y	2212-67-1	DEDOPGXGGQYYMW-UHFFFAOYSA-N	187.30
DTXSID2036588	Nonoxynol	Y	26027-38-3	NA	NA
DTXSID4024270	Prochloraz	Y	67747-09-5	TVLSRXXIMLFWEO-UHFFFAOYSA-N	376.66
DTXSID9023706	Trilostane	N	13647-35-3	KVJXBPDAXMEYOA-CXANFOAXSA-N	329.44
DTXSID4036770	Human chorionic gonadotropin	N	NA	NA	NA
Supplemental Reference Chemicals					
DTXSID0020523	2,4-Dinitrophenol	Y	51-28-5	UFBJCMHMOXMLKC-UHFFFAOYSA-N	184.11
DTXSID7020182	Bisphenol A	Y	80-05-7	IISBACLAFKSPIT-UHFFFAOYSA-N	228.29
DTXSID2022880	Danazol	Y	17230-88-5	POZRVZJJTULAOH-LHZXLZLDSA-N	337.46
DTXSID5020607	Di(2-ethylhexyl) phthalate	Y	117-81-7	BJQHLKABXJIVAM-UHFFFAOYSA-N	390.56
DTXSID7020479	Dimethoate	Y	60-51-5	MCWXGJITAZMZEV-UHFFFAOYSA-N	229.25
DTXSID2032390	Fenarimol	Y	60168-88-9	NHOWDZOIZKMVAI-UHFFFAOYSA-N	331.20
DTXSID3020625	Finasteride	Y	98319-26-7	DBEPLOCGEIEOCV-WSBQPABSSA-N	372.55
DTXSID7032004	Flutamide	Y	13311-84-7	MKXKFYHWDHIYRV-UHFFFAOYSA-N	276.21
DTXSID5022308	Genistein	Y	446-72-0	TZBJGXHYKVUXJN-UHFFFAOYSA-N	270.24
DTXSID1024122	Glyphosate	N	1071-83-6	XDDAORKBJWWYJS-UHFFFAOYSA-N	169.07
DTXSID7029879	Ketoconazole	Y	65277-42-1	XMAYWYJOQHXEEK-OZXSUGGESA-N	531.43
DTXSID5023322	Mifepristone	Y	84371-65-3	VKHAHZOOUSRJNA-GCNJZUOMSA-N	429.60
DTXSID1021166	Piperonyl butoxide	Y	51-03-6	FIPWRIJSWJWJAI-UHFFFAOYSA-N	338.44

DTXSID	Preferred Name	HT H295 R Data	CASRN	INCHI Key	Average MW
DTXSID6022341	Prometon	Y	1610-18-0	ISEUFVQQFVOBCY-UHFFFAOYSA-N	225.30
DTXSID6034186	Spironolactone	Y	52-01-7	LXMSZDCAJNLERA-ZHYRCANASA-N	416.58
DTXSID4021391	Tricresyl phosphate	Y	1330-78-5	NA	NA

Interpretation of the OECD Inter-Laboratory Validation Results

E2 and T were measured as biomarkers of estrogen and androgen biosynthesis, respectively. These data were analyzed per OECD TG 456 (Hecker et al., 2011; OECD, 2011). For normally distributed data, an analysis of variance (ANOVA) was performed and differences from vehicle control were evaluated using a Dunnett's test. For data that were not normally distributed, as evaluated by standard probability plots or Shapiro-Wilk's test, a Kruskal-Wallis test followed by a Mann-Whitney *U* test was employed (see details in (see details in Hecker et al., 2011). These data are summarized in (Hecker et al., 2011) as part of the OECD inter-laboratory validation study (OECD, 2010b), and were extracted for this comparison. A lowest effect concentration (LOEC) was reported for each laboratory. However, there was an error in the published work, and these LOECs from Tables 3 and 4 of (Hecker et al., 2011) were really in µM units (not µg/mL as reported; confirmed via personal communication with Dr. M. Hecker and by raw data in the submitted validation study report). If no LOEC was reported, the LOEC was assigned a value of 'not detected' (ND). E2 and T were annotated as being increased (up) or decreased (dn). For the core chemicals, in the event that the results of ≥ 2 of the 5 laboratories qualitatively disagreed, an effect on E2 or T was considered equivocal. For the 16 supplemental chemicals, a response was considered equivocal if the anticipated response failed to match qualitatively between the two laboratories.

Interpretation of the HT H295R Results

E2 (assay component CEETOX_H295R_ESTRADIOL) and T (assay component CEETOX_H295R_TESTO) were used as biomarkers of estrogen and androgen biosynthesis, respectively. The data used for this comparison were analyzed by ANOVA as described above. Per the procedure in (Hecker et al., 2011), chemicals were indicated as positives, but were flagged accordingly, if they fell into any of the following categories: (1) effects were seen at only the maximum concentration; (2) effects were observed for a minimum of two consecutive concentrations, but with the highest concentration corresponding to a loss in cell viability; (3) effects were seen at two consecutive concentrations, but no effect was seen at the highest concentration tested; or (4) positive effects were seen, but they were within 1.5-fold of control.

Calculation of Confusion Matrices

Confusion matrices were constructed for E2 and T for increased and decreased production, using the OECD inter-laboratory validation results (Hecker et al., 2011) (see Tables 3 and 4 of the publication) as the source of 'true' positives and negatives. The HT H295R assay data, analyzed by an ANOVA and post-hoc Dunnett's procedure, along with the OECD logic used for positive responses (Hecker et al., 2011; Hecker et al., 2008), were used for comparison. Equivocal data from the OECD inter-laboratory validation results for the specific effect type were excluded from the calculation of sensitivity, specificity, and accuracy; increased and decreased T and increased and decreased E2 sets excluded 4, 1, 4, and 2 equivocal results, respectively, yielding 21, 24, 21, and 23 chemicals total in the analysis of these effect types. A set of revised confusion matrices and associated sensitivity, specificity, and accuracy values were also generated following removal of nonoxynol-9 (due to uncertainties in the substance evaluated for the OECD inter-laboratory validation) from all effect types and letrozole from decreased T, leaving 20, 22, 20, and 22 chemicals for increased and decreased T and E2, respectively. The sensitivity or true positive rate was calculated as:

Equation 9

$$\frac{true\ positives}{true\ positives + false\ negatives}$$

The specificity or true negative rate was calculated as:

Equation 10

$$\frac{true\ negatives}{true\ negatives + false\ positives}$$

And finally, the accuracy was calculated as:

Equation 11

$$\frac{true\ positives + true\ negatives}{total\ number\ of\ chemicals\ for\ effect\ type}$$

3.3 Results

The results of this study include the analyzed hormone concentration-response outputs using significance testing by ANOVA and post-hoc Dunnett's results for all chemicals assayed in HT H295R assay, a comparison of the results for chemicals included in the OECD inter-

laboratory validation and HT H295R assay, and the pathway-based results from computation of the maximum mean Mahalanobis distance (maxmMd) for each concentration of each chemical.

3.3.1 Analysis of HT H295R Data by ANOVA and Post-Hoc Dunnett's Test

An ANOVA and post-hoc Dunnett's test was completed for raw hormone quantification data (converted to µM units) collected for 766 chemicals samples, comprised of 654 unique chemicals with concentration-response data. The complete results of this analysis are provided in Supplemental File 4 as a table of the p-values from the ANOVA procedure. Supplemental File 5 contains a summary of the significant effects of a chemical sample for each hormone, denoted as a zero for no effect or a one for a significant effect. Supplemental File 6 contains binary strings that represent the significant effects ($p \leq 0.05$) by concentration for each chemical-steroid hormone analyte pair. These binary strings were used to determine when significant effects were observed for a given hormone, *i.e.* when two consecutive concentrations demonstrated significant effects, or if a significant effect was demonstrated at the top concentration only, a chemical was labeled as a "positive" response for a particular steroid hormone analyte. The complete graphical results are presented in Supplemental File 7, with concentrations that demonstrated a significant effect of treatment colored red, and dotted horizontal lines demarcating the 1.5-fold control boundaries.

The number of positive chemicals, and the positive percentage of the library tested in concentration response, are summarized in **Table 3-4**. The relatively high rate of hits for the chemical library (for many steroids exceeding 50% of the tested library) screened in concentration-response was expected, as chemicals screened in concentration-response were selected predominantly from positive responses in single concentration screening (with positive responses for ≥ 3 steroid hormones for approximately 80% of the chemicals screened in concentration-response). All of the p-values by steroid hormone analyte for each comparison of concentrations for a chemical, and binary assessment of the positive/negative behavior of each chemical for each steroid hormone analyte, are presented in Supplemental Files 3 and 4. An example of the ANOVA results for the prototypical pathway inhibitor, prochloraz, are presented in **Of 654** chemical samples, positive hit rates for the 11 hormones used in this analysis ranged from 57.5 to 85.9%.

Figure 3-3. The high positive rate (**Table 3-4**) was further explored via determination of the correlation of the effects and residuals between steroid hormone analytes, discussed in subsequent explanation of the Mahalanobis distance results.

The results of the ANOVA analysis for all steroid hormone analyte data were also considered in terms of how each chemical may have affected different hormone classes across the steroid biosynthetic pathway, *i.e.* progestagen, glucocorticoid, androgen, or estrogen production. Considering these steroid hormone classes (highlighted in **Figure 3-1**), the results for the 654 chemicals evaluated are represented in a Venn diagram (**Figure 3-4**) to illustrate the number of chemicals that affected each hormone class or combination of classes. Of the 654 chemicals with concentration-response data amenable to ANOVA, 25 chemicals failed to produce a positive result on any hormone; the remaining 628 chemicals produced a positive result on at least one hormone class. Three hundred five chemicals, or 47% of chemicals tested in concentration-response, demonstrated positive results for at least one hormone from each of the four classes. This finding is not unexpected, as chemicals evaluated in concentration-response were largely pre-selected for demonstrated effects in single concentration screening for 3 to 4 hormone analytes. Interestingly, few chemicals affected only estrogens (estrone and E2; 9 chemicals) or androgens (androstenedione and T; 1 chemical), or both (1 chemical), even though 4 hormone analytes comprise these two classes combined. Due to the relatively high percentage of the screened chemicals that affected androgens or estrogens in addition to glucocorticoid and/or progestagens, it appears that integration of data for the glucocorticoid and progestagen hormone measurements with the data for estrogen and androgen hormone measurements may provide important information on the magnitude of overall steroid biosynthetic pathway perturbation. Sixty-eight chemicals, or approximately 10% of the chemicals screened in concentration-response, affected progestagens only (13), glucocorticoids only (10), or progestagens and glucocorticoids only (45). Thus, consideration of glucocorticoid and progestagen hormone levels in the HT H295R assay may identify chemicals that perturb portions of the steroid biosynthesis pathway expressed in H295R cells that are overlooked in the H295R assay when only E2 and T are reported.

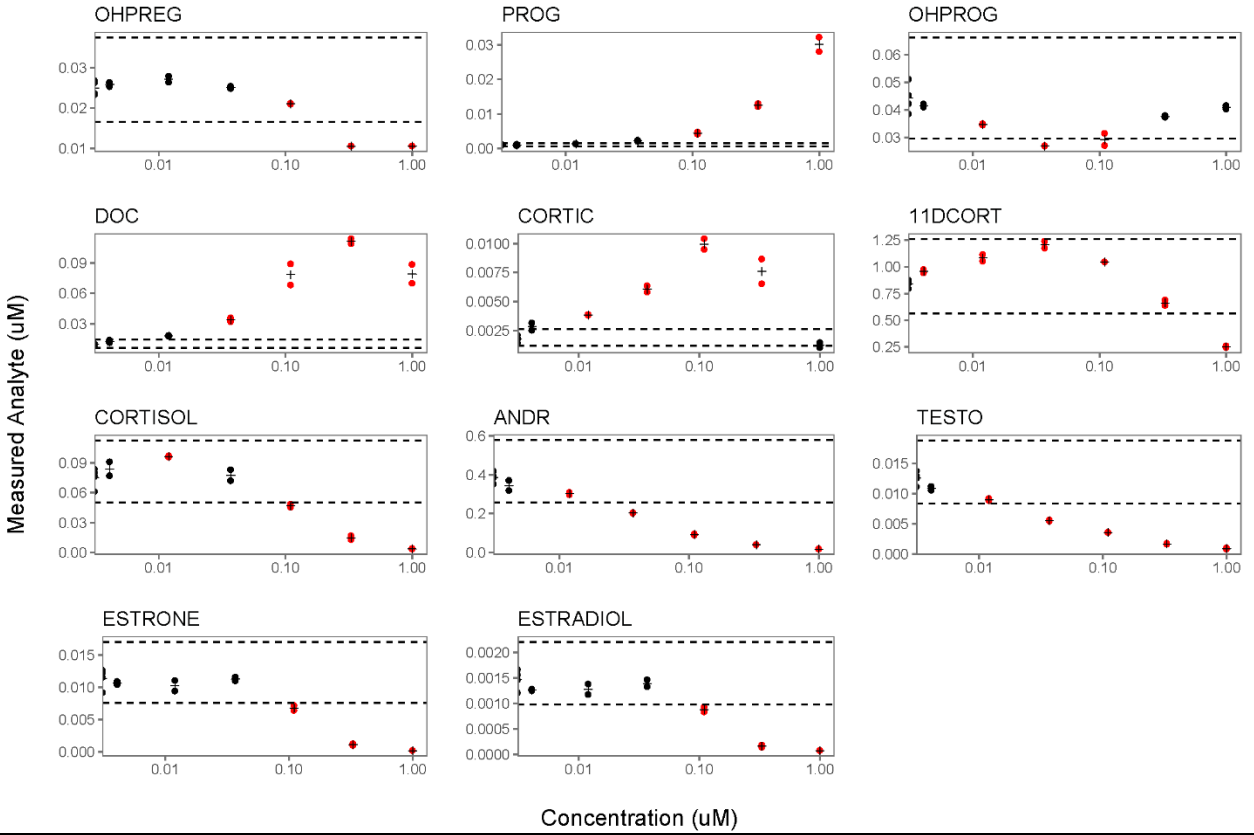
Table 3-4 Positive ANOVA (Plus Post Hoc Dunnett's Test) Results by Steroid Hormone Analyte

#	Steroid Hormone Analyte	Abbreviation	# Positive Chemical Samples	% of Tested Library
1	OH-Pregnenolone	OHPREG	387	59.2
2	Progesterone	PROG	509	77.8
3	OH-Progesterone	OHPROG	562	85.9
4	DOC	DOC	511	78.1
5	Corticosterone	CORTIC	386	59.0
6	11-deoxycortisol	11DCORT	504	77.1

7	Cortisol	CORTISOL or CORT	376	57.5
8	Androstenedione	ANDR	438	67.0
9	Testosterone	TESTO or T	397	60.7
10	Estrone	ESTRONE or E1	425	65.0
11	Estradiol	ESTRADIOL or E2	408	62.4

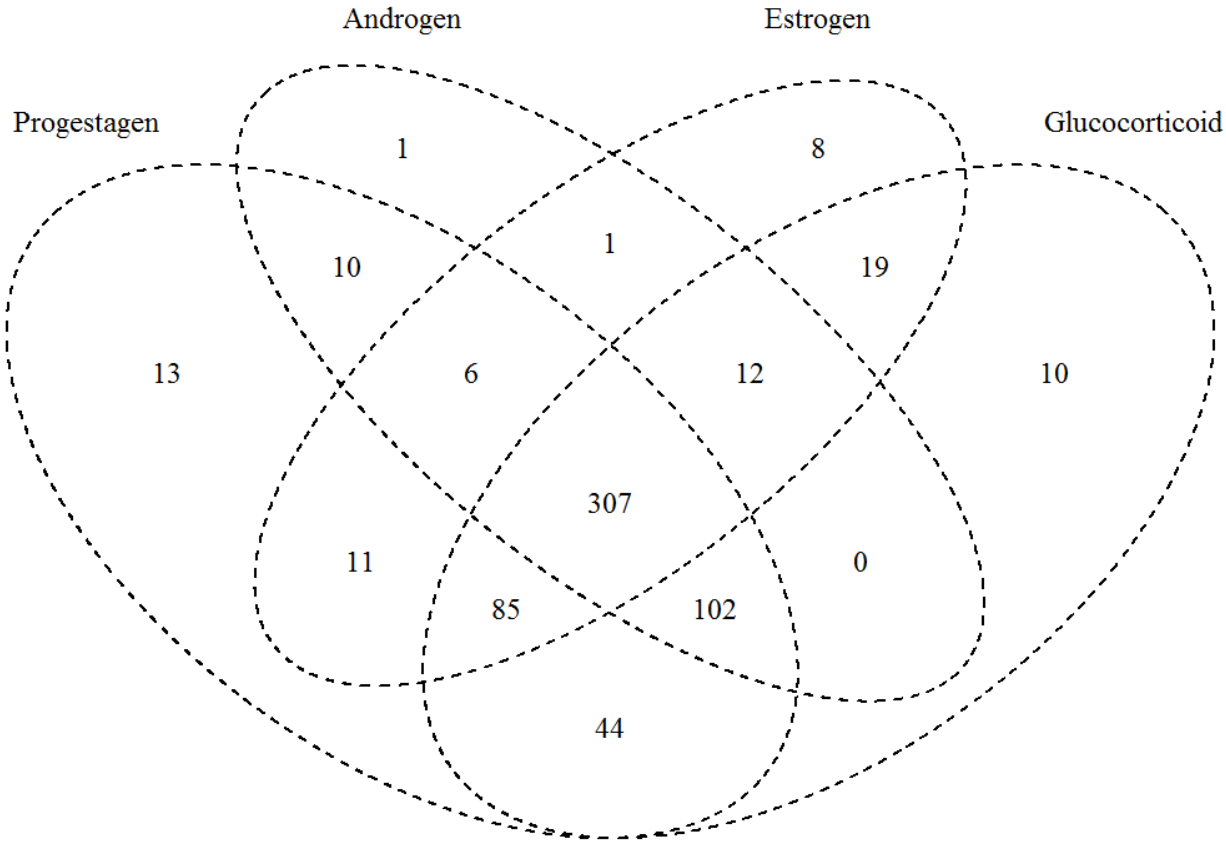
Of 654 chemical samples, positive hit rates for the 11 hormones used in this analysis ranged from 57.5 to 85.9%.

Figure 3-3 Example Visualizations of the ANOVA and Post Hoc Testing Results for Prochloraz



Replicates and the mean response value are denoted as filled circles and plus signs, respectively. Symbols in red indicate data points that were significant ($p<0.05$). Dashed horizontal lines indicate ± 1.5-fold vs. DMSO control to give additional context for low magnitude, but positive, responses. Data are plotted as concentration (µM) of prochloraz versus the measured steroid hormone analyte concentration (µM).

Figure 3-4 Venn Diagram of ANOVA Results for Effects on Steroid Hormone Synthesis, Grouped by Steroid Class



The number of chemicals with positive results for progestagens (OH-pregnenolone, progesterone, OH-progesterone), glucocorticoids (DOC, Corticosterone, 11-deoxycortisol, Cortisol), androgens (androstenedione, T), and estrogens (estrone, E2) are shown. A total of 629 chemical samples are represented in the Venn diagram (24 chemicals tested in concentration-response with data available for analysis failed to produce positive ANOVA results for any hormone class).

3.3.2 Pathway-Based Results Using the Mahalanobis Distance Approach

The Mahalanobis distance adjusts the distances, or effect sizes, for the variance and covariance among the hormone measures at each concentration, thereby accounting for knowledge of the interrelatedness of the steroid hormone measurements (**Figure 3-2**). The use of Mahalanobis distance rather than Euclidean distance allows for correction of the covariance matrix to account for the correlation among the residuals of the hormone responses for each chemical-concentration pair. To support selection of the Mahalanobis distance as a basis for the

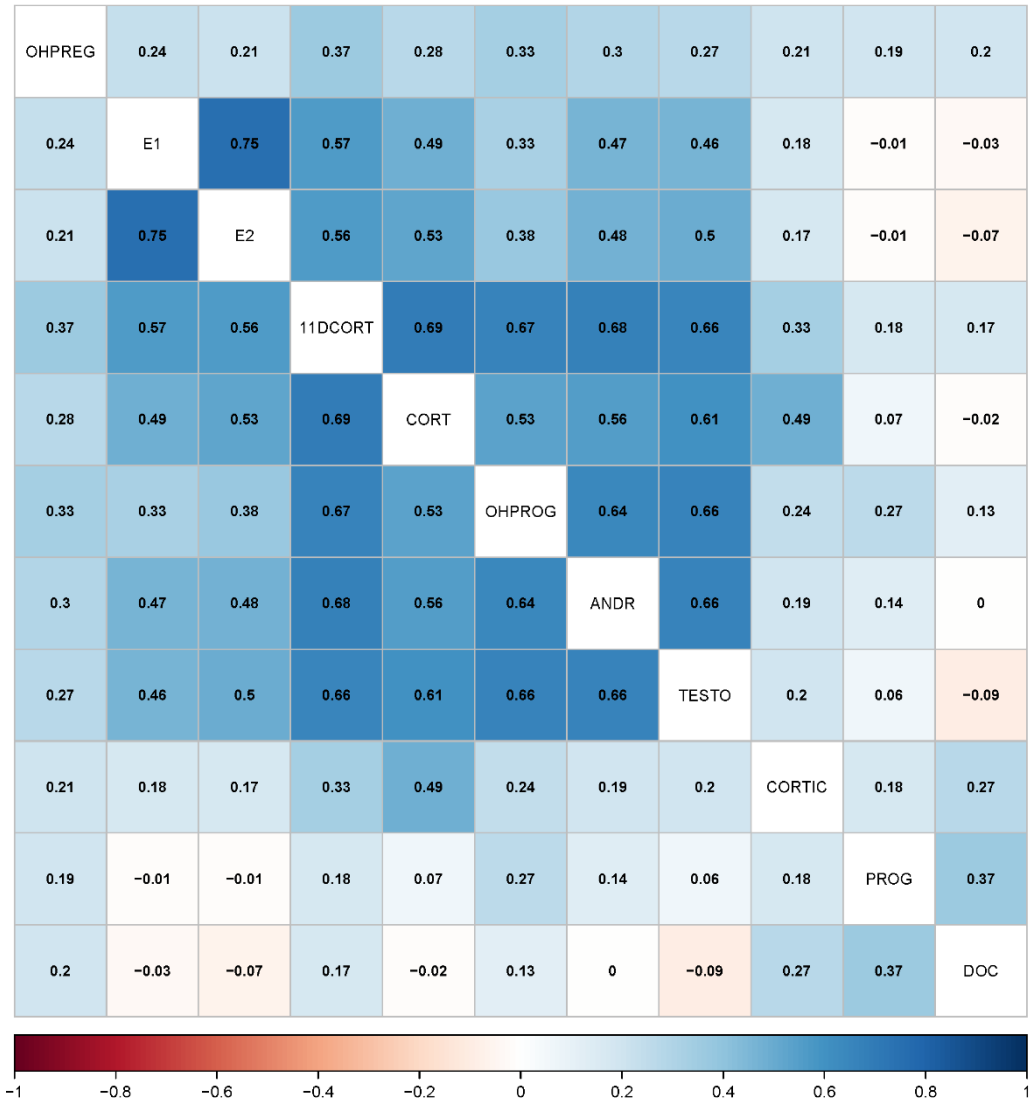
new statistical approach, the correlation matrix corresponding to the covariance matrix used in calculation of the mMd for the steroid hormone analytes was examined. As anticipated from knowledge of the steroidogenesis pathway in H295R cells (**Figure 3-1**), the covariance of the residuals for several steroid hormone analytes in the HT H295R assay were highly correlated with one another (**Figure 3-5**). For example, the residuals for estrone and E2 were highly correlated (Pearson's R = 0.75), as were androstenedione and T (R = 0.66). Residuals for cortisol and 11-deoxycortisol were also highly correlated (R = 0.69). In contrast, the residuals for both progesterone and DOC had very weak correlations, in some cases negative correlations, with residuals for all of the steroid hormones measured. This correlation matrix demonstrated high correlation of the residuals of many of the steroid hormone measures, which suggests that the Mahalanobis distance is one appropriate analysis metric for interpretation of these data.

The results from measurement of 11 steroid hormone analytes were used in the derivation of the mean Mahalanobis distance (mMd) at each concentration for chemicals screened in concentration-response. Radar plots were used to visualize the response for a single chemical for these 11 assayed hormones, with examples for atrazine, benfluralin, and mifepristone illustrated in **Figure 3-6** (radar plots for all tested chemicals available in Supplemental File 8). Next to the example radar plots in **Figure 3-6**, the plot of the estimated mMd by concentration is shown, with a horizontal red dashed line to indicate the critical limit. If a mMd exceeds the critical limit, it is considered a positive result for this pathway approach. The maxmMd is the maximum of the set of mMd values produced for all concentrations of a tested chemical. Atrazine moderately affected a number of hormones, including estrogens, progestagens, glucocorticoids, and androgens, yielding a moderate adjusted maxmMd of 3.14. Benfluralin provides an example of a chemical with a negative pathway result, as the maxmMd failed to exceed the critical limit (adjusted maxmMd of -0.14). In contrast to the moderate effects of atrazine on multiple steroid hormones, mifepristone strongly modulated progestagens with significant effects on progesterone and OH-progesterone and moderate but non-significant trends on glucocorticoids and androgens, resulting in a relatively high adjusted maxmMd of 33. The steroid hormone response data, annotated by the ANOVA results, and plots of the mMd for all tested chemicals are available as Supplemental File 7.

To provide context for the relative maxmMd responses, the distribution of the maxmMd values for the 766 chemical samples with concentration-response data that cleared the cell viability filter is illustrated in **Figure 3-7**. These maxmMd values are adjusted for the critical value (maxmMd – critical value = adjusted maxmMd), such that a positive maxmMd should be

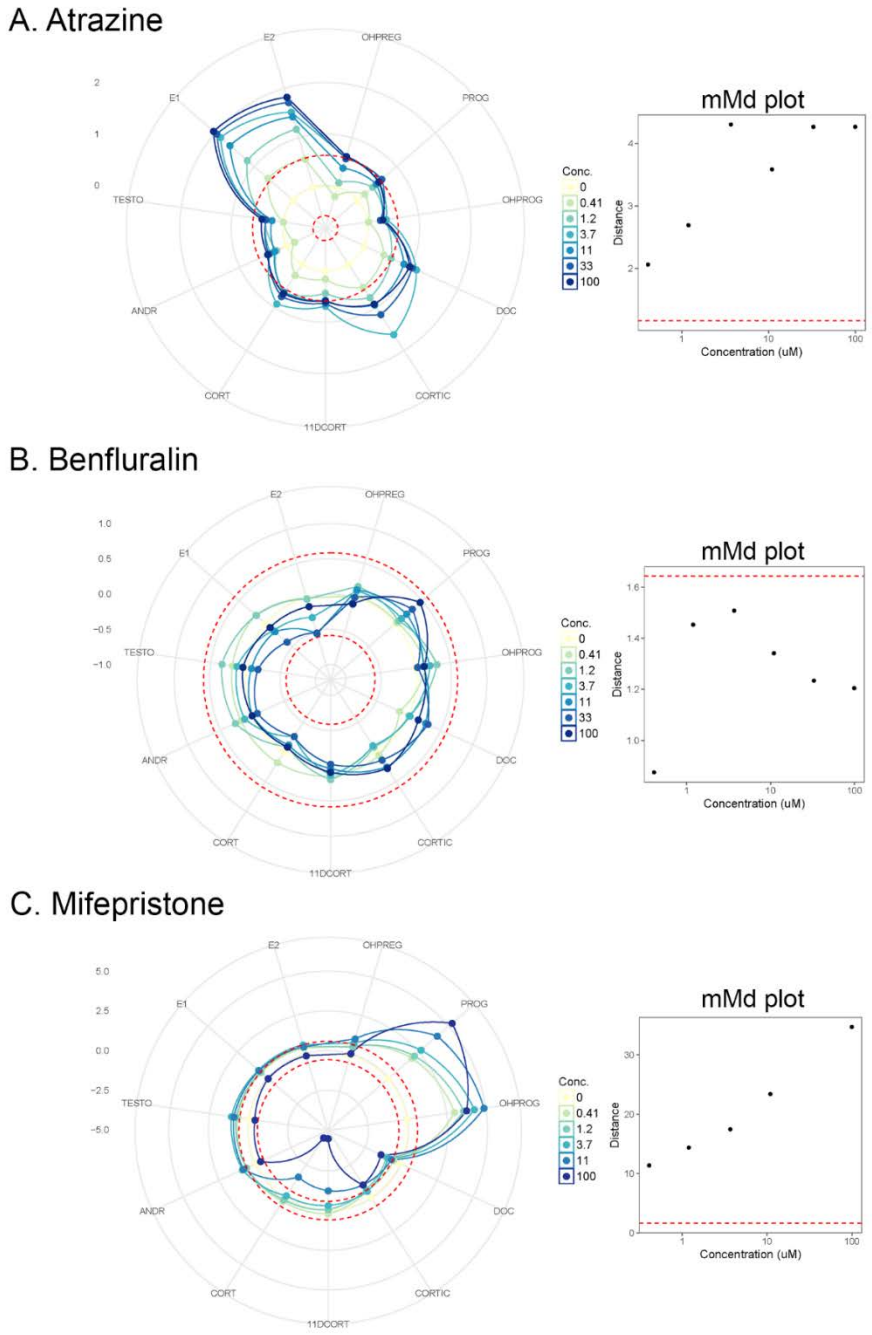
greater than zero. The range of adjusted-maxmMd values for this dataset is -0.64 – 51.8. The median of the distribution, 3.52, is annotated by a vertical dashed red line. The mean of the distribution was 5.92. The distribution would likely be more informative if the chemical set had not been pre-selected predominantly from single concentration screening for positives; however, the relative rank within this distribution appears to be informative based on stratification of the limited number of OECD reference chemicals, with the additional caveat that these reference chemicals have typically only been evaluated for effects on E2 and T. A comparison of the maxmMd for all of the OECD reference chemicals as well as a comparison of the steroid hormone positive hit count versus maxmMd value are presented in the next section. All of the maxmMd values, the critical values, and the adjusted maxmMd values are provided by chemical sample in Supplemental File 9.

Figure 3-5 Heatmap Summarizing Correlation of Steroid Hormone Analyte Responses and Residuals



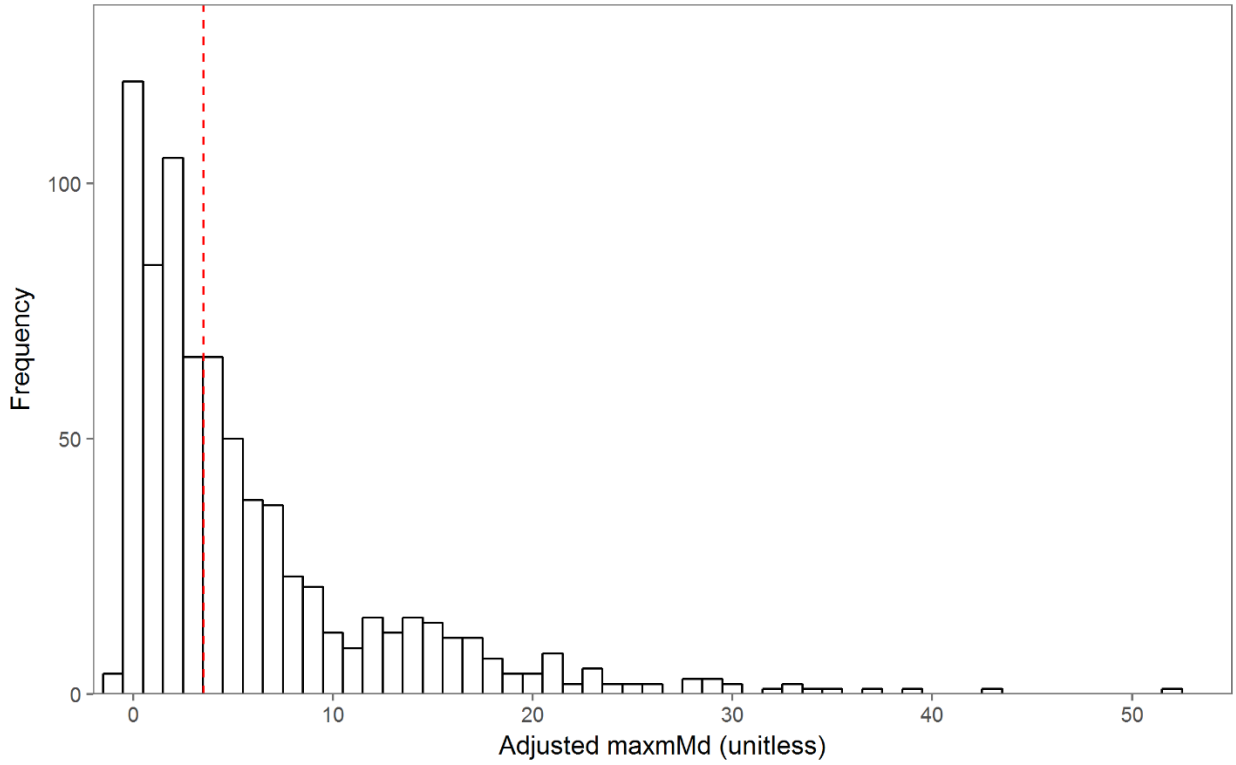
The correlation plots are provided for the correlation matrix corresponding to the covariance matrix values for each steroid hormone.

Figure 3-6 Example Radar Plots of the 11-Dimensional Dataset Used to Derive a Mean Mahalanobis Distance (mMd) for Each Concentration Assayed



The 11 steroid hormone analytes are represented as the 'spokes' of the radar plot, and each concentration of the chemical is annotated by a different color. The dotted, concentric circles denote ± 1.5-fold control as threshold to contextualize the responses, as the y-axes vary by chemical to allow for visualization of the relative magnitude of effects. The numbers on the left of each radar plot denotes the fold change values of the major gridlines of the plots. Next to each radar plot is a plot of mMd by concentration, with the critical limit for the mMd annotated using a horizontal dashed red line. A. atrazine (CASRN 1912-24-9); B. benfluralin (CASRN 1861-40-1); C. mifepristone (CASRN 84371-65-3). Radar plots and mMd plots are supplied for all chemicals in Supplemental File 8.

Figure 3-7 Frequency Distribution of the maxmMd Values for the 655 Chemicals with Concentration-Response HT H295R Screening Data



The maximum mMd (maxmMd) values were adjusted for the critical value (*i.e.*, the maxmMd – critical value for the specific chemical). The median of the distribution of adjusted maxmMd values (3.52) is denoted by a red, dashed vertical line. The range of the distribution is -0.64 – 51.8.

3.3.3 Comparison and Evaluation of the ANOVA and maxmMd Results

Comparison of the HT H295R data with the OECD inter-laboratory validation results

Utilizing an ANOVA procedure and a post-hoc Dunnett's test enabled a comparison of the HT H295R screening data with the summary results available from the OECD inter-laboratory validation (Hecker et al., 2011). A detailed comparison of the effects on estrogen synthesis and androgen synthesis is illustrated in Supplemental File 10 Tables A and B, respectively, and summarized by confusion matrices and a table of sensitivity and specificity values by effect type in **Figure 3-8**. For the confusion matrix, a chemical was excluded from the sensitivity and specificity calculations if the OECD inter-laboratory validation results for E2 or T in a particular direction were equivocal. OECD inter-laboratory results for a chemical were considered equivocal if there was significant disagreement among labs, as specified here: (1) two or more laboratories failed to detect a LOEC for a "core" reference chemical tested in all five laboratories; or, (2) if only one of two laboratories reported a LOEC for the "supplemental"

reference chemicals that were tested in only two labs. A revised confusion matrix along with sensitivity, specificity, and accuracy values were also generated based on exclusion of one chemical, nonoxynol-9, from all effect types, and letrozole from decreased T (**Figure 3-8**).

Confusion matrices summarizing the comparison of OECD inter-laboratory validation results and the HT H295R screening data analyzed by ANOVA, excluding the OECD inter-laboratory equivocal results by effect type, demonstrated sensitivities of 0.75 and 0.80, specificities of 0.85 and 0.94, and accuracies of 0.81 and 0.91 for increased and decreased estradiol, respectively (**Figure 3-8**). For T synthesis, sensitivities of 1 and 0.55, specificities of 0.90 and 0.92, and accuracies of 0.90 and 0.75 were observed for increased and decreased T, respectively. Revision of the confusion matrices to exclude nonoxynol-9 and letrozole (from decreased T only) increased the sensitivity for decreased T to 0.67. It should be noted that the reference chemical sets were not balanced, with strong weighting toward true negatives and limited true positives. True positives ranged from only > 5% to approximately 29% of the result sets used for the confusion matrices. Further, inclusion of the supplemental reference chemicals, tested in only two laboratories for the OECD inter-laboratory validation, was complicated by additional equivocal findings due to discordance between labs.

Qualitative comparison of the effects of the OECD reference chemicals on E2 synthesis in both the OECD inter-laboratory validation and HT H295R assay demonstrated good concordance (Supplemental File 10). For increased E2 for the core reference chemicals, one chemical had equivocal findings (butylparaben), and of the remaining nine chemicals, eight chemicals agreed (aminoglutehimide, atrazine, benomyl, forskolin, letrozole, molinate, nonoxynol-9, and prochloraz). For decreased E2 for the core reference chemicals, there were no equivocal findings, and eight of the ten chemicals agreed (atrazine, benomyl, butylparaben, ethylene dimethanesulfonate, forskolin, letrozole, molinate, and prochloraz). Five of the 15 supplemental reference chemicals with data for comparison produced equivocal results for effects on E2 synthesis in the OECD inter-laboratory validation: three chemicals, dimethoate, flutamide, and tricresyl phosphate demonstrated equivocal findings for increased estradiol, and two chemicals, fenarimol and finasteride, demonstrated equivocal findings for decreased estradiol. For these five chemicals, the ‘true’ result is uncertain. Three of the 15 chemicals produced equivocal results for increased E2, leaving 12 chemicals for comparison; of these 12, nine chemicals agreed for increased E2 (bisphenol A, danazol, di(2-ethylhexyl)phthalate, 2,4-dinitrophenol, fenarimol, finasteride, ketoconazole, prometon, spironolactone). Two chemicals (fenarimol and finasteride) were equivocal for decreased E2, leaving 13 chemicals for

comparison; of these 13 chemicals, all agreed for decreased E2 (bisphenol A, danazol, di(2-ethylhexyl)phthalate, dimethoate 2,4-dinitrophenol, flutamide, genistein, ketoconazole, mifepristone, piperonyl butoxide, prometon, spironolactone, and tricresyl phosphate).

Qualitative comparison of the effects of the OECD reference chemicals on T synthesis were similarly concordant. For increased T for the core reference chemicals, two chemicals had equivocal findings (atrazine and butylparaben), and of the remaining eight chemicals, six chemicals agreed (aminoglutehimide, forskolin, letrozole, molinate, nonoxynol-9, prochloraz). For decreased T for the core reference chemicals, there were no equivocal findings, and eight of the ten chemicals agreed (aminoglutehimide, atrazine, benomyl, butylparaben, ethylene dimethanesulfonate, forskolin, molinate, and prochloraz). However, if nonoxynol-9 is excluded based on uncertainty regarding the chemical identity, and letrozole is excluded as the MTC in the HT H295R assay (14 µM) is less than the LOECs reported by the OECD inter-laboratory study (100 µM), then eight of eight core reference chemicals agree for decreased T. Two of the 15 supplemental chemicals produced equivocal results for increased T, leaving 13 chemicals for comparison; all of which agreed for increased T (bisphenol A, danazol, di(2-ethylhexyl)phthalate, dimethoate, 2,4-dinitrophenol, fenarimol, finasteride, flutamide, genistein, ketoconazole, piperonyl butoxide, prometon, and spironolactone). One of the 15 supplemental chemicals produced equivocal results for decreased T, leaving 14 chemicals for comparison; of these 14, ten chemicals agreed for decreased T (bisphenol A, di(2-ethylhexyl)phthalate, dimethoate, flutamide, genistein, ketoconazole, mifepristone, prometon, spironolactone, and tricresyl phosphate).

Equivocal findings and discordances included the following chemicals:

(a) Aminoglutehimide was likely a borderline positive for decreased E2 in the OECD inter-laboratory validation; three of the five labs reported a LOEC at the greatest non-cytotoxic concentration (100 µM) with no concentration-response, and one lab reported a LOEC that was annotated as not significantly different from control (p-value of 0.051). Aminoglutehimide was negative for E2 effects in the HT H295R assay, but it did significantly decrease several hormones (11-deoxycortisol, DOC, progesterone, OH-progesterone, androstenedione, and testosterone) and increase progesterone at 100 µM. These responses produced a weak pathway positive, with a low but significant adjusted maxmMd (1.56), and so would not constitute a false negative for effects on steroid biosynthesis when using all of the available screening data.

(b) Atrazine induced increases in T were not detected by two of five laboratories, and though the HT H295R assay was a positive with curve-fit warning flags, the pathway analysis produced a significant adjusted maxmMd (3.14) as atrazine moderately, but significantly, affected 10 of the 11 hormones in the pathway.

(c) Benomyl was negative in the OECD inter-laboratory validation for effects on T, but produced a borderline positive in the HT H295R assay for increased T (*i.e.*, effects were not concentration-dependent and failed to exceed the threshold of 1.5-fold control). The adjusted maxmMd was positive but small (0.16).

(d) Butylparaben was negative for effects on E2 in the HT H295R assay, but produced equivocal results for increased E2 in the OECD inter-laboratory validation, as three of five labs failed to detect a LOEC. Three of five laboratories in the OECD validation failed to detect a LOEC for butylparaben-induced increases in T, and the HT H295R T results were negative; however, butylparaben was a pathway positive (adjusted maxmMd = 4.64), as it significantly affected two progestagen hormones (progesterone and OH-progesterone) in the pathway.

(e) Danazol decreased T in the HT H295R assay but was negative in the OECD inter-laboratory validation; danazol in this comparison is classed as a false positive, but appeared to affect several hormones across the pathway in the HT H295R in a concentration-consistent manner (adjusted maxmMd = 15.3 – 21.5).

(f) 2,4-Dinitrophenol decreased T in the OECD inter-laboratory validation with LOECs for the two laboratories that ranged five orders of magnitude on a log10 scale (0.0001-100 µM). 2,4-dinitrophenol was negative in the HT H295R assay, but was screened only at the MTC (10 µM); no concentration-response data were available for pathway-based analysis and so a maxmMd value was not computed.

(g) Ethylene dimethanesulfonate (EDS) was negative in the OECD inter-laboratory validation for effects on E2 synthesis, but was a conditional positive in the HT H295R assay for increased E2; though multiple concentrations were positive, the effects were not concentration-responsive and were not significant at the maximum concentration; further these effects did not exceed 1.5-fold of the control. As such, this positive result for EDS in the HT H295R assay was a borderline positive. EDS was also negative in the OECD inter-laboratory validation for effects on T, but produced a conditional or borderline positive in the HT H295R assay for increased T (*i.e.*, effects were not concentration-responsive and failed to exceed the threshold of 1.5-fold control). Supportive of these borderline findings

for E2 and T is the negative result for the pathway-based approach due to a maxmMd that failed to exceed the critical limit (adjusted maxmMd of -0.433).

(h) Finasteride decreased T in the OECD inter-laboratory validation; though it failed to significantly perturb T in the HT H295R assay (only one concentration, 10 µM, was significant), it significantly affected production of OH-pregnenolone, progesterone, OH-progesterone, DOC, 11-deoxycortisol, and androstenedione, yielding a pathway positive (adjusted maxmMd of 12.3).

(i) Genistein increased E2 in the OECD inter-laboratory validation, but failed to increase E2 in the HT H295R assay. Genistein did produce a strong pathway positive, based on significant effects on OH-pregnenolone, progesterone, OH-progesterone, DOC, 11-deoxycortisol, cortisol, androstenedione, and T, with a significant, high adjusted maxmMd (31.8). One concentration, 11.11 µM, appeared to significantly increase estrone and estradiol, but did not meet the minimum criteria for a positive result (two consecutive concentrations with significant results or the highest non-cytotoxic concentration with significant results). Genistein was a strong positive using a pathway approach.

(j) Letrozole was reported to decrease T, but all 5 laboratories in the OECD inter-laboratory validation reported a LOEC at the maximum tested concentration only (100 µM), which exceeded the MTC used in the HT H295R to maintain cell viability (14 µM). Based on differences in the concentration range tested, letrozole was excluded from the confusion matrix for decreased T. Letrozole was maintained in the confusion matrices for the other effect types that would not have been affected by inability to screen up to 100 µM. Letrozole, a pharmacologic CYP19A1 inhibitor, inhibited estrone and E2 production at sub-micromolar concentrations such that these hormones dropped below the LLOQ in addition to moderate effects on several other hormones in the pathway (adjusted maxmMd = 12.4).

(k) Mifepristone increased E2 in the OECD inter-laboratory validation, but failed to increase E2 in the HT H295R assay. Mifepristone produced significant effects on two hormones (progesterone, OH-progesterone), with trends toward decreased DOC, corticosterone, 11-deoxycortisol, cortisol. The responses across the pathway produced a high adjusted maxmMd (33.1).

(l) Nonoxynol-9 was negative for effects on E2 synthesis in the OECD validation, but positive in the HT H295R assay for decreased E2; it is unclear if this was a false positive in the HT H295R assay or not due to uncertainties associated with the identity of the substance

tested in the OECD validation (see Methods for detailed discussion). The magnitude of the effect on E2 synthesis was low. Nonoxynol-9 decreased T in the OECD validation, but was negative in HT H295R; however, the LOEC reported by four of the five labs in the OECD validation (10 µM) exceeded the maximum tested concentration in HT H295R (0.4 µM) (one lab failed to detect a LOEC, and all reported LOECs reflected a single significant concentration). Nonoxynol-9 was just barely positive in the pathway-based approach (adjusted maxmMd of 0.078). Uncertainties regarding the chemical substance, and the disparity in the tested concentration range due to cytotoxicity concerns, supported revision of the confusion matrix to exclude nonoxynol-9.

(m) Piperonyl butoxide was negative in the OECD inter-laboratory validation but produced a conditional positive for increased E2 in the HT H295R assay, with multiple concentrations significantly different from control that failed to exceed 1.5-fold of the control. Piperonyl butoxide minimally decreased T synthesis in the OECD inter-laboratory validation at 10 µM. Piperonyl butoxide failed to affect T synthesis in the HT H295R assay, but did demonstrate minor effects on a number of hormones in the pathway, often without a monotonic concentration-response, yielding a weak pathway positive and adjusted maxmMd of 2.30.

Combined comparison of E2 and T results and maxmMd for OECD reference chemicals

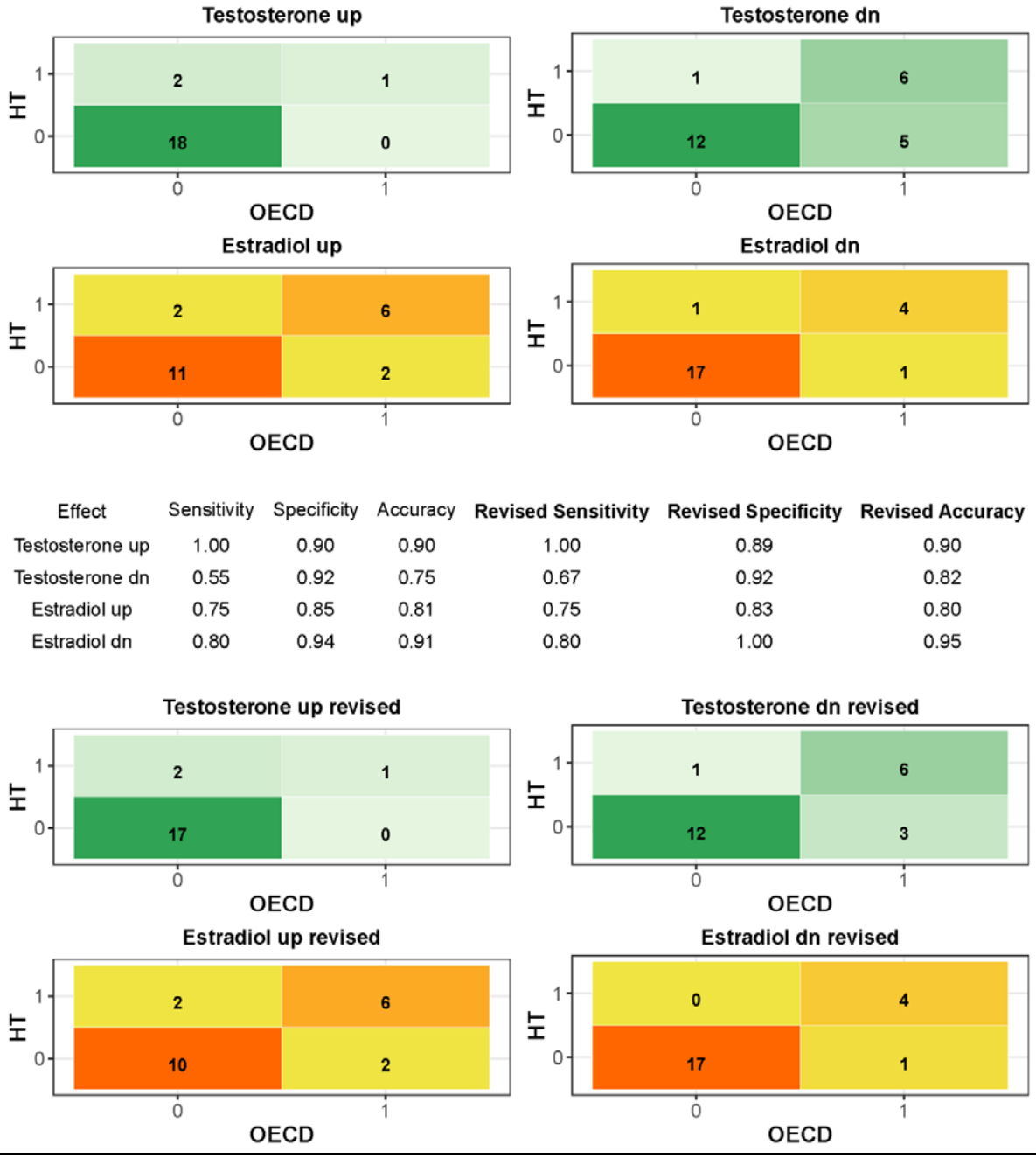
A summary comparison of the OECD inter-laboratory and HT H295R results for E2 and T for each reference chemical is provided in **Figure 3-9** along with a positive or negative designation for the pathway-based maxmMd analysis. In **Figure 3-9**, the reference chemicals are rank-ordered by $\log_{10}$-maxmMd. The maxmMd value appears to separate known strong steroidogenesis disruptors largely comprised of pharmacological modulators of hormone biosynthesis (*e.g.*, mifepristone, prochloraz, ketoconazole, danazol, letrozole) from moderate disruptors (*e.g.*, atrazine, molinate, di(2-ethylhexyl-phthalate) and from non-active chemicals (*e.g.*, EDS). However, effects of these reference chemicals on progestagen and glucorticoid biosynthesis is unknown in some cases. Known activities of these reference chemicals, approximations of the magnitude of perturbation, the quadrants of the steroid biosynthesis pathway perturbed in the HT H295R assay, the number of steroid hormones perturbed in the HT H295R assay, and the maxmMd values are briefly summarized in Supplemental File 11.

Consideration of the maxmMd as a ranking metric

A data-driven approach to understanding the added value of the maxmMd metric involved comparison of the number of steroid hormones significantly affected by a chemical using the ANOVA-based logic with the maxmMd value for that chemical. A boxplot of the maxmMd values, binned by the steroid hormone hit-count, is presented in **Figure** 3-10. The primary purpose of this visualization is to demonstrate that the sum of steroid hormone hits does not necessarily relay the magnitude of the effect of a test chemical on the set of 11 steroid hormones, whereas the maxmMd value allows for quantitative distinction of chemicals that affect similar numbers of hormones but with varying efficacy. The median of the maxmMd values generally increased as the steroid hormone hit-count increased; however, the maxmMd values enabled distinction of chemicals with the same steroid hormone hit-count, in some cases by more than one order of magnitude on a $\log_{10}$ scale. For example, both tricresyl phosphate and letrozole significantly perturbed synthesis of seven steroid hormones in the set based on the ANOVA logic employed, but their adjusted maxmMd values were 0.94 and 12.4, respectively. Mifepristone significantly affected only two steroid hormones, but with great magnitude, such that it had a high maxmMd. BPA was replicated on 3 plates in two different screening blocks, and across these three replicates, perturbed five to seven hormones based on minor effects for a few steroid hormones near the threshold for positive activity; however, the maxmMd values were relatively stable (adjusted maxmMd values for BPA ranged from 4.21 to 5.22). Open symbols in **Figure 3-10** indicate chemicals with maxmMd values that failed to exceed the critical value, *i.e.*, pathway-based negatives; these negatives are distributed across steroid hit count bins of zero to six, indicating that though effects of low magnitude may produce positive results in the ANOVA-based logic, the maxmMd provides a more quantitatively robust indicator of pathway perturbation than the sum of steroid hormone hitcalls. Of the OECD reference chemicals, EDS yielded a negative adjusted maxmMd value, with a corresponding steroid hormone hit-count of six. Benomyl, nonoxynol-9, and tricresyl phosphate produced weak pathway positives with adjusted maxmMd values of 0.16, 0.078, and 0.94 that corresponded to steroid hormone hit-counts of four, four, and seven. Conversely, small trends in the data for multiple steroid hormones that are not significant may result in a positive maxmMd; in the case of dimethoate, the adjusted maxmMd is 0.12, just above zero, indicating a very low pathway response that corresponds to no significant steroid hormone perturbations by the ANOVA-based logic. Thus, the maxmMd value appears to provide added value above steroid hormone hit-count alone for description of the magnitude of steroid biosynthesis pathway effects.

As suggested by the appearance of BPA on three plates across two separate screening blocks, the maxmMd appears to be a relatively stable metric when compared to steroid hormone hit counts. One hundred seven chemicals were screened in more than one screening block (all other chemicals appeared in technical duplicate in one screening block only). Ninety-four of the 107 chemicals (87.9%) replicated a positive (maxmMd > critical value) or negative (maxmMd < critical value) pathway response across blocks. In contrast, the average recall for the 11 steroid hormone hit-calls across replicate blocks using the OECD guideline-based logic was approximately 65% (Supplemental File 12). For this 107 chemical subset with data from replicate blocks, estimation of the residual standard deviation (0.33) suggests that the 95% confidence interval for predicting the maxmMd would be the maxmMd value multiplied or divided by 1.93 (Supplemental File 12). For bisphenol A, with a median maxmMd of 5.98, the 95% confidence interval around this value would be 3.10 to 11.5. Further evaluation of the reproducibility of the maxmMd metric suggests that the standard deviation between maxmMd values derived from replicate blocks is slightly decreased in value for increased maxmMd values. Therefore, the calculated 95% confidence interval likely represents a conservative estimate of the variability of the maxmMd metric. The reproducibility of the maxmMd values across blocks is described in further detail in Supplemental File 12.

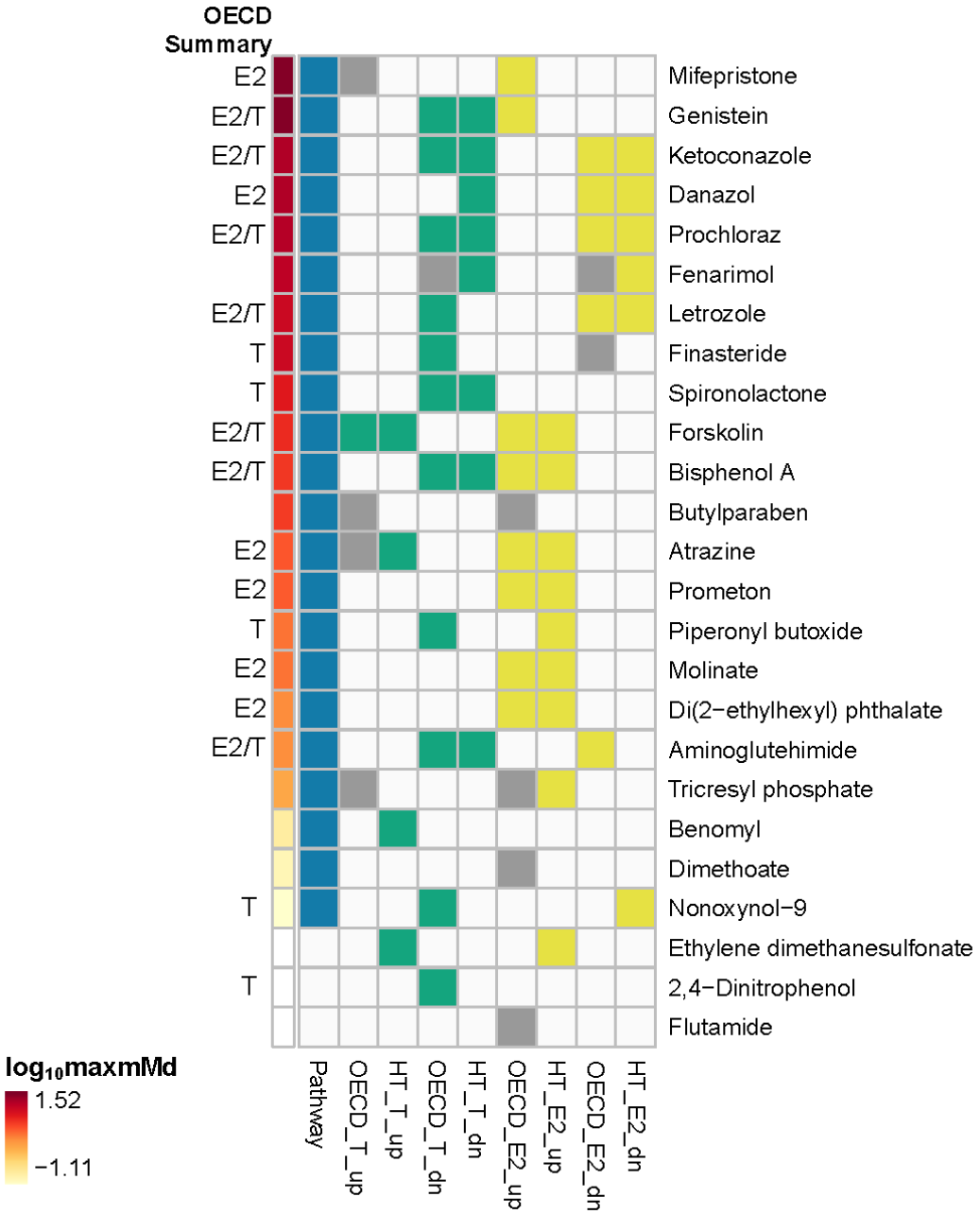
Figure 3-8 Confusion Matrices for Effects on T and E2



Effect	Sensitivity	Specificity	Accuracy	Revised Sensitivity	Revised Specificity	Revised Accuracy
Testosterone up	1.00	0.90	0.90	1.00	0.89	0.90
Testosterone dn	0.55	0.92	0.75	0.67	0.92	0.82
Estradiol up	0.75	0.85	0.81	0.75	0.83	0.80
Estradiol dn	0.80	0.94	0.91	0.80	1.00	0.95

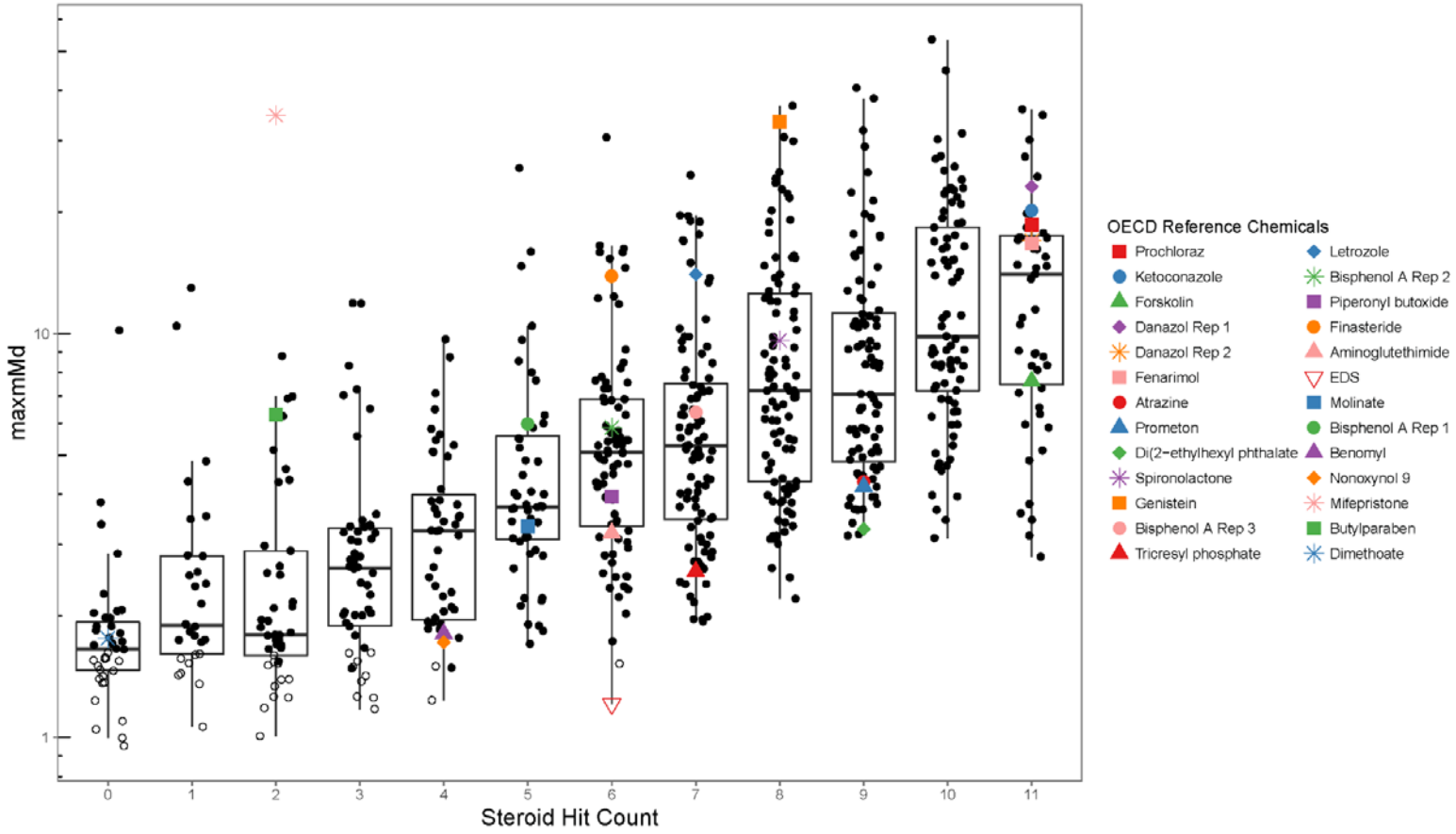
The OECD inter-laboratory validation study results (Hecker et al., 2011) were interpreted as true outcomes, and the HT H295R results analyzed by ANOVA with a post-hoc Dunnett's test were interpreted as predicted outcomes. Four effect types were considered: increased (up) and decreased (dn) testosterone (T) and estradiol (E2). The number of chemicals included for each effect type varied because chemicals with equivocal results for the effect type (4 for T up, 1 for T down, 4 for E2 up, 2 for E2 down) were removed. Revised confusion matrices present the comparison without nonoxynol-9 and omitting letrozole from testosterone dn.

Figure 3-9 Geometric Tiling to Compare the OECD Validation and HT H295R Results



For each chemical in the core and supplemental OECD chemical reference sets, a binary comparison of the OECD inter-laboratory validation result (OECD_) and the HT H295R results (HT_) is presented. Positive E2 responses are blocked as yellow, positive T responses are blocked as green, equivocal responses in the OECD inter-laboratory validation are blocked as gray, and negatives are blocked as white. Blue blocks denote positive pathway responses (defined as the maxmMd exceeding the critical limit for a chemical), and the annotation bar ranks all of the chemicals in the set by their $\log_{10}$ maxmMd from high (red) to low (yellow), white blocks indicating negative pathway results. "OECD Summary" is a text annotation to indicate whether an effect (up or dn) was observed for E2 or T in the OECD inter-laboratory validation.

Figure 3-10 Boxplot of Adjusted maxmMd Values Versus Sum of Steroid Hormone Positive Responses



The maxmMd values for all 654 chemicals were binned by steroid hit count (ranging from 0 to 11 steroid hormones, as analyzed by the ANOVA logic employed herein), with the y-axis is $\log_{10}$-scaled. OECD reference chemicals are annotated within the plot. Closed symbols for all chemicals, including OECD reference chemicals, indicate positive maxmMd values that exceeded the critical value; open symbols for all chemicals, including OECD reference chemicals, indicate negative maxmMd values.

3.4 Discussion

The current work demonstrates the utility of the HT-H295R screening assay as an alternative for the OECD-validated, low throughput H295R assay (OECD TG 456). The ANOVA analysis and logic used herein for the HT-H295R dataset to determine effects on the steroid biosynthesis pathway enabled a direct comparison of the OECD inter-laboratory validation data and the HT-H295R data. This detailed, performance-based comparison highlights good concordance of results, with accuracies that range 0.75 – 0.91 for effects on E2 and T. Understanding that E2 and T provide limited perspective on the impact of chemicals on the steroidogenesis pathway present in H295R cells, this work also presents a novel evaluation of hormone data from more of the steroid biosynthesis pathway. To integrate 11 steroid hormone analytes for pathway-level analysis using the HT-H295R assay data, a mean Mahalanobis distance (mMd) was computed for each chemical concentration screened. The mMd provided a set of unitless values from which the maximum mean Mahalanobis distance (maxmMd) could be calculated across the concentration range screened. We suggest that this maxmMd may be useful for prioritizing chemicals by the relative magnitude of their overall impact on the steroid biosynthesis pathway. Thus, this work, through demonstration of the HT-H295R as an alternative and a novel data analysis approach, advances efforts to rapidly identify and prioritize large numbers of chemicals as potential steroidogenesis disruptors for further evaluation or confirmatory screening.

Evaluation of the concordance of the OECD reference chemical effects on E2 and T synthesis in the OECD inter-laboratory validation exercise and the HT-H295R screening campaign demonstrated similarity in the findings, despite some differences in experimental assay design. In addition, it also underscored some of the thematic challenges of comparing alternative screening approaches to traditional methods. The OECD reference chemical set was heavily weighted with “true” negatives for E2 or T, yielding relatively high specificity values (0.85-0.94). However, only one of the 25 chemicals with data from the OECD inter-laboratory validation that were screened in the HT-H295R assay (EDS and benomyl) were negative in the OECD inter-laboratory validation for effects on both E2 and T synthesis (Supplemental File 11). The OECD reference chemical set was also limited in the number of “true” positives; however, the sensitivity values (0.55-1.0) demonstrated that the HT-H295R assay was capable of detecting

these chemical effects on E2 and T alone. The sensitivity without adjustment for decreased T was 0.55, but increased to 0.67 if nonoxynol-9 and letrozole were omitted (for reasons of chemical uncertainty and a LOEC > the MTC, respectively).

Using a pathway approach based on the maxmMd, rather than solely measures of E2 and T, appeared to increase screening sensitivity and identified chemicals as pathway positives that were potential HT-H295R false negatives for effects on E2 (aminoglutehimide failed to decrease E2; mifepristone and genistein failed to increase E2), and T (2,4-dintrophenol, finasteride, and piperonyl butoxide failed to decrease T). One hypothesis for the false negative findings for mifepristone and genistein and increased E2 is that the HT-H295R system may be slightly less sensitive to E2 increases due to pre-stimulation with forskolin. However, a critical strength of collecting data for multiple steroid hormones in the pathway and combining these data into a single metric, the maxmMd, is that weak effects on multiple hormones, or strong effects on one or two hormones, can contribute to a pathway-based positive. Indeed, all of the aforementioned potential false negatives were pathway positives using this approach. In addition to the need for a higher number of curated reference chemicals with data from multiple studies on which to base evaluations, it would be helpful to have reference chemicals to better evaluate the steroidogenesis pathway as a whole, including known negatives for the entire pathway, and chemicals with effects on glucocorticoid and progestagen synthesis. The small number of “true” negatives for both E2 and T in the OECD reference chemical set, and a lack of information regarding “true” pathway negatives, limits determination of the negative predictive value of the maxmMd approach. Another challenge in comparing these datasets includes the variability in the reference data set; data insufficient for comparison due to laboratory disagreements, and reported potency and efficacy values that were highly variable, are difficult to evaluate for validation purposes. However, ranking screened chemicals by the magnitude of perturbation induced across the steroid biosynthesis pathway appears to represent an effective and efficient means of understanding the priority of particular chemicals within a list, above and beyond tabulation of the number of steroid hormones perturbed (Figure 3-10). As suggested in Figure 3-10, though the maxmMd generally increased with increasing number of steroid hormones affected, the maxmMd metric appeared to provide the ability to distinguish chemicals with the same steroid hormone hit count. Further, as detailed in Supplemental File 11, the maxmMd appeared to distinguish strong modulators of steroidogenesis (e.g., mifepristone, genistein, prochloraz,

ketoconazole, danazol, letrozole, with adjusted maxmMds ranging from 33.1 to 12.4) from moderate modulators of steroidogenesis (e.g., BPA, butylparaben, atrazine, prometon, with adjusted maxmMd values ranging from 5.22 to 3.10) and minor or borderline modulators (e.g., piperonyl butoxide, molinate, benomyl, and nonoxynol with maxmMd values ranging from 2.30 to 0.078) or negative chemicals showing no effect on steroidogenesis (e.g., EDS, flutamide, 2,4-dinitrophenol, with adjusted maxmMd values of ≤0 or NA). As with nearly any alternative approach, additional reference chemicals with full steroid biosynthesis pathway information would enable additional consideration of the quantitative and qualitative value of using the maxmMd approach.

In addition to the ToxCast HT-H295R screening implementation, other research efforts have measured multiple steroid hormones in the pathway expressed in H295R cells (Hansen et al., 2017; Tonoli et al., 2015; Maglich et al., 2014; Abdel-Khalik et al., 2013; Nielsen et al., 2012; Rijk et al., 2012; Zhang et al., 2011). However, to date, the ToxCast screening implementation remains the largest publicly reported screening effort in terms of number of chemicals and concentrations evaluated for effects on steroid hormones. The number of recent reports that measure multiple hormones in the steroidogenesis pathway support the concept that the synthesis of steroids other than just E2 and T contribute important insight into chemically-induced steroidogenesis disruption. Existing computational models for chemical modulation of interdependent hormone profiles in H295R cells have employed a systems biology approach, incorporating biological and kinetic information to quantitatively estimate the anticipated levels of multiple steroid hormones in the pathway following chemical exposure (Saito et al., 2016; Breen et al., 2010). The necessary time course information used to inform such a pathway-based model has not been generated on the HT-H295R assay. Thus, the current work uses an empirical approach to statistically integrate screening data for 11 steroid hormones and compute a summary value (i.e., mMd) for each chemical concentration screened. Using mMd values statistically accounts for the correlation of the residuals of the steroid hormone measures rather than using information about the enzyme reaction kinetics to describe their interrelatedness. The concentration-response behavior of the mMd values was also condensed to a single value (i.e., maxmMd), which may be useful in prioritizing chemicals to more accurately reflect their effects on the broader steroidogenesis pathway. If a systems biology model is required to adequately interpret the data for regulatory decisions, future work would be needed to develop time-course

information for control chemicals in the HT-H295R assay to inform model parameters. Further, the potential interaction of biological mechanisms beyond cholesterol transport and enzymatic steroid synthesis reactions, e.g., the contributions of steroid hormone nuclear receptors expressed in H295R cells such as the glucocorticoid and androgen receptors (Robitaille et al., 2015; Asser et al., 2014; Yanes and Romero, 2009; Hecker et al., 2006) is the subject of ongoing research, and suggests additional mechanisms that could be included in a model.

As demonstrated in the current study, the maxmMd may be a useful for prioritizing chemicals for additional testing, but areas of uncertainty in applying these pathway data should be noted. For example, more work is needed to understand how to translate *in vitro* steroidogenesis findings to prediction of *in vivo* effects. In previous validation efforts for guideline-based H295R assays, *in vitro* results have not always predicted the effect or correct direction of effect for serum steroid hormone findings (Paul Friedman et al., 2016; Lebaron et al., 2014; Hecker et al., 2011). To more accurately extrapolate to *in vivo* effects, a broader range of endpoints need to be considered and integrated including the absorption, distribution, metabolism, and excretion of a chemical, aromatase activity and other enzymes necessary for steroidogenesis (e.g., conversion of T to 5α-dihydrotestosterone by steroid-5α-reductase), indicators of cholesterol transport, and markers of mitochondrial toxicity.

Another area of uncertainty in understanding how HT-H295R assay data might be translated involves interpretation of changes in progestagen and glucocorticoid hormones. Changes detected in the synthesis of progestagen and glucocorticoid hormones may inform hypothetical mechanisms of action, particularly for potential inhibition of enzymes that act early in the steroidogenesis pathway in H295R cells, e.g. steroidogenic acute regulatory protein (StAR), CYP11A1, and 3β-hydroxysteroid dehydrogenases. Further, given that H295R cells present a dynamic system, one might hypothesize that modulation of progestagens in particular would eventually propagate to changes in downstream glucocorticoid, estrogen, and androgen production *in vitro* given enough exposure and time (Saito et al., 2016). *In vitro* studies have suggested that H295R cells are useful for identifying chemicals that may perturb only progestagens and/or glucocorticoids and modulate hypothalamic-pituitary-adrenal axis function *in vivo*, leading to pathologies associated with hyper- or hypofunction of the adrenal (Strajhar et al., 2017; Oskarsson et al., 2016). However, the database of animal toxicology information to connect these *in vitro* findings and *in vivo* measures is lacking. Instead, if it was important to

focus a particular prioritization task for chemicals that may affect estrogen and androgen synthesis specifically, chemicals with effects on estrogens and/or androgens could be ranked using the maxmMd approach, i.e. separating chemicals that affected progestagens and/or glucocorticoids only (i.e., an absence of effects on any other hormones in the pathway) into a list for future consideration. From the 654 chemicals with data for the maxmMd analysis, 596 chemicals had positive maxmMd responses, and of these, 10 chemicals affected only estrogen and/or androgen synthesis, and 67 chemicals affected only progestagen and/or glucocorticoid synthesis. One conclusion from the Venn diagram presented in Figure 4 is that though it is possible to identify chemicals that only perturb estrogen and/or androgen synthesis, most of the chemicals in the screened set affected other steroid hormones as well, and using these data to evaluate the magnitude of overall pathway effect appears useful. Interestingly, examples of the chemicals that affected only synthesis of progestagens and/or glucocorticoids include butylparaben, in line with an independent report of its activity in the H295R model (Taxvig et al., 2008), and prednisone, which has known clinical interactions with the mineralocorticoid receptor as a glucocorticoid prodrug (Ferraldeschi et al., 2013). Thus, excluding chemicals that only affect progestagen and/or glucocorticoid synthesis from prioritization tasks may exclude chemicals with activities of potential interest, and using the maxmMd approach for the whole pathway would be more inclusive. The ratio of observed positives in this screened chemical set by steroid hormone class might shift if a naïve screening approach was taken without pre-selecting positive chemicals based on single concentration screening. However, this high rate of pathway positives in this pre-selected set demonstrates the original success of the HT-H295R ToxCast screening workflow in terms of identifying chemicals that may disrupt steroidogenesis by performing single concentration screening followed by concentration-response screening.

The work described herein demonstrates the performance of the HT-H295R assay as an alternative to the OECD TG 456 H295R assay, and proposes use of a novel statistical approach to integrate the information from 11 steroid hormones in the pathway to yield a relative rank of steroidogenesis perturbation. The approach based on Mahalanobis distances accounts for the correlation of the residuals of the hormone measures. A clear advantage of the mean Mahalanobis distance approach is that the concentration at which effects across the pathway begin to occur can be identified. Further, the pathway analysis approach appears to increase the sensitivity of detecting chemicals that are known to perturb the pathway. As the database of

reference chemicals for perturbation of *in vitro* steroidogenesis grows, further characterization of the strengths and weaknesses of this approach will develop. Potential use of the maxmMd in prioritization tasks represents a data-driven option for evaluating lists of chemicals for putative effects on steroidogenesis.

3.5 Limitations of the Current Assay and Future Refinements

3.5.1 Metabolic Capacity

Limitations of the HT H295R assay includes the lack of or limited metabolic capacity of the system. The ability to incorporate metabolic competence into high-throughput screening assays is the subject of ongoing research at the US EPA Office of Research and Development. In addition, the US EPA has partnered with the National Toxicology Program (NTP) and the National Center for Advancing Translational Science (NCATS) to engage the broader scientific and commercial communities in tackling the issue of incorporating metabolism into cell-based and cell-free assay systems through the release of a challenge competition (EPA news release). Further, chemical structure-based models are under development to identify chemicals predicted to undergo transformation to more bioactive metabolites. It should be noted that the low throughput H295R assay currently in the EDSP Tier 1 battery also lacks metabolic competency. Ability to make predictions for chemicals beyond the domain of applicability for the current assay set.

3.5.2 Chemical Library Restrictions

Currently the HT H295R assay and associated maxmMd analysis are limited by the chemical libraries made available to the ToxCast program. The current libraries are restricted to DMSO-soluble chemicals. Future plans also include expanding chemical testing to a water-soluble library.

3.6 Future Use of the HT H295R Assay and maxmMd Analysis

3.6.1 Prioritization and Risk Assessment

Although the HT H295R assay results and maxmMd analysis have demonstrated ability to prioritize environmental compounds for potential to disrupt steroidogenesis on a hazard basis, they should be integrated with exposure estimates for decision making in a risk assessment framework (Paul Friedman et al., 2016; Teeguarden et al., 2016; U.S. EPA, 2015f). The integrated bioactivity and exposure (IBER) methodology was presented to a FIFRA SAP in

December 2014 (U.S. EPA, 2014a); in this approach, the bioactivity based on the HT assays and computational models are put in a dose context and compared with high-throughput exposure estimates (Wambaugh et al., 2014) to prioritize chemicals for further evaluation.

3.6.2 Alternative to Other EDSP Tier 1 Assays

To ultimately interpret disruption of *in vitro* steroidogenesis that includes potential adverse *in vivo* outcomes, a broader range of endpoints need to be considered and integrated in the analysis. For example, aromatase activity and other enzymes necessary for steroidogenesis (e.g., conversion of T to 5α-dihydrotestosterone by steroid-5α-reductase), indicators of cholesterol transport, and markers of mitochondrial toxicity need to be incorporated in a systematic and quantitative manner. Further, interpretation of changes in progestagen and glucocorticoid hormones need to be better understood. *In vitro* studies have suggested that H295R cells are useful for identifying chemicals that may perturb only progestagens and/or glucocorticoids and modulate hypothalamic-pituitary-adrenal axis function *in vivo,* leading to pathologies associated with hyper- or hypofunction of the adrenal (Strajhar et al., 2017; Oskarsson et al., 2016). However, the database of animal toxicology information to connect these *in vitro* findings and *in vivo* measures is lacking. Additional study will be required to put the full range of *in vitro* perturbations in an *in vivo* hazard context.

Future interpretation of the HT H295R assay activity with respect to potential *in vivo* effects will also require application of high-throughput *in vitro* toxicokinetic assays and *in vitro*-to-*in vivo* extrapolation (IVIVE) approaches to provide an *in vivo* dose context (Wambaugh et al., 2015; Wetmore et al., 2015; Chang et al., 2014; Wambaugh et al., 2014; Rotroff et al., 2010b). The high-throughput toxicokinetic assays and IVIVE approaches allow the conversion steady state blood concentrations equivalent to *in vitro* potency estimates (uM) from the HT H295R assay and maxmMd analysis into *in vivo* administered doses (mg/kg/d). These efforts can be used to validate additional high-throughput *in vitro* assays and computational models to predict more complex *in vivo* effects.

3.7 Performance-Based Approach to Establishing Confidence: Considerations & Conclusions for the HT H295R Assay and Mahalanobis Distance Analysis

3.7.1 The Performance-Based Approach

This chapter has demonstrated the performance of the HT H295R assay compared with the OECD TG 456 H295R assay, and proposes use of a novel statistical approach to integrate the

information from 11 steroid hormones to quantify the overall impact of a chemical on the steroidogenesis pathway. The narrative in Section 1.5 provides an overall framework for establishing confidence for new, alternative approaches. Regarding the HT H295R assay and integrated statistical analysis, each of the considerations have been met in the regulatory context of screening and prioritization of chemicals for purposes of evaluating potential for disruption of steroidogenesis:

- *Mechanistically and/or Biologically Relevant Assays:* The steroidogenesis assay employing the human adrenocarcinoma cell line (H295R) is an *in vitro* method for detecting chemical disruption of the catalytic events of steroidogenesis, and has been used predominantly to predict chemical perturbation of 17β-estradiol (E2) and testosterone (T) synthesis (OECD, 2011). The H295R cell line demonstrates the biological characteristics of zonally undifferentiated human fetal adrenal cells, but produces steroid hormones found in adult adrenal cortex, ovaries, and testes (Gracia et al., 2006; Gazdar et al., 1990). H295R cells have been used to evaluate effects of xenobiotics on hormone production, as well as steroidogenic enzyme activity and expression (Maglich et al., 2014; Hilscherova et al., 2004; Sanderson et al., 2000). Many different mechanisms of disruption may be captured in measurement of hormones from the H295R assay. The HT H295R variation of the assay maintains the mechanistic and biological relevance of the original H295R assay.
- *Reliability considering accepted best practices within the given field:* Reliability of the data and the model presented here must be interpreted within the context of being fit for the purpose of screening for prioritization of large numbers of chemicals. Ten of the twelve OECD core reference chemicals used for validation and fifteen of the sixteen supplemental reference chemicals in the ToxCast chemical library were evaluated in the HT H295R assay, and produced qualitatively similar results to the validated low-throughput assay with accuracies of 0.90/0.75 and 0.81/0.91 for increased/decreased testosterone production and increased/decreased estradiol production, respectively. The maxmMd has been proposed as a potentially useful summary value to represent the magnitude of changes observed in the production of up to 11 steroid hormones in the HT-H295R assay. The reliability of the maxmMd analysis was assessed based on chemicals assayed in more than one block. Block was defined as the experimental screening date,

with eight screening blocks used to generate the currently available dataset. On a qualitative basis, 94 of the 107 chemicals (87.9%) screened in more than one block replicated a pathway positive (i.e., maxmMd > critical value) or negative (i.e., maxmMd < critical value) pathway response across blocks. In contrast, the average recall for the 11 steroid hormone hit-calls across replicate blocks (using the OECD guideline-based analysis procedure) was approximately 65%.

- *Transparency:* Supplemental files and figures listed in Section 3.5 contain the detailed information about the assays and the data used, results, supporting information, and associated analyses. The operating procedure has been defined in detail in a previously published work (Karmaus et al., 2016). The R scripts (R Core Team, 2015) used to analyze the data, calculate the maxmMd values, and generate figures are available statically as Supplemental File 13. Updated versions (including the version for this Whitepaper) are available at: ftp://newftp.epa.gov/COMPTOX/NCCT_Publication_Data/Haggard/2017_Prediction_of_H295R_steroidogenesis_Pathway_Perturbation/.
- *Uncertainties and Limitations:* Uncertainty in the maxmMd was estimated based on confidence intervals defined for the 107 block-replicated chemicals. The range of maxmMd values for this 107 chemical subset was 0.996 to 34.7, spanning from weak to strong modulators of steroid biosynthesis. Estimation of the residual standard deviation (0.33) for these 107 chemicals suggests that the 95% confidence interval for predicting the maxmMd would be the maxmMd value multiplied or divided by 1.93. As an example, for bisphenol A, with a median maxmMd of 5.98, the 95% confidence interval around this value would be 3.10 to 11.5. Further evaluation of the quantitative reproducibility of the maxmMd metric suggests that the standard deviation between maxmMd values derived from replicate blocks is slightly decreased in value for increased maxmMd values. I.e., as the maxmMd value increases, there is likely greater confidence in that value. Therefore, the calculated 95% confidence interval likely represents a conservative estimate of the variability of the maxmMd metric. The limitations of the HT H295R assay and the maxmMd as a summary metric for quantifying the overall perturbations in the steroidogenesis pathway have been outlined in Sections 3.4 and 3.5.

- *Reference Chemicals:* Ten of the twelve OECD core reference chemicals used for validation and fifteen of the sixteen supplemental reference chemicals were evaluated in the HT H295R assay and using the maxmMd approach. These represent the most comprehensive list of reference chemicals currently available.
- *Peer Review:* The 2017 FIFRA SAP will review this work. Further information on the function of the HT-H295R assay has been previously published (Karmaus et al., 2016), while newer has been recently submitted for publication (Haggard et al.).

3.7.2 Summary and Conclusions

Considering the entire steroidogenic pathway as a means of identifying potential endocrine disrupters is not a new idea. Previously, computational models of chemical modulation of the interdependent hormone profiles in H295R cells have employed a systems biology approach, incorporating biological and kinetic information to quantitatively estimate the anticipated levels of multiple steroid hormones in the pathway following chemical exposure (Saito et al., 2016; Breen et al., 2010). These existing models from the cited publications help in determining the mechanism of action by, for instance, estimating which steroidogenic enzymes are affected.

To integrate 11 steroid hormones for pathway-level analysis using the HT H295R assay data, a mMd was computed for each chemical concentration screened. The concentration response behavior of the mMd values was condensed to a single value, the maxmMd. The pathway approach confers the following advantages:

- The maxmMd can be useful for prioritizing chemicals, more accurately reflecting a chemical's effects on the broader steroidogenesis pathway. Ranking screened chemicals by the magnitude of perturbation induced across the steroid biosynthesis pathway appears to represent an effective and efficient means of understanding the priority of particular chemicals within a list, above and beyond tabulation of the number of steroid hormones perturbed.
- The concentration at which effects across the pathway begin to occur can be identified.
- Measuring 11 hormones along the steroidogenesis pathway allows more biological information than measuring only E2 and T.
- Using a pathway approach, rather than measures of E2 and T only, appeared to increase screening sensitivity and identified chemicals as pathway positives that were potential HT

false negatives for effects on E2 (aminoglutehimide failed to decrease E2; mifepristone and genistein failed to increase E2), and T (2,4-dinitrophenol, finasteride, and piperonyl butoxide failed to decrease T).

- As discussed in Sections 3.2.2 and 3.2.3, cumulative pathway response can be more sensitive than measuring only E2 and T. Weak effects on multiple hormones, or strong effects on one or two hormones, can contribute to a pathway-based positive. Indeed, all of the aforementioned potential false negatives were pathway positives using this approach.

In addition to the ToxCast HT H295R screening implementation, other research efforts have measured multiple steroid hormones in the pathway expressed in H295R cells (Hansen et al., 2017; Tonoli et al., 2015; Maglich et al., 2014; Abdel-Khalik et al., 2013; Nielsen et al., 2012; Rijk et al., 2012; Zhang et al., 2011), but to date the ToxCast screening implementation remains the largest publicly reported screening effort in terms of number of chemicals and concentrations evaluated for effects on 11 steroid hormones. The previous cited literature that measure multiple hormones in the steroidogenesis pathway support the concept that the synthesis of steroids other than E2 and T may contribute important insight into chemically-induced steroidogenesis disruption.

Based on characteristics described here, EPA concludes that the HT-H295R assay performs equally, if not better, than the low-throughput OECD-validated H295R assay and therefore, the HT H295R assay can be used as an alternative/replacement for the low-throughput assay with respect to measurement of *in vitro* E2 and T biosynthesis.

Further, the HT-H295R assay provides additional steroid hormone biosynthesis data (11 steroid hormones used in this analysis) that when integrated statistically may yield an increased understanding of chemical bioactivity in the H295R assay over the measurement of E2 and T alone.

EPA is soliciting comment on the utility of these additional data for steroid hormones beyond E2 and T, as well as the use of a Mahalanobis distance-based statistical approach (and the maxmMd summary value) to support screening and prioritization.

3.8 Supplemental Files

Supplemental File 1: Normalized MTT cell viability data (multi-concentration response level 3 data from invitrodb_v3) for all 654 chemicals.

Supplemental File 2: Plots of MTT cell viability data (with multi-concentration response level 6 data from invitrodb_v3) for all 654 chemicals.

Supplemental File 3: Master data table of steroid hormone data used in the analyses.

Supplemental File 4: P-value results of ANOVA for 654 chemicals. The p-values for each chemical concentration comparison using a post-hoc Dunnett's comparison are reported.

Supplemental File 5: Binary logic describing the positive or negative result, based on the OECD TG 456 criteria for positives, for each chemical-hormone pair.

Supplemental File 6: Binary string logic used to implement the OECD TG 456-based analysis for positive responses.

Supplemental File 7: Steroid hormone response plots and mean Mahalanobis distance plots for 654 chemicals with multi-concentration response data. Chemical concentrations that were significantly different from control are annotated in red in the steroid hormone response plots.

Supplemental File 8: Radar plots by chemical for the 11 steroid hormone analytes.

Supplemental File 9: Maximum mean Mahalanobis distance (maxmMd) by chemical. The maxmMd, the critical value, and adjusted-maxmMd are provided.

Supplemental File 10: Comparison of the OECD inter-laboratory validation results and HT H295R for E2 and T synthesis. A binary comparison of these data (positive or negative) and additional notes are provided.

Supplemental File 11: OECD Reference Chemical Performance in HT H295R versus OECD inter-laboratory results and literature-reported results. The maxmMd, steroid class and number of steroid hormones affected, and a summary of the OECD inter-laboratory activity and literature-reported activity are provided.

Supplemental File 12: Evaluation of the reproducibility of the maxmMd metric across replicate experimental blocks for 107 chemicals.

Supplemental File 13 R scripts (R Core Team, 2015) used to create the analyses in this chapter as a PDF

4. Thyroid Conceptual Framework

4.1 Introduction

This chapter presents the status of the developing EDSP strategy for a thyroid conceptual framework to identify potential thyroid disrupting chemicals (TDCs). The chapter outlines known thyroid-related pathways, reviews thyroid-related molecular initiating events (MIEs) in an adverse outcome pathway context, and presents the status of a developing set of HT assays for a subset of these thyroid-related MIEs.

The thyroid hormones (THs) thyroxine (T4) and triiodothyronine (T3) are evolutionarily conserved molecules found in all vertebrates and some invertebrates (Laudet, 2011; Crockford,

2009; Heyland and Moroz, 2005). THs are involved in a number of critical physiological processes of homeostasis and development. In most vertebrate species, THs control or regulate cellular and tissue signaling pathways involved in: growth, differentiation, development, maintenance and function of the nervous system, lungs, immune system and other organs, and homeostatic processes involved in energy metabolism. In addition, there are species-specific processes regulated by THs, including thermoregulation in endothermic mammals, and metamorphosis in amphibians and some fish species. A complex set of integrated physiological processes are responsible for the normal regulation of circulating and tissue concentrations of THs including: thyroidal iodide uptake; hypothalamic/pituitary control of TH synthesis; TH storage in and release from the thyroid; circulatory transport via serum TH binding proteins; cellular uptake via TH specific and non-specific membrane transporters; cellular deiodination of T4 to T3; binding of T3 to thyroid hormone receptors (TRs); subsequent activation of TR-dependent transcriptional processes; and, catabolic degradation by hepatic and nephritic enzymes followed by elimination into the bile or urine (Murk et al., 2013; Capen and Martin, 1989).

Due to the complex nature of thyroid hormone regulation and action, chemicals can act on heterogeneous molecular targets and signaling pathways (Murk et al., 2013; Crofton, 2008; Köhrle, 2008; Zoeller and Tan, 2007; Brucker-Davis, 1998; Hurley, 1998; Capen and Martin, 1989). The EPA has previously demonstrated (U.S. EPA, 2015f) that estrogenic activity from nuclear hormone receptor (estrogen receptor (ER)) activity and resultant cellular signaling pathways correctly predicts over 85% of chemicals known to produce positive findings in estrogen related *in vivo* assays (*i.e.*, uterotrophic assay). However, thyroid hormone receptor activity fails to predict the vast majority of thyroid hormone related findings in *in vivo* studies. This appears due to the high ligand specificity of the TH receptors (Freitas et al., 2014) and the ability of chemicals to act with differential sensitivities on the multiplicity of non-TH receptor elements within the thyroid pathways (Hallinger et al., 2017; Paul Friedman et al., 2016; Paul et al., 2014; Brucker-Davis, 1998; Capen, 1997). Consequently, a comprehensive pathway-based approach, that incorporates screening for potential interaction with multiple MIEs, is needed to effectively screen for TDCs (OECD, 2014; Murk et al., 2013).

4.2 Molecular Initiating Events, Key Events and Related Adverse Outcome Pathways

Molecular and cellular processes associated with thyroid homeostasis are known to be altered by xenobiotics and include: hypothalamic and pituitary feedback control; iodine transport and syntheses of thyroid hormones in the thyroid; serum transport of THs; cellular uptake and

metabolism of THs; activation of thyroid hormone receptors; and catabolism and excretion of THs. The molecular targets for these processes, as well as downstream consequences, can be represented as adverse outcome pathways (AOPs) (Ankley et al., 2010). The initial point of chemical-biological interaction within the organism, or molecular initiating event (MIE), can elicit a cascade of key events (KEs) leading to an adverse outcome (AO) of regulatory concern. These linkages can be causative or correlative, and if quantitative information is available, can be used to describe key event relationships (KERs). Thus, AOPs can inform the basis for a decision framework that allows for the use of data collected in models that include MIEs and KEs that may not also include an adverse outcome in order to inform approaches for prioritization and evaluation to identify potential TDCs. The OECD launched a program for crowd-sourced development of AOPs as the central element of a toxicological knowledge framework being built to support chemical risk assessment based on mechanistic reasoning (OECD, 2017). To support this effort, a web-based platform called AOP wiki (OECD, 2016) organizes the curated information into a system that enables standardization of the descriptions as well as connections between AOPs based on shared KEs or AOs.

As a first step, the EDSP has identified 15 potential MIEs for thyroid-based AOPs and established their significance in the thyroid pathways (**Table 4-1**). **Table 4-1** is divided into sections representing different processes within the thyroid network (*i.e.* biosynthesis; transporters; peripheral tissue; receptors) and lists the MIEs associated with each process. The references support the thyroid pathway relevance for each MIE and provide evidence: *1*) that the MIE regulates biologically important processes; *2*) that the MIE, when disturbed, can lead to adverse outcome(s); and *3*) of environmental chemicals interacting with the MIE.

Table 4-1 Potential MIEs for Thyroid-Based AOPs

The column titled "Thyroid Pathway Relevance" establishes that the molecular component is involved with important biological processes in the thyroid pathways. The column titled "References" documents the biological significance of each MIE in the thyroid pathways and/or provides evidence of chemical interactions with the MIE. Adapted from (OECD, 2014; Murk et al., 2013)).

MIE	Thyroid Pathway Relevance	References
TH Biosynthesis in Thyroid		
Sodium-Iodide Symporter (NIS)	Transmembrane glycoprotein which regulates iodide uptake in thyroid follicular cells and is the first step in TH synthesis.	(Hallinger et al., 2017; Mclanahan et al., 2014; Di Bernardo et al., 2011; Cianchetta et al., 2010; Waltz et al., 2010; Lecat-Guillet et al., 2008; Rhoden et al., 2008; Wolff, 1998)
Thyroperoxidase (TPO)	Enzyme secreted into the thyroid colloid which	(Paul-Friedman et al., 2017; Paul et al., 2014; Vickers et al., 2012; Verhaeghe et al., 2008;

MIE	Thyroid Pathway Relevance	References
	oxidizes iodide ions for addition onto thyroglobulin catalyzing the formation of THs.	Schmutzler et al., 2007; Freyberger and Ahr, 2006; Zoeller and Crofton, 2005; Chang and Doerge, 2000; Capen, 1998; Doerge et al., 1998; Divi et al., 1997; Divi and Doerge, 1994; Doerge and Decker, 1994; Doerge and Takazawa, 1990) See also Thyroperoxidase in Table 4-3
Pendrin	Anion exchange protein mediating the efflux of iodide across the apical membrane of the thyrocyte.	(Bizhanova and Kopp, 2011; Zoeller, 2007; Dossena et al., 2006; Taylor et al., 2002)
Dual Oxidase (DUOX)	Enzyme exposed to colloid which co-localizes with TPO and generates peroxide necessary for TH synthesis.	(Carvalho and Dupuy, 2013; Massart et al., 2011; Fortunato et al., 2010; Vigone et al., 2005; Moreno et al., 2002)
Iodotyrosine Deiodinase (IYD)	Deiodinase enzyme in the apical plasma membrane of the colloid which catalyzes deiodination of iodinated tyrosines to recycle iodide.	(Renko et al., 2016; Shimizu, 2014; Shimizu et al., 2013; Buss et al., 2012; Mctamney and Rokita, 2009; Moreno et al., 2008; Krause et al., 2007; Mani et al., 2007; Gnidehou et al., 2004; Ito et al., 1986)
TH Transporters		
Serum TH-Binding Proteins	Three serum proteins (thyroxine-binding globulin; transthyretin; albumin) responsible for binding and transporting THs.	(Aqai et al., 2012; Montaño et al., 2012; Cao et al., 2011; Cao et al., 2010; Hedge et al., 2009; Marchesini et al., 2008; Hamers et al., 2006; Marchesini et al., 2006; Hallgren and Darnerud, 2002; Robbins, 2000; Cheek et al., 1999; Brouwer et al., 1998; Palha et al., 1997; Lans et al., 1994; Brouwer and Vandenberg, 1986; Brouwer et al., 1986; Nilsson and Peterson, 1975; Nilsson et al., 1975)
Membrane Transporters	Solute carrier (SLC) gene family proteins transport TH across plasma membranes. TH-specific transporters include monocarboxylate transporters (MCT) 8 & 10 and organic anion-transporting protein (OATP) 1C1.	(Murk et al., 2013; Noyes et al., 2013; Braun et al., 2012; Visser et al., 2011; Ianculescu et al., 2010; Kinne et al., 2009; Westholm et al., 2009; Friesema et al., 2008; Heuer, 2007; Friesema et al., 2003)
TH Peripheral Tissue Metabolism		
Iodothyronine Deiodinase (DIO)	Enzymes deiodinate T4 to T3 (activation) and/or inactivate T4. Three types (DIO I, II and III) function in a tissue-specific and temporal manner to modulate TH homeostasis.	(Renko et al., 2015; Pavelka, 2014; Noyes et al., 2013; Xu et al., 2013; Renko et al., 2012; Butt et al., 2011; Noyes et al., 2011; Wang and Stapleton, 2010; Crofton, 2008; Köhrle, 2008; Freyberger and Ahr, 2006; Kuiper et al., 2006; Streckfuss et al., 2005; Zavacki et al., 2005; Fekete et al., 2004; Curcio et al., 2001; Mol et al., 1999; Steinsapir et al., 1998; Morse et al., 1993; Capen and Martin, 1989)

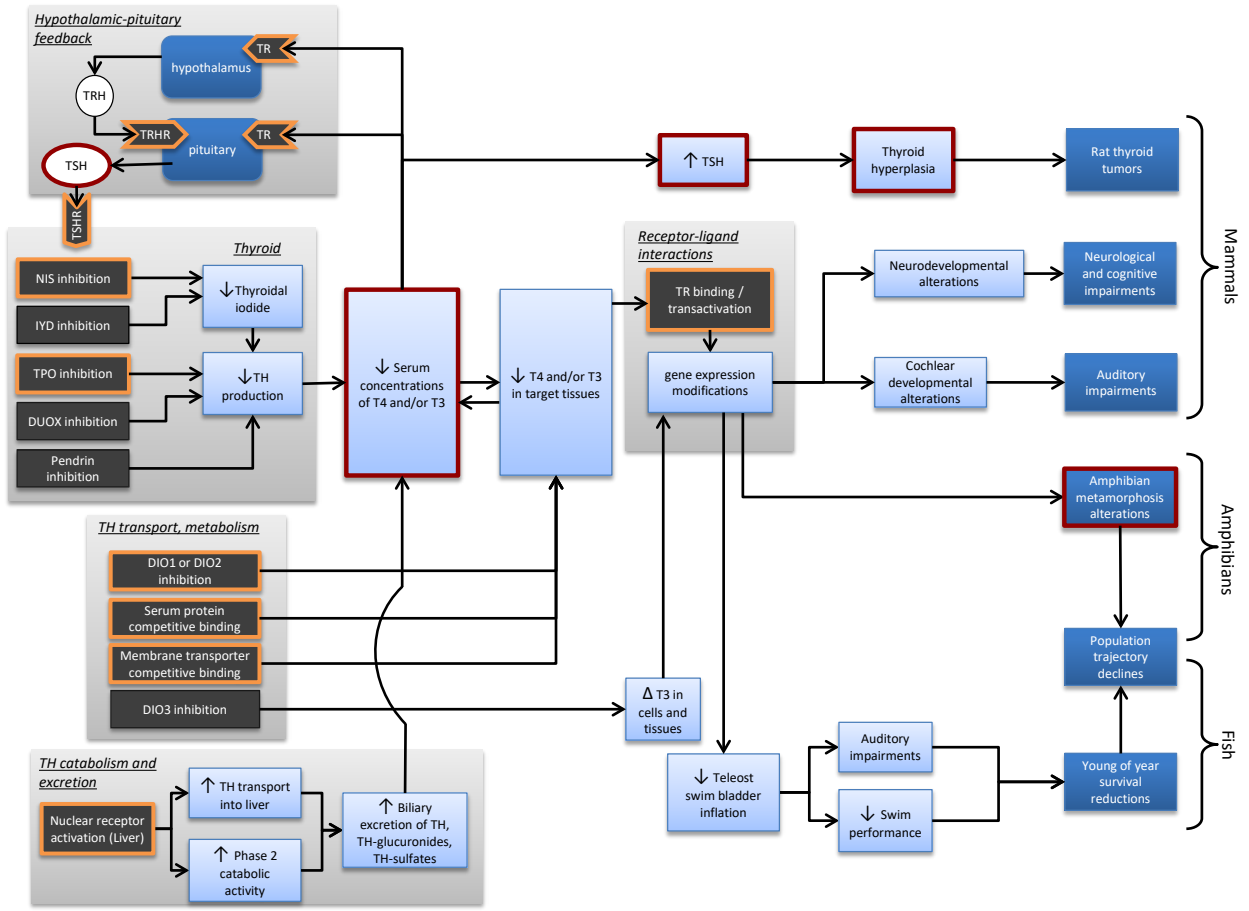
MIE	Thyroid Pathway Relevance	References
Hepatic Nuclear Receptors (NRs)	Mediate Phase 1, 2, and 3 metabolism and disposition of endogenous and exogenous chemicals, contributing to TH homeostasis.	(Paul Friedman et al., 2016; Kavlock et al., 2012; Huang et al., 2011; Raucy and Lasker, 2010; Crofton, 2008; Köhrle, 2008; Zoeller, 2007; Crofton and Zoeller, 2005; Brucker-Davis, 1998; Hurley, 1998; Capen, 1997; Visser, 1996)
Sulfation and Glucuronidation	Sulfation and glucuronidation are important hepatic and nephric pathways that regulate TH catabolism.	(Butt and Stapleton, 2013; Larson et al., 2011; Rotroff et al., 2010a; Yu et al., 2009; Wang and James, 2006; Barter and Klaassen, 1994; Visser et al., 1993)
Alanine Side-Chain	Alanine side-chains of T4 and T3 can be metabolized by oxidative decarboxylation or deamination; deamination produces thyroacetic acids and decarboxylation produces thyroanimines.	(Scanlan, 2009; Wu et al., 2005)
Receptor-Based Targets		
TRH Receptor	G protein-coupled receptors (GPCRs) (TRHR1, TRHR2) primarily in hypothalamus controlling synthesis and release of TSH.	(Engel et al., 2008; Beck-Peccoz et al., 2006) See also TRH Receptor in Table 4-3
TSH Receptor	GPCR primarily on thyroid epithelial cells controls production of THs.	(Jomaa et al., 2013; Gershengorn and Neumann, 2012; Neumann et al., 2009; Titus et al., 2008; Santini et al., 2003) See also TSH Receptor in Table 4-3
TH Receptor	Nuclear receptor activated by T3 regulates gene expression in a wide variety of cell types. Subtypes TRα (1 & 2) and TRβ (1 & 2) show tissue-specific and temporal function.	(Chapo et al., 2007; Gauger et al., 2007; You et al., 2006; Kitamura et al., 2005b; Kitamura et al., 2005a; Gauger et al., 2004; Kitamura et al., 2002; Cheek et al., 1999) See also TH Receptor in Table 4-3
TH Transcription	Many TH signaling pathways are mediated by transcription of TR responsive genes and are critical to normal development and organ system functioning.	(Freitas et al., 2014; Freitas et al., 2011; Huang et al., 2011; Invitrogen, 2010; Hofmann et al., 2009; Kojima et al., 2009; Lema et al., 2008; Zoeller and Crofton, 2005; Cheek et al., 1999) See also TH Transcription in Table 4-3

A proposed AOP network for chemically induced thyroid bioactivity is illustrated in **Figure 4-1**. This diagram illustrates known and postulated AOPs for thyroid disruption beginning with the molecular initiating events and ending with AOs. The completed AOP associated with thyroperoxidase (TPO) (AOP #42) within the OECD AOP Wiki, 'TPO Inhibition and Altered Neurodevelopment' (Crofton et al., 2017) is included in **Figure 4-1**. Additionally, there are a number of previously published mechanisms, modes-of-action, and AOPs for thyroid

disrupting chemicals (TDCs) (Murk et al., 2013; Perkins et al., 2013; U.S. EPA, 2013b; Miller et al., 2009; Crofton, 2008; Zoeller and Crofton, 2005; Brucker-Davis, 1998; Hurley, 1998; Capen and Martin, 1989; Mcclain et al., 1988). There are many MIEs that lead to common downstream key events (KEs), which are linked to a number of adverse outcomes that are species and life-stage specific.

The key purpose of depicting the thyroid system in this manner is that it portrays the variety of MIEs, as well as the interacting nature of the pathways with nodes called key events (*i.e.* AOP network) (Wittwehr et al., 2017; Becker et al., 2015; Villeneuve et al., 2014; Perkins et al., 2013). **Figure 4-1** communicates two important points relevant to screening for potential TDCs. The first point is a summary of knowledge of the biological systems involved in thyroid disruption, which involve physiological systems that include: membrane transporters; serum transporters; multiple enzymes that are responsible for synthesis and catabolism of thyroid hormones; and, nuclear receptors. The second point is a mapping of AOPs that informs identification of biological targets or MIEs.

Figure 4-1 Proposed Adverse Outcome Pathway (AOP) Network for Chemically-Induced Thyroid Bioactivity Showing the Integration of Multiple Individual AOPs



Biological linkages described may be informed by *in vitro*, *in vivo*, or computational data, and may be causal, inferential, or putative depending on the strength of the evidence. Shapes with red borders represent endpoints measured in the EDSP Tier 1 *in vivo* screening assays. MIEs with orange borders have one or more HT assays available in or in development for ToxCast/Tox21 and/or published in the peer reviewed literature. Citations and more details for each MIE and the identified HT assays are provided in **Table 4** and **Table 4-3**, respectively (For purposes of this figure, HT is considered to be 96 well or more plates).

Abbreviations: DUOX = dual oxidase; DIO = iodothyronine deiodinase; IYD = iodotyrosine deiodinase; NIS = sodium-iodide symporter; T4 = thyroxine; T3 = 3,3′,5-triiodothyronine; TH = thyroid hormone; TPO = thyroperoxidase; TR = thyroid hormone receptor; TRHR = thyrotropin releasing hormone receptor; TSH = thyroid stimulating hormone; TSHR = thyroid stimulating hormone receptor

4.3 Screening and Assay Status

Currently, the EDSP uses a two-tiered approach to screen chemicals for potential effects on estrogen, androgen and thyroid hormone systems; however, there are no assays in the EDSP framework that require reporting of thyroid specific mechanistic information, *i.e.*, information on potentially relevant MIEs illustrated in **Figure 4-1** or listed in **Table 4** (U.S. EPA, 2017b).

Assays in Tier 1 are run as a battery to inform on the potential of a chemical to be an endocrine disruptor, while Tier 2 tests provide dose-response information necessary for risk assessment (**Table 1-1**). Tier 1 screening assays that assess thyroid function include the pubertal rodent (female (U.S. EPA, 2009i) and male (U.S. EPA, 2009j) pubertal assays) and amphibian metamorphosis assays (U.S. EPA, 2009b). Tier 2 tests include an extended one generation reproductive rodent assay (OECD, 2012b), a two-generation reproductive quail assay (U.S. EPA, 2015a) and a growth and development assay in frogs (U.S. EPA, 2015c) (**Table 4**).

Table 4-2 List of EDSP Test Guidelines with Endpoints Informing on Thyroid Pathways

Neither Tier 1 nor Tier 2 contains *in vitro* assays informative on the thyroid pathways.

Tier 1 Test Guidelines	Species	Thyroid-Specific Endpoints	Thyroid-Related Endpoints
(Amphibian Metamorphosis Assay (AMA) (OPPTS: 890.1100) U.S. EPA, 2009b)	Frog	Hind Limb Length Developmental Stage Thyroid histology: • Thyroid hypertrophy/atrophy • Follicular cell hypertrophy • Follicular cell hyperplasia • Follicular lumen area • Colloid quality • Follicular cell height/shape.	Body weight Snout to vent length
(Pubertal Development and Thyroid Function in Intact Juvenile/Peripubertal Female Rats (OPPTS: 890.1450) U.S. EPA, 2009i)	Rat	Thyroid weight Serum Total T4 Serum TSH Thyroid histology: • Colloid area • Follicular cell height	Growth Body weight
(Pubertal Development and Thyroid Function in Intact Juvenile/Peripubertal Male Rats (OPPTS: 890.1500) U.S. EPA, 2009j)	Rat	Thyroid weight Serum Total T4 Serum TSH Thyroid histology: • Colloid area • Follicular cell height	Growth Body weight

Tier 2 Test Guidelines	Species	Thyroid-Specific Measures	Thyroid-Related Measures
(Avian Two-Generation Toxicity Test (OCSPP: 890.2100) U.S. EPA, 2015a)	Japanese Quail	Thyroid size Plasma Total T4 Thyroid Total T4 Egg yolk Total T4 Thyroid histology: • Follicular size and shape • Size and relative number of follicular epithelial cells • Relative quantity and quality of colloid	Body weight/Growth
(Larval Amphibian Growth and Development Assay (LAGDA) (OCSPP 890.2300) U.S. EPA, 2015c)	Frog	Development Time to NF Stage 62 (metamorphosis) Thyroid Histology: • Thyroid hypertrophy/atrophy • Follicular cell hypertrophy • Follicular cell hyperplasia	Body weight Snout to vent length
(Extended One-Generation Reproductive Toxicity Study (EOGRTS) (OECD Test No. 443) OECD, 2012b)	Rodent (rat preferred)	Thyroid weight Serum T4 and TSH Full Thyroid Histology	Neurohistopathology** Neurobehavioral tests Brain weight

**These endpoints are not required, but can be triggered.

Mapping the endpoints from these test guidelines to thyroid AOPs reveals these assays measure endpoints downstream from known MIEs, including: serum hormone concentrations and thyroid weight and histopathology in rats and quail; and, progression through developmental stages (metamorphosis) and thyroid histopathology in frogs (**Figure 4-1**). The EDSP does not currently use *in vitro* assays informative of thyroid pathway endpoints (*i.e.* MIEs and KEs) and this may limit efficient screening for potential TDCs.

There have been several reviews (OECD, 2014; Murk et al., 2013; Noyes et al., submitted) aimed at identifying available and potential HT *in vitro* assays for coverage of the MIEs for AOPs that lead to disruption of thyroid pathways. The EPA has an interest in evaluating and utilizing all available data, methods and models to identify potential TDCs, including published literature. As a guide towards this goal, the EDSP has prioritized key MIEs based on thyroid pathway and toxicological relevance, and HT status of each MIE and designated a "Rank" for each MIE (**Table 4-3**).

The ToxCast/Tox21 status for thyroid relevant assays for each MIE is also provided in **Table 4-3** with the designations: *a*) **Existing** – one or more HT assays exists in ToxCast/Tox21 and where results have been published in the peer review literature citations are provided. ToxCast/Tox21 assay information and data are available via (U.S. EPA, 2015e); *b*) **Developing** – one or more HT assays are presently being developed in ToxCast/Tox21 but assay information or results have not yet been made publicly available; *c*) **Promising** – HT assays, or assays amenable to HT, have been published in the peer reviewed literature, but have not yet been incorporated into ToxCast/Tox21; or, *d*) **R&D** (Research & Development) – there are no existing HT assays for the MIE either in ToxCast/Tox21 or the peer reviewed literature, or existing assays will require basic research and development prior to implementation. The overall goal was to determine whether HT assays/technologies currently exist for thyroid related MIEs.

Table 4-3 HT assay status and prioritization ranking of MIEs

The column titled "ToxCast/Tox21 Status" designates the availability of HT assays in ToxCast/Tox21 and in the peer reviewed literature. The column titled "ToxCast/Tox21 Assay Name" provides the names of assays for each MIE that are publicly available in ToxCast/Tox21. ToxCast/Tox21 assay information and data are available via (U.S. EPA, 2015e). The column titled "Rank" indicates prioritization based on toxicological relevance of each MIE (**Table 4-1**) and HT assay status. Orange highlighted MIEs correspond with **Figure 4-1** and have one or more HT assays available or in development in ToxCast/Tox21 and/or published in the peer reviewed literature. (For purposes of this table, HT is considered to be 96 well or more plates). The column titled "References" provides links to peer reviewed publications for HT assays associated with the MIE.

MIE	ToxCast/Tox21 Status*	ToxCast/Tox21 Assay Name	Rank	References
TH Biosynthesis in Thyroid				
Sodium-Iodide Symporter	Developing		High	(Hallinger et al., 2017)
Thyroperoxidase	Existing	(NCCT_TPO_GUA_dn) (NCCT_TPO_AUR_dn)	High	(Paul Friedman et al., 2016; Paul et al., 2014)
Pendrin	R&D		Low	
Dual Oxidase	R&D		Low	
Iodotyrosine Deiodinase	R&D		Low	
TH Transporters				
Serum TH-Binding Proteins	Promising		Medium	(Marchesini et al., 2006)
Membrane Transporters	Promising		Medium	(Dong and Wade, 2017)
TH peripheral tissue metabolism				
Iodothyronine Deiodinase	Developing		High	

MIE	ToxCast/Tox21 Status*	ToxCast/Tox21 Assay Name	Rank	References
Hepatic Nuclear Receptors	Existing	Multiple assays exist	High	
Sulfation and Glucuronidation	R&D		High	
Alanine Side-Chain	R&D		Low	
Receptor-based targets				
TRH Receptor	Existing	(NVS_GPCR_rTRH)	Medium	
TSH Receptor	Existing	(TOX21_TSHR_Agonist_ratio) (TOX21_TSHR_Antagonist_ratio)	Medium	(Titus et al., 2008) (Latif et al., 2016)
TH Receptor	Existing	(NVS_NR_hTRa_Antagonist)	Low	
TH Transcription	Existing	(TOX21_TSHR_Antagonist_ratio) (TOX21_TSHR_Antagonist_ratio) (ATG_THRa1_TRANS_up) (ATG_THRb_TRANS2_up)	Medium	(Freitas et al., 2014)

***** **Existing** – one or more HT assays exists in ToxCast/Tox21 and where results have been published in the peer review literature citations are provided. ToxCast/Tox21 assay information and data are available via (U.S. EPA, 2015e); **Developing** – one or more HT assays are presently being developed in ToxCast/Tox21 but assay information or results have not yet been made publicly available; **Promising** – HT assays, or assays amenable to HT, have been published in the peer reviewed literature, but have not yet been incorporated into ToxCast/Tox21; or, **R&D** (Research & Development) – there are no existing HT assays for the MIE either in ToxCast/Tox21 or the peer reviewed literature, or existing assays will require basic research and development prior to implementation.

Each of the references from **Table 4-3** is briefly reviewed in the following paragraphs.

An *in vitro* radioactive iodide (^{125}I-) uptake (RAIU) assay coupled to a human NIS-expressing HEK293T cell line has been developed and adapted for use as an HT assay in the ToxCast platform to identify chemicals that may inhibit ^{125}I- uptake by NIS (Hallinger et al., 2017). The EPA is preparing to publish the results of NIS inhibition screening for the ToxCast phase I chemicals (293 unique chemicals) and is continuing use of this assay to screen another ~1,500 chemicals (*i.e.* ToxCast phase II and E1K chemicals) in the near future.

The Amplex UltraRed-thyroperoxidase (AUR-TPO) assay which can identify TPO inhibition, was utilized in the ToxCast HT platform to screen 1,074 phase I and II chemicals (Friedman et al., 2016). Two additional assays were screened in parallel to identify possible sources of nonspecific assay signal loss, including chemical reactivity and nonspecific enzyme inhibition. Combining the additional assay data with the AUR-TPO assay results enabled

experimental stratification of the 300 putative *in vitro* TPO inhibitors to distinguish chemicals with high AUR-TPO assay activity that did not demonstrate confounding activities.

Cell-free transport protein-based biosensor inhibition assays were utilized to examine the binding affinity of chemicals to recombinant thyroxine binding globulin (TBG) and transthyretin (TTR) (Marchesini et al., 2008). A library of 62 hormones and environmentally relevant chemicals was tested with the two transport protein-based assays. From the 62 chemicals tested, 25 were active for TTR and 15 were active TBG, producing more than 50% inhibition. Another 9 (TTR) and 10 (TBG) chemicals produced an inhibition greater than 15% but less than 50% and were classified as weak binders.

MDCK cells stably transfected with human MCT8-expression vector (MDCK-MCT8) were used with a non-radioactive method of quantifying T3 uptake by measuring the rate of production of a chromogenic reaction of iodine with cerium and arsenate (Dong and Wade, 2017). This assay and the positive control (bromosulfalein) were used to characterize the responses to a training set of 17 chemicals (ten positives and seven negatives) and ten environmentally relevant reported TDCs. Nine of the ten positive chemicals and all the negative chemicals were confirmed, while only one of the ten environmentally relevant reported TDCs inhibited T3 uptake in the MDCK-MCT8 assay.

An assay using detection of cAMP as a marker of TSHR activation in a HEK293 cell line stably transfected with the human TSHR was utilized to screen 73,180 compounds (Titus et al., 2008). While 210 compounds were determined to be active compounds, the compound library tested was designed to identify small molecule actives which could be pharmacologically relevant. A similar effort utilized a cell line stably transfected with TSHR and a cAMP-response element-luciferase gene fusion construct to identify 62 potential small molecule agonists of the TSHR from a library of 48,224 chemicals (Latif et al., 2016). Neither of these publications represent ToxCast/Tox21 assays, however EPA is currently analyzing data from one ToxCast/Tox21 TSHR assay and is developing orthogonal assays (Paul-Friedman et al., 2017).

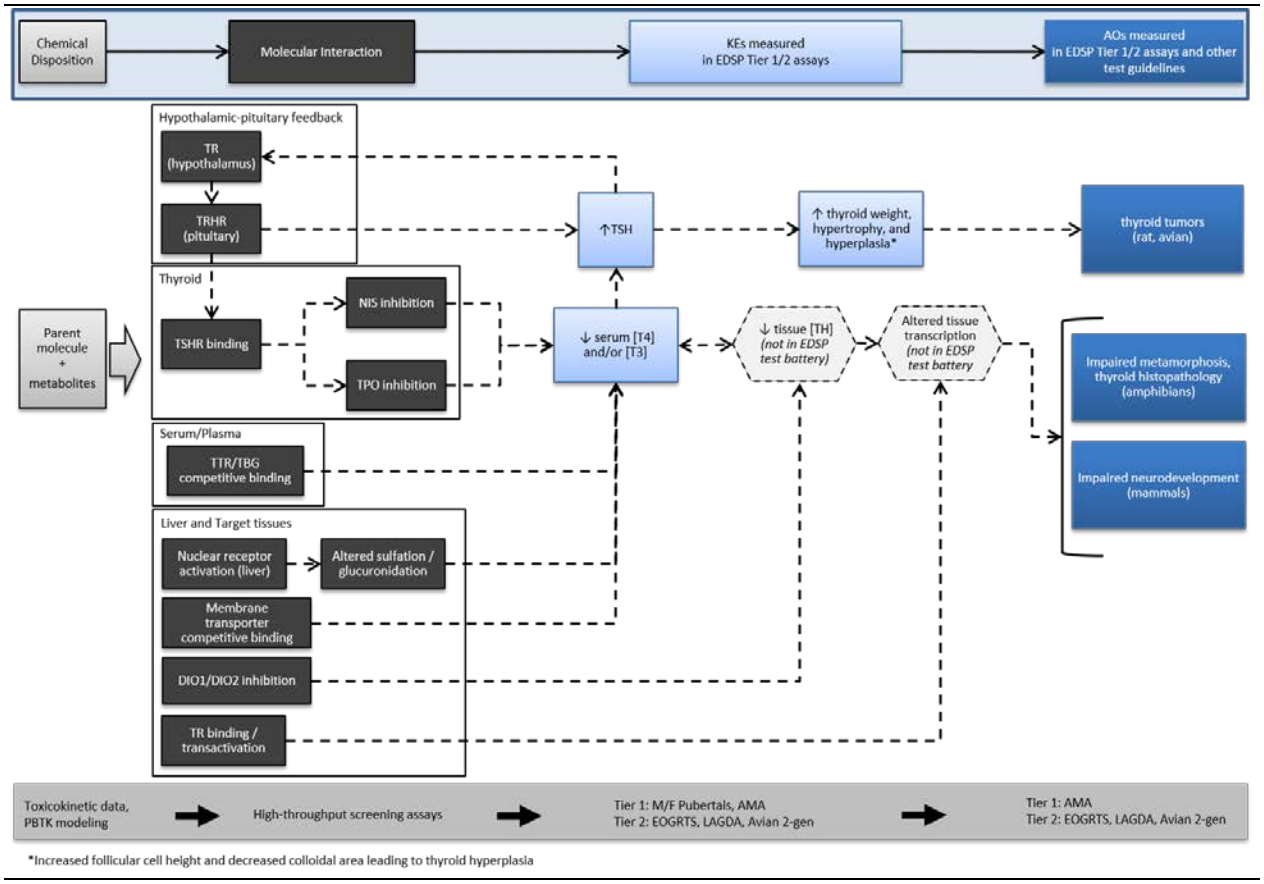
A rat pituitary cell line containing a thyroid receptor response element fused to a luciferase reporter gene (GH3.TRE_Luc) (Freitas et al., 2014) was used by the Tox21 program to screen the Tox21 chemical library (8,500 chemicals) (U.S. EPA, 2015e). Of the 1,280 compounds screened, only 6 were identified as potential TR agonists, and only 4 were identified as potential TR antagonists.

The EPA has also been developing cell-free enzyme activity assays to investigate potential chemical inhibition of DIO I, II and II. Data from a cell-free TRH receptor (TRHR)

binding assay for ToxCast chemical libraries (Phase I and II) are currently available (U.S. EPA, 2015e), and orthogonal, cell-based TRHR assays are in development as part of the Tox21 interagency collaboration. Thyroid hormone receptor (THR) binding and transcriptional-activation assays are also included in the ToxCast/Tox21 programs with data collection completed and data analysis ongoing.

To help visualize the overlap between HT assays informative of MIEs and the KEs and AOs measured in EDSP test guidelines, the EDSP developed a thyroid AOP framework (**Figure 4-2**). This framework is based on the thyroid AOP network, which integrates HT assays informative of MIEs with KEs and other apical endpoints associated with the *in vivo* EDSP Tier 1 and Tier 2 test guidelines.

Figure 4-2 AOP-Informed Framework for the Thyroid Network with Several Initiating MIEs that Can Be Measured Using In Vitro HT Assays, with Linkages to Intermediate KEs that Culminate in Adverse Outcomes (AOs) Identified by In Vivo EDSP Test Guidelines



Hexagonal boxes with dotted lines represent examples of putative KEs that are data gaps and candidates for further characterization.

4.4 Next Steps and Challenges

There are several important challenges and next steps for EDSP as it develops a HT approach for identification of potential TH disrupting chemicals. Chief among these are:

- Identification of reference chemicals for extant assays as well as for assays in development.
- Assay development
- Development of performance-based approaches (see Section 1.5**)**
- Development of an integrative strategy for analysis of assay data.
- Development of a framework for prioritization of chemical screening.

4.4.1 Identification of Reference Chemicals

EDSP is undertaking a systematic literature search for potential reference chemicals as part of a large-scale systematic review of thyroid hormone-regulated pathways. The systematic review has three major aims: Identification of reference chemicals; identification of molecular mechanisms of xenobiotic disruption of TH pathways; and development of a machine-learning-based system for optimization of large scale systematic reviews. While the systematic review project is still in the early planning stages, identification of a number of reference chemicals has already been completed. (Wegner et al., 2016) published an analysis of a set of 34 potential thyroid pathway reference chemicals for use in validation of HT *in vitro* and *in vivo* assays. These chemicals cover a range of mechanisms of action, structures, and potencies. Testing against reference chemicals selected to assess potency will provide additional important information that will be used in development of the integrative strategy for data analysis.

4.4.2 Development of Additional Assays

Assays to identify potential TDCs are lacking for several potential points of the thyroid pathways and will require development (**Table 4**) if complete assessment is to be achieved. All assays have technical limitations, thus multiple assays (2+) for each node are highly desirable. Multiple assays evaluating the same mechanism will allow for identification of possible chemical interference or other limitations that might result in false positive or negative results. Duplicative assays that use different platforms or technologies for signal detection and/or interrogate different parts of the same MIE or KE (orthogonal assays), if available and employed can be used to improve confidence in the results; however, this may not always be possible given technical or resource considerations. Ultimately, EPA envisions that the EDSP will move to full replacement of all Tier 1 *in vivo* assays with *in vitro* assays, specifically high-throughput, and

computational models that will refine, reduce and replace animal testing (U.S. EPA, 2015f). However, these models must be robust, reliable, and validated.

4.4.3 Development of an Integrative Strategy for Assay Data

The EPA has previously developed consensus models that allow use of *in vitro* data to predict chemical interactions with estrogen and androgen receptors (Kleinstreuer et al., 2017a; Judson et al., 2015). These models use quantitative scoring approaches to integrate *in vitro* assay data for multiple receptor pathway events leading to receptor activation. There are currently no such consensus prediction models available for any of the multiple targets or pathways by which thyroid systems may be disrupted. A useful tool to assist with assessment of TDCs would be a systems model that would incorporate parameters to enable toxicity predictions for many MIEs, life stages, and species and incorporate HT data to predict biological AOs. The need for modeling approaches that encompass a set of molecular events within a tissue/organ (*e.g.* thyroid or peripheral tissue) or across several tissues (*e.g.* systems model incorporating plasma transport), and with indications for multiple life stages and species, presents a complex challenge.

As part of a strategy to integrate assay data, there is a need to address how screening for some MIEs in the thyroid AOP network may identify chemicals of interest, but information from additional sources may need to be considered to reduce uncertainty and make more specific predictions (*e.g.* hepatic NR modulators). Model performance would be evaluated through use of reference chemicals that would determine model sensitivity, specificity, and accuracy for predicting thyroid-related bioactivity. These measures would then be used to establish performance-based acceptance criteria which will ensure the relevance and reproducibility of the results. The biological relevance of this approach is based on using assays which evaluate MIEs or KEs in established adverse outcome pathways.

Models that predict AOs directly from HT data will require development of additional toxicokinetic tools. Though qualitative evidence is generally strong that the MIEs described in **Figure 4-2** link to thyroid AOs, data to establish *quantitative* linkages to and from KEs are often lacking. As quantitative data become available, additional modeling approaches to further characterize KE relationships in the thyroid AOP network would be needed to predict thyroid-related AOs (Wittwehr et al., 2017; Becker et al., 2015).

4.4.4 Development of Framework for Prioritization of Chemical Screening

Prioritization of chemicals for EDSP screening has historically been based on human exposure potential and production volume (U.S. EPA, 2012, 1998a). The AOP framework

presented in this white paper (**Figure 4-2**) could be used to prioritize chemicals for targeted testing based on their bioactivity as informed or determined by HT assays or computational models for relevant MIEs. The EDSP plans to develop a robust and detailed AOP-based prioritization framework for potential TDCs and such a prioritization will consider HT bioactivity screening results as well as results from models of higher biological complexity (*in vitro* and/or *in vivo*) and/or combined with other informative descriptors. Therefore, a positive result for a chemical in any single HT thyroid related assay should not necessarily be expected to lead to a Tier 1 test order for male/female pubertal or AMA. Once these HT assay based models and AOP-based prioritization framework are developed, a decision tree flow chart will map next steps based on thyroid bioactivity and empirical and predicted exposure or toxicokinetic parameters (*i.e.* absorption; distribution; metabolism; excretion).

4.5 Conclusions

Several *in vitro* HT assays are now available to help understand the potential for chemicals to interact with thyroid pathways. The thyroid AOP network described in this white paper will inform the integration of *in vitro* HT data to describe and understand relationships between thyroid-based AOPs and outcomes of existing EDSP Tier 1 assays. Next steps will involve designing approaches to use existing HT data to prioritize chemicals for testing in the EDSP Tier 1 screening battery, with future efforts to develop predictive models for using HT data in WoE evaluations.

5. References

Abdel-Khalik, J; Bjorklund, E; Hansen, M. (2013). Development of a solid phase extraction method for the simultaneous determination of steroid hormones in H295R cell line using liquid chromatography-tandem mass spectrometry. J Chromatogr B Analyt Technol Biomed Life Sci 935: 61-69. http://dx.doi.org/10.1016/j.jchromb.2013.07.013

Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle. In E Parzen; K Tanabe; G Kitagawa (Eds.), Selected papers of Hirotugu Akaike (pp. 199-213). New York, NY: Springer. http://dx.doi.org/10.1007/978-1-4612-1694-0_15

Ankley, GT; Bennett, RS; Erickson, RJ; Hoff, DJ; Hornung, MW; Johnson, RD; Mount, DR; Nichols, JW; Russom, CL; Schmieder, PK; Serrrano, JA; Tietge, JE; Villeneuve, DL. (2010). Adverse outcome pathways: a conceptual framework to support ecotoxicology research and risk assessment [Review]. Environ Toxicol Chem 29: 730-741. http://dx.doi.org/10.1002/etc.34

Aqai, P; Fryganas, C; Mizuguchi, M; Haasnoot, W; Nielen, MW. (2012). Triple bioaffinity mass spectrometry concept for thyroid transporter ligands. Anal Chem 84: 6488-6493. http://dx.doi.org/10.1021/ac300543u

Asser, L; Hescot, S; Viengchareun, S; Delemer, B; Trabado, S; Lombes, M. (2014). Autocrine positive regulatory feedback of glucocorticoid secretion: Glucocorticoid receptor directly impacts H295R human adrenocortical cell function. Mol Cell Endocrinol 395: 1-9. http://dx.doi.org/10.1016/j.mce.2014.07.012

Barter, RA; Klaassen, CD. (1994). Reduction of thyroid hormone levels and alteration of thyroid function by four representative UDP-glucuronosyltransferase inducers in rats. Toxicol Appl Pharmacol 128: 9-17. http://dx.doi.org/10.1006/taap.1994.1174

Beck-Peccoz, P; Persani, L; Calebiro, D; Bonomi, M; Mannavola, D; Campi, I. (2006). Syndromes of hormone resistance in the hypothalamic-pituitary-thyroid axis. Best Pract Res Clin Endocrinol Metab 20: 529-546. http://dx.doi.org/10.1016/j.beem.2006.11.001

Becker, RA; Ankley, GT; Edwards, SW; Kennedy, SW; Linkov, I; Meek, B; Sachana, M; Segner, H; Van Der Burg, B; Villeneuve, DL; Watanabe, H; Barton-Maclaren, TS. (2015). Increasing Scientific Confidence in Adverse Outcome Pathways: Application of Tailored Bradford-Hill Considerations for Evaluating Weight of Evidence. Regul Toxicol Pharmacol 72: 514-537. http://dx.doi.org/10.1016/j.yrtph.2015.04.004

Bizhanova, A; Kopp, P. (2011). Controversies Concerning the Role of Pendrin as an Apical Iodide Transporter in Thyroid Follicular Cells. Cell Physiol Biochem 28: 485-490. http://dx.doi.org/10.1159/000335103

Braun, D; Kim, TD; Le Coutre, P; Kohrle, J; Hershman, JM; Schweizer, U. (2012). Tyrosine kinase inhibitors noncompetitively inhibit MCT8-mediated iodothyronine transport. J Clin Endocrinol Metab 97: E100-E105. http://dx.doi.org/10.1210/jc.2011-1837

Breen, MS; Breen, M; Terasaki, N; Yamazaki, M; Conolly, RB. (2010). Computational model of steroidogenesis in human H295R cells to predict biochemical response to endocrine-active chemicals: Model development for metyrapone. Environ Health Perspect 118: 265-272. http://dx.doi.org/10.1289/ehp.0901107

Brouwer, A; Morse, DC; Lans, MC; Schuur, AG; Murk, AJ; Klasson-Wehler, E; Bergman, Å; Visser, TJ. (1998). Interactions of persistent environmental organohalogens with the thyroid hormone system: Mechanisms and possible consequences for animal and human

health [Review]. Toxicol Ind Health 14: 59-84. http://dx.doi.org/10.1177/074823379801400107
Brouwer, A; Vandenberg, KJ. (1986). Binding of a metabolite of 3,4,3',4'-tetrachlorobiphenyl to transthyretin reduces serum vitamin-a transport by inhibiting the formation of the protein complex carrying both retinol and thyroxine. Toxicol Appl Pharmacol 85: 301-312. http://dx.doi.org/10.1016/0041-008X(86)90337-6
Brouwer, A; Vandenberg, KJ; Blaner, WS; Goodman, DS. (1986). TRANSTHYRETIN (PREALBUMIN) BINDING OF PCBS, A MODEL FOR THE MECHANISM OF INTERFERENCE WITH VITAMIN A AND THYROID-HORMONE METABOLISM. Chemosphere 15: 1699-1706. http://dx.doi.org/10.1016/0045-6535(86)90457-1
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: 8804-8814. http://dx.doi.org/10.1021/acs.est.5b02641
Brucker-Davis, F. (1998). Effects of environmental synthetic chemicals on thyroid function [Review]. Thyroid 8: 827-856. http://dx.doi.org/10.1089/thy.1998.8.827
Buss, JM; Mctamney, PM; Rokita, SE. (2012). Expression of a soluble form of iodotyrosine deiodinase for active site characterization by engineering the native membrane protein from Mus musculus. Protein Sci 21: 351-361. http://dx.doi.org/10.1002/pro.2020
Butt, CM; Stapleton, HM. (2013). Inhibition of Thyroid Hormone Sulfotransferase Activity by Brominated Flame Retardants and Halogenated Phenolics. Chem Res Toxicol 26: 1692-1702. http://dx.doi.org/10.1021/tx400342k
Butt, CM; Wang, D; Stapleton, HM. (2011). Halogenated phenolic contaminants inhibit the in vitro activity of the thyroid-regulating deiodinases in human liver. Toxicol Sci 124: 339-347. http://dx.doi.org/10.1093/toxsci/kfr117
Cao, J; Guo, LH; Wan, B; Wei, Y. (2011). In vitro fluorescence displacement investigation of thyroxine transport disruption by bisphenol A. J Environ Sci 23: 315-321.
Cao, J, ie; Lin, Y; Guo, LH; Zhang, A, iQ; Wei, Y, in; Yang, Y, u. (2010). Structure-based investigation on the binding interaction of hydroxylated polybrominated diphenyl ethers with thyroxine transport proteins. Toxicology 277: 20-28. http://dx.doi.org/10.1016/j.tox.2010.08.012
Capen, CC. (1997). Mechanistic data and risk assessment of selected toxic end points of the thyroid gland [Review]. Toxicol Pathol 25: 39-48. http://dx.doi.org/10.1177/019262339702500109
Capen, CC. (1998). Correlation of mechanistic data and histopathology in the evaluation of selected toxic endpoints of the endocrine system [Review]. Toxicol Lett 102-103: 405-409.
Capen, CC; Martin, SL. (1989). The effects of xenobiotics on the structure and function of thyroid follicular and C-cells [Review]. Toxicol Pathol 17: 266-293.
Carvalho, DP; Dupuy, C. (2013). Role of the NADPH oxidases DUOX and NOX4 in thyroid oxidative stress. Eur Thyroid J 2: 160-167. http://dx.doi.org/10.1159/000354745
Chang, HC; Doerge, DR. (2000). Dietary genistein inactivates rat thyroid peroxidase in vivo without an apparent hypothyroid effect. Toxicol Appl Pharmacol 168: 244-252. http://dx.doi.org/10.1006/taap.2000.9019
Chang, X; Kleinstreuer, N; Ceger, P; Hsieh, JH; Allen, D; Casey, W. (2014). Application of Reverse Dosimetry to Compare In Vitro and In Vivo Estrogen Receptor Activity. 1: 33-44. http://dx.doi.org/10.1089/aivt.2014.0005

Chapo, J; Peng, Y; Pitts, KR. (2007). A phosphorimager-based filter binding thyroid hormone receptor competition assay for chemical screening. J Pharmacol Toxicol Methods 56: 28-33.
Cheek, AO; Kow, K; Chen, J; Mclachlan, JA. (1999). Potential mechanisms of thyroid disruption in humans: Interaction of organochlorine compounds with thyroid receptor, transthyretin, and thyroid-binding globulin. Environ Health Perspect 107: 273-278. http://dx.doi.org/10.2307/3434593
Cianchetta, S; di Bernardo, J; Romeo, G; Rhoden, KJ. (2010). Perchlorate transport and inhibition of the sodium iodide symporter measured with the yellow fluorescent protein variant YFP-H148Q/I152L. Toxicol Appl Pharmacol 243: 372-380. http://dx.doi.org/10.1016/j.taap.2009.12.004
Crockford, SJ. (2009). Evolutionary roots of iodine and thyroid hormones in cellcell signaling. Integr Comp Biol 49: 155-166. http://dx.doi.org/10.1093/icb/icp053
Crofton, KM. (2008). Thyroid disrupting chemicals: Mechanisms and mixtures [Review]. Int J Androl 31: 209-223. http://dx.doi.org/10.1111/j.1365-2605.2007.00857.x
Crofton, KM; Gilbert, M; Paul-Friedman, KP; Demeneix, B; Marty, MS; Zoeller, RT. (2017). Inhibition of thyroperoxidase and subsequent adverse neurodevelopmental outcomes in mammals. Available online at https://aopwiki.org/aops/42
Crofton, KM; Zoeller, RT. (2005). Mode of action: neurotoxicity induced by thyroid hormone disruption during development--hearing loss resulting from exposure to PHAHs [Review]. Crit Rev Toxicol 35: 757-769. http://dx.doi.org/10.1080/10408440591007304
Curcio, C; Baqui, MMA; Salvatore, D; Rihn, BH; Mohr, S; Harney, JW; Larsen, PR; Bianco, AC. (2001). The human Type 2 iodothyronine deiodinase is a selenoprotein highly expressed in a mesothelioma cell line. J Biol Chem 276: 30183-30187. http://dx.doi.org/10.1074/jbc.C100325200
De Maesschalck, R; Jouan-Rimbaud, D; Massart, DL. (2000). The Mahalanobis distance. Chemometr Intell Lab Syst 50: 1-18. http://dx.doi.org/Doi 10.1016/S0169-7439(99)00047-7
Di Bernardo, J; Iosco, C; Rhoden, KJ. (2011). Intracellular anion fluorescence assay for sodium/iodide symporter substrates. Anal Biochem 415: 32-38. http://dx.doi.org/10.1016/j.ab.2011.04.017
Divi, RL; Chang, HC; Doerge, DR. (1997). Anti-thyroid isoflavones from soybean: Isolation, characterization, and mechanisms of action. Biochem Pharmacol 54: 1087-1096. http://dx.doi.org/10.1016/s0006-2952(97)00301-8
Divi, RL; Doerge, DR. (1994). Mechanism-based inactivation of lactoperoxidase and thyroid peroxidase by resorcinol derivatives. Biochemistry 33: 9668-9674. http://dx.doi.org/10.1021/bi00198a036
Doerge, DR; Chang, HC; Divi, RL; Churchwell, MI. (1998). Mechanism for inhibition of thyroid peroxidase by leucomalachite green. Chem Res Toxicol 11: 1098-1104. http://dx.doi.org/10.1021/tx970226o
Doerge, DR; Decker, CJ. (1994). Inhibition of peroxidase-catalyzed reactions by arylamines: mechanism for the anti-thyroid action of sulfamethazine. Chem Res Toxicol 7: 164-169. http://dx.doi.org/10.1021/tx00038a008
Doerge, DR; Takazawa, RS. (1990). Mechanism of thyroid peroxidase inhibition by ethylenethiourea. Chem Res Toxicol 3: 98-101. http://dx.doi.org/10.1021/tx00014a003

Dong, H; Wade, MG. (2017). Application of a nonradioactive assay for high throughput screening for inhibition of thyroid hormone uptake via the transmembrane transporter MCT8. Toxicol In Vitro 40: 234-242.
Dossena, S; Rodighiero, S; Vezzoli, V; Bazzini, C; Sironi, C; Meyer, G; Furst, J; Ritter, M; Garavaglia, ML; Fugazzola, L; Persani, L; Zorowka, P; Storelli, C; Beck-Peccoz, P; Botta, G; Paulmichl, M. (2006). Fast fluorometric method for measuring pendrin (SLC26A4) Cl-/I- transport activity. Cell Physiol Biochem 18: 67-74. http://dx.doi.org/10.1159/000095164
Engel, S; Skoumbourdis, AP; Childress, J; Neumann, S; Deschamps, JR; Thomas, CJ; Colson, AO; Costanzi, S; Gershengorn, MC. (2008). A virtual screen for diverse ligands: Discovery of selective G protein-coupled receptor antagonists. J Am Chem Soc 130: 5115-5123. http://dx.doi.org/10.1021/ja0776201
Fekete, C; Gereben, B; Doleschall, M; Harney, JW; Dora, JM; Bianco, AC; Sarkar, S; Liposits, Z; Rand, W; Emerson, C; Kacskovics, I; Larsen, PR; Lechan, RM. (2004). Lipopolysaccharide induces type 2 iodothyronine deiodinase in the mediobasal hypothalamus: Implications for the nonthyroidal illness syndrome. Endocrinology 145: 1649-1655. http://dx.doi.org/10.1210/en.2003-1439
Ferraldeschi, R; Sharifi, N; Auchus, RJ; Attard, G. (2013). Molecular pathways: Inhibiting steroid biosynthesis in prostate cancer. Clin Cancer Res 19: 3353-3359. http://dx.doi.org/10.1158/1078-0432.CCR-12-0931
Filer, D. (2015). The Toxcast analysis pipeline: An R package for processing and modeling chemical screening data. Available online at https://www3.epa.gov/research/COMPTOX/R_Programming_Language_Nov_2015/Pipeline_Overview.pdf
Filer, DL; Kothiya, P; Setzer, RW; Judson, RS; Martin, MT. (2016). tcpl: The ToxCast Pipeline for High-Throughput Screening Data. Bioinformatics 33: 618-620. http://dx.doi.org/10.1093/bioinformatics/btw680
Fortunato, RS; De Souza, ECL; Hassani, RAE; Boufraqech, M; Weyemi, U; Talbot, M; Lagente-Chevallier, O; De Carvalho, DP; Bidart, JM; Schlumberger, M; Dupuy, C. (2010). Functional consequences of dual oxidase-thyroperoxidase interaction at the plasma membrane. J Clin Endocrinol Metab 95: 5403-5411. http://dx.doi.org/10.1210/jc.2010-1085
Freitas, J; Cano, P; Craig-Veit, C; Goodson, ML; Furlow, JD; Murk, AJ. (2011). Detection of thyroid hormone receptor disruptors by a novel stable in vitro reporter gene assay. Toxicol In Vitro 25: 257-266. http://dx.doi.org/10.1016/j.tiv.2010.08.013
Freitas, J; Miller, N; Mengeling, BJ; Xia, M; Huang, R; Houck, K; Rietjens, IM; Furlow, JD; Murk, AJ. (2014). Identification of thyroid hormone receptor active compounds using a quantitative high-throughput screening platform. 8: 36-46. http://dx.doi.org/10.2174/2213988501408010036
Freyberger, A; Ahr, HJ. (2006). Studies on the goitrogenic mechanism of action of N,N,N',N'-tetramethylthiourea. Toxicology 217: 169-175. http://dx.doi.org/10.1016/j.tox.2005.09.005
Friedman, KP; Watt, ED; Hornung, MW; Hedge, JM; Judson, RS; Crofton, KM; Houck, KA; Simmons, SO. (2016). Tiered High-Throughput Screening Approach to Identify Thyroperoxidase Inhibitors Within the ToxCast Phase I and II Chemical Libraries. Toxicol Sci 151: 160-180. http://dx.doi.org/10.1093/toxsci/kfw034

Friesema, EC; Jansen, J; Jachtenberg, JW; Visser, WE; Kester, MH; Visser, TJ. (2008). Effective cellular uptake and efflux of thyroid hormone by human monocarboxylate transporter 10. Mol Endocrinol 22: 1357-1369. http://dx.doi.org/10.1210/me.2007-0112

Friesema, ECH; Ganguly, S; Abdalla, A; Fox, JEM; Halestrap, AP; Visser, TJ. (2003). Identification of monocarboxylate transporter 8 as a specific thyroid hormone transporter. J Biol Chem 278: 40128-40135. http://dx.doi.org/10.1074/jbc.M300909200

Gauger, KJ; Giera, S; Sharlin, DS; Bansal, R; Iannacone, E; Zoeller, RT. (2007). Polychlorinated biphenyls 105 and 118 form thyroid hormone receptor agonists after cytochrome P4501A1 activation in rat pituitary GH3 cells. Environ Health Perspect 115: 1623-1630. http://dx.doi.org/10.1289/ehp.10328

Gauger, KJ; Kato, Y; Haraguchi, K; Lehmler, H, -J; Robertson, LW; Bansal, R; Zoeller, RT. (2004). Polychlorinated biphenyls (PCBs) exert thyroid hormone-like effects in the fetal rat brain but do not bind to thyroid hormone receptors. Environ Health Perspect 112: 516-523.

Gazdar, AF; Oie, HK; Shackleton, CH; Chen, TR; Triche, TJ; Myers, CE; Chrousos, GP; Brennan, MF; Stein, CA; La Rocca, RV. (1990). Establishment and characterization of a human adrenocortical carcinoma cell line that expresses multiple pathways of steroid biosynthesis. Cancer Res 50: 5488-5496.

Gershengorn, MC; Neumann, S. (2012). Update in TSH Receptor Agonists and Antagonists. 97: 4287-4292. http://dx.doi.org/10.1210/jc.2012-3080

Gnidehou, S; Caillou, B; Talbot, M; Ohayon, R; Kaniewski, J; Noel-Hudson, MS; Morand, S; Agnangji, D; Sezan, A; Courtin, F; Virion, A; Dupuy, C. (2004). Iodotyrosine dehalogenase 1 (DEHAL1) is a transmembrane protein involved in the recycling of iodide close to the thyroglobulin iodination site. FASEB J 18: 1574-1576. http://dx.doi.org/10.1096/fj.04-2023fje

Gracia, T; Hilscherova, K; Jones, PD; Newsted, JL; Zhang, X; Hecker, M; Higley, EB; Sanderson, JT; Yu, RM; Wu, RS; Giesy, JP. (2006). The H295R system for evaluation of endocrine-disrupting effects. Ecotoxicol Environ Saf 65: 293-305. http://dx.doi.org/10.1016/j.ecoenv.2006.06.012

Gronemeyer, H; Gustafsson, JA; Laudet, V. (2004). Principles for modulation of the nuclear receptor superfamily [Review]. Nat Rev Drug Discov 3: 950-964. http://dx.doi.org/10.1038/nrd1551

Haggard, DE; Karmaus, AL; Martin, MT; Judson, RS; Setzer, RW; Friedman, KP. Using a high-throughput steroidogenesis assay to predict effects on estrogen and androgen production or overall perturbation of the steroid biosynthesis pathway.

Hallgren, S; Darnerud, PO. (2002). Polybrominated diphenyl ethers (PBDEs), polychlorinated biphenyls (PCBs) and chlorinated paraffins (CPs) in rats-testing interactions and mechanisms for thyroid hormone effects. Toxicology 177: 227-243. http://dx.doi.org/10.1016/S0300-483X(02)00222-6

Hallinger, DR; Murr, AS; Buckalew, AR; Simmons, SO; Stoker, TE; Laws, SC. (2017). Development of a screening approach to detect thyroid disrupting chemicals that inhibit the human sodium iodide symporter (NIS). Toxicol In Vitro 40: 66-78. http://dx.doi.org/http://dx.doi.org/10.1016/j.tiv.2016.12.006

Hamers, T; Kamstra, JH; Sonneveld, E; Murk, AJ; Kester, MH; Andersson, PL; Legler, J; Brouwer, A. (2006). In vitro profiling of the endocrine-disrupting potency of brominated flame retardants. Toxicol Sci 92: 157-173. http://dx.doi.org/10.1093/toxsci/kfj187

Hansen, CH; Larsen, LW; Sorensen, AM; Halling-Sorensen, B; Styrishave, B. (2017). The six most widely used selective serotonin reuptake inhibitors decrease androgens and increase estrogens in the H295R cell line. Toxicol In Vitro 41: 1-11. http://dx.doi.org/10.1016/j.tiv.2017.02.001

Hartig, PC; Bobseine, KL; Britt, BH; Cardon, MC; Lambright, CR; Wilson, VS; Gray, LE. (2002). Development of two androgen receptor assays using adenoviral transduction of MMTV-luc reporter and/or hAR for endocrine screening. Toxicol Sci 66: 82-90. http://dx.doi.org/10.1093/toxsci/66.1.82

Hartig, PC; Cardon, MC; Blystone, CR; Gray, LE; Wilson, VS. (2008). High throughput adjustable 96-well plate assay for androgen receptor binding: a practical approach for EDC screening using the chimpanzee AR. Toxicol Lett 181: 126-131. http://dx.doi.org/10.1016/j.toxlet.2008.07.008

Hecker, M; Giesy, J; Timm, G. (2008). Multi-Laboratory Validation of the H295R Steroidogenesis Assay to Identify Modulators of Testosterone and Estradiol Production. 89.

Hecker, M; Hollert, H; Cooper, R; Vinggaard, AM; Akahori, Y; Murphy, M; Nellemann, C; Higley, E; Newsted, J; Laskey, J; Buckalew, A; Grund, S; Maletz, S; Giesy, J; Timm, G. (2011). The OECD validation program of the H295R steroidogenesis assay: Phase 3. Final inter-laboratory validation study. Environ Sci Pollut Res Int 18: 503-515. http://dx.doi.org/10.1007/s11356-010-0396-x

Hecker, M; Newsted, JL; Murphy, MB; Higley, EB; Jones, PD; Wu, R; Giesy, JP. (2006). Human adrenocarcinoma (H295R) cells for rapid in vitro determination of effects on steroidogenesis: hormone production. Toxicol Appl Pharmacol 217: 114-124. http://dx.doi.org/10.1016/j.taap.2006.07.007

Hedge, JM; Devito, MJ; Crofton, KM. (2009). In vivo acute exposure to polychlorinated biphenyls: effects on free and total thyroxine in rats. Int J Toxicol 28: 382-391. http://dx.doi.org/10.1177/1091581809344631

Heuer, H. (2007). The importance of thyroid hormone transporters for brain development and function. Best Pract Res Clin Endocrinol Metab 21: 265-276. http://dx.doi.org/10.1016/j.beem.2007.03.003

Heyland, A; Moroz, LL. (2005). Cross-kingdom hormonal signaling: an insight from thyroid hormone functions in marine larvae [Review]. J Exp Biol 208: 4355-4361. http://dx.doi.org/10.1242/jeb.01877

Hilscherova, K; Jones, PD; Gracia, T; Newsted, JL; Zhang, X; Sanderson, JT; Yu, RM; Wu, RS; Giesy, JP. (2004). Assessment of the effects of chemicals on the expression of ten steroidogenic genes in the H295R cell line using real-time PCR. Toxicol Sci 81: 78-89. http://dx.doi.org/10.1093/toxsci/kfh191

Hofmann, PJ; Schomburg, L; Köhrle, J. (2009). Interference of endocrine disrupters with thyroid hormone receptor-dependent transactivation. Toxicol Sci 110: 125-137. http://dx.doi.org/10.1093/toxsci/kfp086

Huang, R; Xia, M; Cho, MH; Sakamuru, S; Shinn, P; Houck, KA; Dix, DJ; Judson, RS; Witt, KL; Kavlock, RJ; Tice, RR; Austin, CP. (2011). Chemical genomics profiling of environmental chemical modulation of human nuclear receptors. Environ Health Perspect 119: 1142-1148. http://dx.doi.org/10.1289/ehp.1002952

Hurley, PM. (1998). Mode of carcinogenic action of pesticides inducing thyroid follicular cell tumors in rodents. Environ Health Perspect 106: 437-445. http://dx.doi.org/10.1289/ehp.98106437

Ianculescu, AG; Friesema, ECH; Visser, TJ; Giacomini, KM; Scanlan, TS. (2010). Transport of thyroid hormones is selectively inhibited by 3-iodothyronamine. Mol Biosyst 6: 1403-1410. http://dx.doi.org/10.1039/b926588k

ICCVAM (Interagency Coordinating Committee on the Validation of Alternative Methods). (2003). ICCVAM evaluation of in vitro test methods for detecting potential endocrine disruptors: Estrogen receptor and androgen receptor binding and transcriptional activation assays. (NIH Publication No. 03-4503). RTP, NC. https://ntp.niehs.nih.gov/iccvam/docs/endo_docs/edfinalrpt0503/edfinrpt.pdf

Invitrogen. (2010). GeneBLAzer. TR Alpha HEK 293T DA and TR Alpha-UAS-bla HEK 293T Cell-based Assay. http://tools.invitrogen.com/content/sfs/manuals/geneblazer_TRalpha_UASblaHEK293T_man.pdf. [Accessed 12/20/16].

Ito, A; Watanabe, H; Naito, M; Aoyama, H; Nakagawa, Y; Fujimoto, N. (1986). Induction of thyroid tumors in (C57BL/6N x C3H/N)F1 mice by oral administration of 9-3',4',5',6'-tetrachloro-o-carboxy phenyl-6-hydroxy-2,4,5,7-tetraiodo-3-isoxanthone sodium (Food Red 105, rose bengal B). J Natl Cancer Inst 77: 277-281.

Jomaa, B; Aarts, J; de Haan, LHJ; Peijnenburg, A; Bovee, TFH; Murk, AJ; Rietjens, I. (2013). In Vitro Pituitary and Thyroid Cell Proliferation Assays and Their Relevance as Alternatives to Animal Testing. 30: 293-307.

Judson, R; Houck, K; Martin, M; Richard, AM; Knudsen, TB; Shah, I; Little, S; Wambaugh, J; Setzer, RW; Kothya, P; Phuong, J; Filer, D; Smith, D; Reif, D; Rotroff, D; Kleinstreuer, N; Sipes, 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 152: 323-339. http://dx.doi.org/10.1093/toxsci/kfw092

Judson, RS; Kavlock, RJ; Setzer, RW; Hubal, EA; Martin, MT; Knudsen, TB; Houck, KA; Thomas, RS; Wetmore, BA; Dix, DJ. (2011). Estimating toxicity-related biological pathway altering doses for high-throughput chemical risk assessment. Chem Res Toxicol 24: 451-462. http://dx.doi.org/10.1021/tx100428e

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. Toxicol Sci 148: 137-154. http://dx.doi.org/10.1093/toxsci/kfv168

Karmaus, AL; Toole, CM; Filer, DL; Lewis, KC; Martin, MT. (2016). High-Throughput Screening of Chemical Effects on Steroidogenesis Using H295R Human Adrenocortical Carcinoma Cells. Toxicol Sci 150: 323-332. http://dx.doi.org/10.1093/toxsci/kfw002

Kavlock, R; Chandler, K; Houck, K; Hunter, S; Judson, R; Kleinstreuer, N; Knudsen, T; Martin, M; Padilla, S; Reif, D; Richard, A; Rotroff, D; Sipes, N; Dix, D. (2012). Update on EPA's ToxCast program: Providing high throughput decision support tools for chemical risk management. Chem Res Toxicol 25: 1287-1302. http://dx.doi.org/10.1021/tx3000939

Kinne, A; Roth, S; Biebermann, H; Kohrle, J; Gruters, A; Schweizer, U. (2009). Surface translocation and tri-iodothyronine uptake of mutant MCT8 proteins are cell type-dependent. J Mol Endocrinol 43: 263-271. http://dx.doi.org/10.1677/jme-09-0043

Kitamura, S; Jinno, N; Ohta, S; Kuroki, H; Fujimoto, N. (2002). Thyroid hormonal activity of the flame retardants tetrabromobisphenol A and tetrachlorobisphenol A. Biochem Biophys Res Commun 293: 554-559. http://dx.doi.org/10.1016/S0006-291X(02)00262-0

Kitamura, S; Jinno, N; Suzuki, T; Sugihara, K; Ohta, S; Kuroki, H; Fujimoto, N. (2005a). Thyroid hormone-like and estrogenic activity of hydroxylated PCBs in cell culture. Toxicology 208: 377-387. http://dx.doi.org/10.1016/j.tox.2004.11.037

Kitamura, S; Kato, T; Iida, M; Jinno, N; Suzuki, T; Ohta, S; Fujimoto, N; Hanada, H; Kashiwagi, K; Kashiwagi, A. (2005b). Anti-thyroid hormonal activity of tetrabromobisphenol A, a flame retardant, and related compounds: Affinity to the mammalian thyroid hormone receptor, and effect on tadpole metamorphosis. Life Sci 76: 1589-1601. http://dx.doi.org/10.1016/j.lfs.2004.08.030

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. (2017a). Development and validation of a computational model for androgen receptor activity. Chem Res Toxicol 30: 946-964. http://dx.doi.org/10.1021/acs.chemrestox.6b00347

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. (2017b). Supporting information: Development and validation of a computational model for androgen receptor activity [Supplemental Data]. Chem Res Toxicol 30.

Kleinstreuer, NC; Ceger, PC; Allen, DG; Strickland, J; Chang, X; Hamm, JT; Casey, WM. (2016). A curated database of rodent uterotrophic bioactivity. Environ Health Perspect 124: 556-562. http://dx.doi.org/10.1289/ehp.1510183

Knudsen, TB; Houck, KA; Sipes, NS; Singh, AV; Judson, RS; Martin, MT; Weissman, A; Kleinstreuer, NC; Mortensen, HM; Reif, DM; Rabinowitz, JR; Setzer, RW; Richard, AM; Dix, DJ; Kavlock, RJ. (2011). Activity profiles of 309 ToxCast chemicals evaluated across 292 biochemical targets. Toxicology 282: 1-15. http://dx.doi.org/10.1016/j.tox.2010.12.010

Köhrle, J. (2008). Environment and endocrinology: The case of thyroidology [Review]. Ann Endocrinol 69: 116-122. http://dx.doi.org/10.1016/j.ando.2008.02.008

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

Kojima, H; Takeuchi, S; Uramaru, N; Sugihara, K; Yoshida, T; Kitamura, S. (2009). Nuclear Hormone Receptor Activity of Polybrominated Diphenyl Ethers and Their Hydroxylated and Methoxylated Metabolites in Transactivation Assays Using Chinese Hamster Ovary Cells. Environ Health Perspect 117: 1210-1218. http://dx.doi.org/10.1289/ehp.0900753

Krause, K; Karger, S; Gimm, O; Sheu, SY; Dralle, H; Tannapfel, A; Schmid, KW; Dupuy, C; Fuhrer, D. (2007). Characterisation of DEHAL1 expression in thyroid pathologies. Eur J Endocrinol 156: 295-301. http://dx.doi.org/10.1530/eje-06-0596

Kuiper, G; Klootwijk, W; Dubois, GM; Destree, O; Darras, VM; Van Der Geyten, S; Demeneix, B; Visser, TJ. (2006). Characterization of recombinant Xenopus laevis type I iodothyronine deiodinase: substitution of a proline residue in the catalytic center by

serine (Pro132Ser) restores sensitivity to 6-propyl-2-thiouracil. Endocrinology 147: 3519-3529. http://dx.doi.org/10.1210/en.2005-0711

Lans, MC; Spiertz, C; Brouwer, A; Koeman, JH. (1994). Different competition of thyroxine binding to transthyretin and thyroxine-binding globulin by hydroxy-PCBs, PCDDs and PCDFs. Eur J Pharmacol 270: 129-136. http://dx.doi.org/10.1016/0926-6917(94)90054-X

Larson, B; Kelts, JL; Banks, P; Cali, JJ. (2011). Automation and miniaturization of the Bioluminescent UGT-Glo assay for screening of UDP-glucuronosyltransferase inhibition by various compounds. The Association for Laboratory Automation Journal 16: 38-46. http://dx.doi.org/10.1016/j.jala.2010.05.005

Latif, R; Lau, Z; Cheung, P; Felsenfeld, DP; Davies, TF. (2016). The "TSH Receptor Glo Assay" - A High-Throughput Detection System for Thyroid Stimulation. Front Endocrinol (Lausanne) 7: 9. http://dx.doi.org/10.3389/fendo.2016.00003

Laudet, V. (2011). The Origins and Evolution of Vertebrate Metamorphosis [Review]. Curr Biol 21: R726-R737. http://dx.doi.org/10.1016/j.cub.2011.07.030

Lebaron, MJ; Coady, KK; O'Connor, JC; Nabb, DL; Markell, LK; Snajdr, S; Sue Marty, M. (2014). Key learnings from performance of the U.S. EPA Endocrine Disruptor Screening Program (EDSP) Tier 1 in vitro assays. Birth Defects Res B Dev Reprod Toxicol 101: 23-42. http://dx.doi.org/10.1002/bdrb.21094

Lecat-Guillet, N; Merer, G; Lopez, R; Pourcher, T; Rousseau, B; Ambroise, Y. (2008). Small-molecule inhibitors of sodium iodide symporter function. Chembiochem 9: 889-895. http://dx.doi.org/10.1002/cbic.200700682

Lema, SC; Dickey, JT; Schultz, IR; Swanson, P. (2008). Dietary exposure to 2,2',4,4'-tetrabromodiphenyl ether (PBDE-47) alters thyroid status and thyroid hormone-regulated gene transcription in the pituitary and brain. Environ Health Perspect 116: 1694-1699. http://dx.doi.org/10.1289/ehp.11570

Lynch, C; Sakamuru, S; Huang, R; Stavreva, D; Varticovski, L; Hager, G; Judson, R; Houck, K; Kleinstreuer, NC; Casey, WM; Paules, RS; Simeonov, A; Xia, M. (2016). Identifying environmental chemicals as agonists of the androgen receptor by applying a quantitative high-throughput screening platform. Toxicology 385: 48-58. http://dx.doi.org/10.1016/j.tox.2017.05.001

Maglich, JM; Kuhn, M; Chapin, RE; Pletcher, MT. (2014). More than just hormones: H295R cells as predictors of reproductive toxicity. Reprod Toxicol 45: 77-86. http://dx.doi.org/10.1016/j.reprotox.2013.12.009

Mani, AR; Ippolito, S; Moreno, JC; Visser, TJ; Moore, KP. (2007). The metabolism and dechlorination of chlorotyrosine in vivo. J Biol Chem 282: 29114-29121. http://dx.doi.org/10.1074/jbc.M704270200

Mansouri, K; Abdelaziz, A; Rybacka, A; Roncaglioni, A; Tropsha, A; Varnek, A; Zakharov, A; Worth, A; Richard, AM; Grulke, CM; Trisciuzzi, D; Fourches, D; Horvath, D; Benfenati, E; Muratov, E; Wedebye, EB; Grisoni, F; Mangiatordi, GF; Incisivo, GM; Hong, H; Ng, HW; Tetko, IV; Balabin, I; Kancherla, J; Shen, J; Burton, J; Nicklaus, M; Cassotti, M; Nikolov, NG; Nicolotti, O; Andersson, PL; Zang, Q; Politi, R; Beger, RD; Todeschini, R; Huang, R; Farag, S; Rosenberg, SA; Slavov, S; Hu, X; Judson, RS. (2016). CERAPP: Collaborative Estrogen Receptor Activity Prediction Project. Environ Health Perspect 124: 1023-1033. http://dx.doi.org/10.1289/ehp.1510267

Marchesini, GR; Meimaridou, A; Haasnoot, W; Meulenberg, E; Albertus, F; Mizuguchi, M; Takeuchi, M; Irth, H; Murk, AJ. (2008). Biosensor discovery of thyroxine transport

disrupting chemicals. Toxicol Appl Pharmacol 232: 150-160. http://dx.doi.org/10.1016/j.taap.2008.06.014
Marchesini, GR; Meulenberg, E; Haasnoot, W; Mizuguchi, M; Irth, H. (2006). Biosensor recognition of thyroid-disrupting chemicals using transport proteins. Anal Chem 78: 1107-1114. http://dx.doi.org/10.1021/ac051399i
Mardia, KV; Kent, JT; Bibby, JM. (1979). Multivariate analysis. New York, NY: Academic Press.
Martin, MT; Dix, DJ; Judson, RS; Kavlock, RJ; Reif, DM; Richard, AM; Rotroff, DM; Romanov, S; Medvedev, A; Poltoratskaya, N; Gambarian, M; Moeser, M; Makarov, SS; Houck, KA. (2010). Impact of environmental chemicals on key transcription regulators and correlation to toxicity end points within EPA's ToxCast program. Chem Res Toxicol 23: 578-590. http://dx.doi.org/10.1021/tx900325g
Massart, C; Hoste, C; Virion, A; Ruf, J; Dumont, JE; Van Sande, J. (2011). Cell biology of H2O2 generation in the thyroid: investigation of the control of dual oxidases (DUOX) activity in intact ex vivo thyroid tissue and cell lines. Mol Cell Endocrinol 343: 32-44. http://dx.doi.org/10.1016/j.mce.2011.05.047
Mcclain, RM; Posch, RC; Bosakowski, T; Armstrong, JM. (1988). Studies on the mode of action for thyroid gland tumor promotion in rats by phenobarbital. Toxicol Appl Pharmacol 94: 254-265.
Mclanahan, ED; White, P; Flowers, L; Schlosser, PM. (2014). The use of PBPK models to inform human health risk assessment: Case study on perchlorate and radioiodide human lifestage models. Risk Anal 34: 356-366. http://dx.doi.org/10.1111/risa.12101
Mctamney, PM; Rokita, SE. (2009). A mammalian reductive deiodinase has broad power to dehalogenate chlorinated and brominated substrates. J Am Chem Soc 131: 14212-14213. http://dx.doi.org/10.1021/ja906642n
Miller, MD; Crofton, KM; Rice, DC; Zoeller, RT. (2009). Thyroid-disrupting chemicals: interpreting upstream biomarkers of adverse outcomes [Review]. Environ Health Perspect 117: 1033-1041. http://dx.doi.org/10.1289/ehp.0800247
Mol, KA; Van Der Geyten, S; Kuhn, ER; Darras, VM. (1999). Effects of experimental hypo- and hyperthyroidism on iodothyronine deiodinases in Nile tilapia, Oreochromis niloticus. Fish Physiol Biochem 20: 201-207.
Montaño, M; Cocco, E; Guignard, C; Marsh, G; Hoffmann, L; Bergman, A; Gutleb, AC; Murk, AJ. (2012). New approaches to assess the transthyretin binding capacity of bioactivated thyroid hormone disruptors. Toxicol Sci 130: 94-105. http://dx.doi.org/10.1093/toxsci/kfs228
Moreno, JC; Bikker, H; Kempers, MJE; Van Trotsenburg, P; Baas, F; De Vijlder, JJM; Vulsma, T; Ris-Stalpers, C. (2002). Inactivating mutations in the gene for thyroid oxidase 2 (THOX2) and congenital hypothyroidism. N Engl J Med 347: 95-102. http://dx.doi.org/10.1056/NEJMoa012752
Moreno, JC; Klootwijk, W; Van Toor, H; Pinto, G; D'Alessandro, M; Leger, A; Goudie, D; Polak, M; Gruters, A; Visser, TJ. (2008). Mutations in the iodotyrosine deiodinase gene and hypothyroidism. N Engl J Med 358: 1811-1818. http://dx.doi.org/10.1056/NEJMoa0706819
Morse, DC; Groen, D; Veerman, M; van Amerongen, CJ; Koëter, HB; Smits van Prooije, AE; Visser, TJ; Koeman, JH; Brouwer, A. (1993). Interference of polychlorinated biphenyls

in hepatic and brain thyroid hormone metabolism in fetal and neonatal rats. Toxicol Appl Pharmacol 122: 27-33. http://dx.doi.org/10.1006/taap.1993.1168

Murk, AJ; Rijntjes, E; Blaauboer, BJ; Clewell, R; Crofton, KM; Dingemans, MM; Furlow, JD; Kavlock, R; Köhrle, J; Opitz, R; Traas, T; Visser, TJ; Xia, M; Gutleb, AC. (2013). Mechanism-based testing strategy using in vitro approaches for identification of thyroid hormone disrupting chemicals [Review]. Toxicol In Vitro 27: 1320-1346. http://dx.doi.org/10.1016/j.tiv.2013.02.012

Nakamura, T; Imada, T. (2005). Multiple comparison procedure of Dunnetts type for multivariate normal means. JJSCS 18: 21-32. http://dx.doi.org/http://doi.org/10.5183/jjscs1988.18.21

National Academies of Sciences, E; Studies, DoEaL; Toxicology, BoESa; Evaluations, CoIsCSiR-B. (2017). Using 21st Century Science to Improve Risk-Related Evaluations. In Using 21st Century Science to Improve Risk-Related Evaluations. Washington (DC): National Academies Press (US). http://dx.doi.org/10.17226/24635

Neumann, S; Huang, WW; Titus, S; Krause, G; Kleinau, G; Alberobello, AT; Zheng, W; Southall, NT; Inglese, J; Austin, CP; Celi, FS; Gavrilova, O; Thomas, CJ; Raaka, BM; Gershengorn, MC. (2009). Small-molecule agonists for the thyrotropin receptor stimulate thyroid function in human thyrocytes and mice. Proc Natl Acad Sci USA 106: 12471-12476. http://dx.doi.org/10.1073/pnas.0904506106

Nielsen, FK; Hansen, CH; Fey, JA; Hansen, M; Jacobsen, NW; Halling-Sørensen, B; Björklund, E; Styrishave, B. (2012). H295R cells as a model for steroidogenic disruption: a broader perspective using simultaneous chemical analysis of 7 key steroid hormones. Toxicol In Vitro 26: 343-350. http://dx.doi.org/10.1016/j.tiv.2011.12.008

Nilsson, SF; Peterson, PA. (1975). Studies on thyroid hormone-binding proteins. I. The subunit structure of human thyroxine-binding globulin and its interaction with ligands. J Biol Chem 250: 8543-8553.

Nilsson, SF; Rask, L; Peterson, PA. (1975). Studies on thyroid hormone-binding proteins. II. Binding of thyroid hormones, retinol-binding protein, and fluorescent probes to prealbumin and effects of thyroxine on prealbumin subunit self association. J Biol Chem 250: 8554-8563.

Noyes, PD; Hinton, DE; Stapleton, HM. (2011). Accumulation and debromination of decabromodiphenyl ether (BDE-209) in juvenile fathead minnows (Pimephales promelas) induces thyroid disruption and liver alterations. Toxicol Sci 122: 265-274. http://dx.doi.org/10.1093/toxsci/kfr105

Noyes, PD; Lema, SC; Macaulay, LJ; Douglas, NK; Stapleton, HM. (2013). Low Level Exposure to the Flame Retardant BDE-209 Reduces Thyroid Hormone Levels and Disrupts Thyroid Signaling in Fathead Minnows. Environ Sci Technol 47: 10012-10021. http://dx.doi.org/10.1021/es402650x

Noyes, PD; Paul-Friedman, KP; Haselman, JT; Barone Jr., S; Crofton, KM; Gilbert, ME; Hornung, MW; Laws, SC; Simmons, SO; Stoker, TE; Tietge, JE; Degitz, SJ. Framework for screening chemicals for thyroid bioactivity using in vitro high-throughput data and adverse outcome pathways.

OECD (Organisation for Economic Co-operation and Development). (2009a). Test No. 231: Amphibian metamorphosis assay. In OECD Guidelines for the Testing of Chemicals, Section 2: Effects on Biotic Systems. Paris. http://dx.doi.org/10.1787/9789264076242-en

OECD (Organisation for Economic Co-operation and Development). (2009b). Test No. 440: Uterotrophic bioassay in rodents: A short-term screening test for oestrogenic properties (Korean version). 1. http://dx.doi.org/http://dx.doi.org/10.1787/9789264067417-en
OECD (Organisation for Economic Co-operation and Development). (2009c). Test no. 441: Hershberger bioassay in rats: A short-term screening assay for (anti)androgenic properties. In OECD guidelines for the testing of chemicals, Section 4: Health effects. Paris, France. http://dx.doi.org/10.1787/9789264076334-en
OECD (Organisation for Economic Co-operation and Development). (2009d). Test no. 455: the stably transfected human estrogen receptor-alpha transcriptional activation assay for detection of estrogenic agonist-activity of chemicals. Paris: OECD Publishing. http://dx.doi.org/http://dx.doi.org/10.1787/9789264076372-en
OECD (Organisation for Economic Co-operation and Development). (2010a). Draft report of pre-validation and inter-laboratory validation for androgen receptor (AR) mediated stably transfected transcriptional activation (AR-STTA) assay to detect androgenic and anti-androgenic activities. Chemicals Evaluation and Research Institute, Japan. http://www.oecd.org/chemicalsafety/testing/46755800.pdf
OECD (Organisation for Economic Co-operation and Development). (2010b). Multi-Laboratory Validation Report of the H295r Steroidogenesis Assay to Identify Modulators of Testosterone and Estradiol Production. (ENV/JM/MONO(2010)31). Paris, France: OECD Publishing. http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2010)31&doclanguage=en
OECD (Organisation for Economic Co-operation and Development). (2011). Test no. 456: h295r steroidogenesis assay. Paris: OECD Publishing. http://dx.doi.org/http://dx.doi.org/10.1787/9789264122642-en
OECD (Organisation for Economic Co-operation and Development). (2012a). Test no. 229: Fish short term reproduction assay. (229). Paris, France: OECD Publishing. http://dx.doi.org/10.1787/9789264185265-en
OECD (Organisation for Economic Co-operation and Development). (2012b). Test no. 443: extended one-generation reproductive toxicity study (rat). Paris: OECD Publishing. http://dx.doi.org/http://dx.doi.org/10.1787/9789264122550-en
OECD (Organisation for Economic Co-operation and Development). (2014). New scoping document on in vitro and ex vivo assays for the identification of modulators of thyroid hormone signaling. (ENV/JM/MONO(2014)23). Paris, France: OECD Environment, Health and Safety Publications. http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2014)23&doclanguage=en
OECD (Organisation for Economic Co-operation and Development). (2016). Collaborative adverse outcome pathway wiki (AOP-wiki). Available online at https://aopwiki.org/
OECD (Organisation for Economic Co-operation and Development). (2017). Adverse outcome pathways, molecular screening and toxicogenomics. Available online at http://www.oecd.org/chemicalsafety/testing/adverse-outcome-pathways-molecular-screening-and-toxicogenomics.htm (accessed July 10, 2017).
Oskarsson, A; Ulleras, E; Ohlsson Andersson, A. (2016). Acetaminophen Increases Aldosterone Secretion While Suppressing Cortisol and Androgens: A Possible Link to Increased Risk of Hypertension. Am J Hypertens 29: 1158-1164. http://dx.doi.org/10.1093/ajh/hpw055

Palha, JA; Hays, MT; Deescobar, GM; Episkopou, V; Gottesman, ME; Saraiva, MJM. (1997). Transthyretin is not essential for thyroxine to reach the brain and other tissues in transthyretin-null mice. Am J Physiol Endocrinol Metab 272: E485-E493.

Paul-Friedman, K; Zhao, J; Huang, R; Xia, M; Crofton, K; Houck, K. (2017). Screening the Tox21 10K library for thyroid stimulating hormone receptor agonist and antagonist activity (Abstract #3423). SOT Annual Meeting, March 12-16, 2017, Baltimore, Maryland.

Paul Friedman, K; Papineni, S; Marty, MS; Yi, KD; Goetz, AK; Rasoulpour, RJ; Kwiatkowski, P; Wolf, DC; Blacker, AM; Peffer, RC. (2016). A predictive data-driven framework for endocrine prioritization: A triazole fungicide case study. Crit Rev Toxicol 46: 1-49. http://dx.doi.org/10.1080/10408444.2016.1193722

Paul, KB; Hedge, JM; Rotroff, DM; Hornung, MW; Crofton, KM; Simmons, SO. (2014). Development of a thyroperoxidase inhibition assay for high-throughput screening. Chem Res Toxicol 27: 387-399. http://dx.doi.org/10.1021/tx400310w

Pavelka, S. (2014). Development of radiometric assays for quantification of enzyme activities of the key enzymes of thyroid hormones metabolism. Physiol Res 63: S133-S140.

Perkins, EJ; Ankley, GT; Crofton, KM; Garcia-Reyero, N; Lalone, CA; Johnson, MS; Tietge, JE; Villeneuve, DL. (2013). Current Perspectives on the Use of Alternative Species in Human Health and Ecological Hazard Assessments [Review]. Environ Health Perspect 121: 1002-1010. http://dx.doi.org/10.1289/ehp.1306638

Pinto, CL; Mansouri, K; Judson, R; Browne, P. (2016). Prediction of Estrogenic Bioactivity of Environmental Chemical Metabolites. Chem Res Toxicol 29: 1410-1427. http://dx.doi.org/10.1021/acs.chemrestox.6b00079

R Core Team (R Development Core Team). (2015). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/

Raucy, JL; Lasker, JM. (2010). Current in vitro high throughput screening approaches to assess nuclear receptor activation. Curr Drug Metab 11: 806-814.

Renko, K; Hoefig, CS; Dupuy, C; Harder, L; Schwiebert, C; Kohrle, J; Schomburg, L. (2016). A nonradioactive DEHAL assay for testing substrates, inhibitors, and monitoring endogenous activity. Endocrinology 157: 4516-4525. http://dx.doi.org/10.1210/en.2016-1549

Renko, K; Hoefig, CS; Hiller, F; Schomburg, L; Köhrle, J. (2012). Identification of iopanoic acid as substrate of type 1 deiodinase by a novel nonradioactive iodide-release assay. Endocrinology 153: 2506-2513. http://dx.doi.org/10.1210/en.2011-1863

Renko, K; Schache, S; Hoefig, CS; Welsink, T; Schwiebert, C; Braun, D; Becker, NP; Kohrle, J; Schomburg, L. (2015). An improved nonradioactive screening method identifies genistein and xanthohumol as potent inhibitors of iodothyronine deiodinases. Thyroid 25: 962-968. http://dx.doi.org/10.1089/thy.2015.0058

Rhoden, KJ; Cianchetta, S; Duchi, S; Romeo, G. (2008). Fluorescence quantitation of thyrocyte iodide accumulation with the yellow fluorescent protein variant YFP-H148Q/I152L. Anal Biochem 373: 239-246. http://dx.doi.org/10.1016/j.ab.2007.10.020

Richard, AM; Judson, RS; Houck, KA; Grulke, CM; Volarath, P; Thillainadarajah, I; Yang, C; Rathman, J; Martin, MT; Wambaugh, JF; Knudsen, TB; Kancherla, J; Mansouri, K; Patlewicz, G; Williams, AJ; Little, SB; Crofton, KM; Thomas, RS. (2016). ToxCast

chemical landscape: Paving the road to 21st century toxicology. Chem Res Toxicol 29: 1225-1251. http://dx.doi.org/10.1021/acs.chemrestox.6b00135
Rijk, JC; Peijnenburg, AA; Blokland, MH; Lommen, A; Hoogenboom, RL; Bovee, TF. (2012). Screening for modulatory effects on steroidogenesis using the human H295R adrenocortical cell line: a metabolomics approach. Chem Res Toxicol 25: 1720-1731. http://dx.doi.org/10.1021/tx3001779
Robbins, J. (2000). Thyroid hormone transport proteins and the physiology of hormone binding. In LE Braverman; D Utiger (Eds.), Werner & Ingbar's the thyroid : A fundamental and clinical text (8th ed., pp. 105-120). Philadelphia, PA: Lippincott Williams and Wilkins.
Robitaille, CN; Rivest, P; Sanderson, JT. (2015). Antiandrogenic mechanisms of pesticides in human LNCaP prostate and H295R adrenocortical carcinoma cells. Toxicol Sci 143: 126-135. http://dx.doi.org/10.1093/toxsci/kfu212
Rotroff, DM; Beam, AL; Dix, DJ; Farmer, A; Freeman, KM; Houck, KA; Judson, RS; Lecluyse, EL; Martin, MT; Reif, DM; Ferguson, SS. (2010a). Xenobiotic-metabolizing enzyme and transporter gene expression in primary cultures of human hepatocytes modulated by ToxCast chemicals. J Toxicol Environ Health B Crit Rev 13: 329-346. http://dx.doi.org/10.1080/10937404.2010.483949
Rotroff, DM; Wetmore, BA; Dix, DJ; Ferguson, SS; Clewell, HJ; Houck, KA; Lecluyse, EL; Andersen, ME; Judson, RS; Smith, CM; Sochaski, MA; Kavlock, RJ; Boellmann, F; Martin, MT; Reif, DM; Wambaugh, JF; Thomas, RS. (2010b). Incorporating human dosimetry and exposure into high-throughput in vitro toxicity screening. Toxicol Sci 117: 348-358. http://dx.doi.org/10.1093/toxsci/kfq220
Saito, R; Terasaki, N; Yamazaki, M; Masutomi, N; Tsutsui, N; Okamoto, M. (2016). Estimation of the mechanism of adrenal action of endocrine-disrupting compounds using a computational model of adrenal steroidogenesis in NCI-H295R cells. J Toxicol 2016: 1-19. http://dx.doi.org/10.1155/2016/4041827
Sanderson, JT; Seinen, W; Giesy, JP; van den Berg, M. (2000). 2-Chloro-s-triazine herbicides induce aromatase (CYP19) activity in H295R human adrenocortical carcinoma cells: A novel mechanism for estrogenicity? Toxicol Sci 54: 121-127. http://dx.doi.org/10.1093/toxsci/54.1.121
Santini, F; Vitti, P; Ceccarini, G; Mammoli, C; Rosellini, V; Pelosini, C; Marsili, A; Tonacchera, M; Agretti, P; Santoni, T; Chiovato, L; Pinchera, A. (2003). In vitro assay of thyroid disruptors affecting TSH-stimulated adenylate cyclase activity. J Endocrinol Invest 26: 950-955. http://dx.doi.org/10.1007/BF03348190
SAS. (2012). What is Mahalanobis distance? Available online
Scanlan, TS. (2009). Minireview: 3-iodothyronamine (T(1)AM): A new player on the thyroid endocrine team? Endocrinology 150: 1108-1111. http://dx.doi.org/10.1210/en.2008-1596
Schmutzler, C; Bacinski, A; Gotthardt, I; Huhne, K; Ambrugger, P; Klammer, H; Schlecht, C; Hoang-Vu, C; Grüters, A; Wuttke, W; Jarry, H; Köhrle, J. (2007). The ultraviolet filter benzophenone 2 interferes with the thyroid hormone axis in rats and is a potent in vitro inhibitor of human recombinant thyroid peroxidase. Endocrinology 148: 2835-2844. http://dx.doi.org/10.1210/en.2006-1280
Shimizu, R; Yamaguchi, M; Uramaru, N; Kuroki, H; Ohta, S; Kitamura, S; Sugihara, K. (2013). Structure-activity relationships of 44 halogenated compounds for iodotyrosine deiodinase-inhibitory activity. Toxicology 314: 22-29. http://dx.doi.org/10.1016/j.tox.2013.08.017

Shimizu, R, yo. (2014). Iodotyrosine Deiodinase, a Novel Target of Environmental Halogenated Chemicals for Disruption of the Thyroid Hormone System in Mammals. Biol Pharm Bull 37: 1430-1434.

Sipes, NS; Martin, MT; Kothiya, P; Reif, DM; Judson, RS; Richard, A; Houck, KA; Dix, DJ; Kavlock, RJ; Knudsen, TB. (2013). Profiling 976 ToxCast Chemicals across 331 Enzymatic and Receptor Signaling Assays. Chem Res Toxicol 26: 878-895. http://dx.doi.org/10.1021/tx400021

Steinsapir, J; Harney, J; Larsen, PR. (1998). Type 2 iodothyronine deiodinase in rat pituitary tumor cells is inactivated in proteasomes. J Clin Invest 102: 1895-1899. http://dx.doi.org/10.1172/jci4672

Strajhar, P; Tonoli, D; Jeanneret, F; Imhof, RM; Malagnino, V; Patt, M; Kratschmar, DV; Boccard, J; Rudaz, S; Odermatt, A. (2017). Steroid profiling in H295R cells to identify chemicals potentially disrupting the production of adrenal steroids. Toxicology 381: 51-63. http://dx.doi.org/10.1016/j.tox.2017.02.010

Streckfuss, F; Hamann, I; Schomburg, L; Michaelis, M; Sapin, R; Klein, MO; Kohrle, J; Schweizer, U. (2005). Hepatic deiodinase activity is dispensable for the maintenance of normal circulating thyroid hormone levels in mice. Biochem Biophys Res Commun 337: 739-745. http://dx.doi.org/10.1016/j.bbrc.2005.09.102

Taxvig, C; Vinggaard, AM; Hass, U; Axelstad, M; Boberg, J; Hansen, PR; Frederiksen, H; Nellemann, C. (2008). Do parabens have the ability to interfere with steroidogenesis? Toxicol Sci 106: 206-213. http://dx.doi.org/10.1093/toxsci/kfn148

Taylor, JP; Metcalfe, RA; Watson, PF; Weetman, AP; Trembath, RC. (2002). Mutations of the PDS gene, encoding pendrin, are associated with protein mislocalization and loss of iodide efflux: Implications for thyroid dysfunction in pendred syndrome. J Clin Endocrinol Metab 87: 1778-1784. http://dx.doi.org/10.1210/jc.87.4.1778

Teeguarden, JG; Tan, YM; Edwards, SW; Leonard, JA; Anderson, KA; Corley, RA; Kile, ML; Simonich, SM; Stone, D; Tanguay, RL; Waters, KM; Harper, SL; Williams, DE. (2016). Completing the Link between Exposure Science and Toxicology for Improved Environmental Health Decision Making: The Aggregate Exposure Pathway Framework. Environ Sci Technol 50: 4579-4586. http://dx.doi.org/10.1021/acs.est.5b05311

Titus, S; Neumann, S; Zheng, W; Southall, N; Michael, S; Klumpp, C; Yasgar, A; Shinn, P; Thomas, CJ; Inglese, J; Gershengorn, MC; Austin, CP. (2008). Quantitative high-throughput screening using a live-cell cAMP assay identifies small-molecule Agonists of the TSH receptor. J Biomol Screen 13: 120-127. http://dx.doi.org/10.1177/1087057107313786

Tonoli, D; Furstenberger, C; Boccard, J; Hochstrasser, D; Jeanneret, F; Odermatt, A; Rudaz, S. (2015). Steroidomic Footprinting Based on Ultra-High Performance Liquid Chromatography Coupled with Qualitative and Quantitative High-Resolution Mass Spectrometry for the Evaluation of Endocrine Disrupting Chemicals in H295R Cells. Chem Res Toxicol 28: 955-966. http://dx.doi.org/10.1021/tx5005369

U.S. EPA (U.S. Environmental Protection Agency). (1998a). Endocrine disruptor screening and testing advisory committee final report [EPA Report]. Washington, DC. https://www.epa.gov/sites/production/files/2015-08/documents/exesum14.pdf

U.S. EPA (U.S. Environmental Protection Agency). (1998b). Health effects test guidelines: 870.3800: Reproduction and fertility effects (Rat) [EPA Report]. (EPA 712C98208).

Washington, DC. https://www.regulations.gov/document?D=EPA-HQ-OPPT-2009-0156-0018
U.S. EPA (U.S. Environmental Protection Agency). (2007). Draft list of initial pesticide active ingredients and pesticide inerts to be considered for screening under the Federal Food, Drug, and Cosmetics Act. Fed Reg 72: 33486-33503.
U.S. EPA (U.S. Environmental Protection Agency). (2008). FIFRA Scientific Advisory Panel; Notice of public meeting. Fed Reg 73: 4216-4218.
U.S. EPA (U.S. Environmental Protection Agency). (2009a). Endocrine disruptor screening program test guidelines oppts 890.1550: steroidogenesis (human cell line h295r) [EPA Report]. (EPA 640-C-09-003). Docket # EPA-HQ-OPPT-2009-05676-0011.pdf: U.S. EPA, Office of Chemical Safety and Pollution Prevention. https://www.regulations.gov/document?D=EPA-HQ-OPPT-2009-0576-0011
U.S. EPA (U.S. Environmental Protection Agency). (2009b). Endocrine Disruptor Screening Program test guidelines: OPPTS 890.1100: Amphibian metamorphosis (frog) [EPA Report]. (EPA 640-C-09-003). Washington, D.C.: U.S. EPA; Office of Chemical Safety and Pollution Prevention. https://www.regulations.gov/document?D=EPA-HQ-OPPT-2009-0576-0002
U.S. EPA (U.S. Environmental Protection Agency). (2009c). Endocrine Disruptor Screening Program test guidelines: OPPTS 890.1150: Androgen receptor binding (rat prostate cytosol) [EPA Report]. (EPA 640-C-09-003). Washington, D.C.: U.S. EPA; Office of Chemical Safety and Pollution Prevention. https://www.regulations.gov/document?D=EPA-HQ-OPPT-2009-0576-0003
U.S. EPA (U.S. Environmental Protection Agency). (2009d). Endocrine Disruptor Screening Program test guidelines: OPPTS 890.1200: Aromatase (human recombinant) [EPA Report]. (EPA 740-C-09-004). Washington, D.C.: U.S. EPA, Office of Chemical Safety and Pollution Prevention. https://www.regulations.gov/document?D=EPA-HQ-OPPT-2009-0576-0004
U.S. EPA (U.S. Environmental Protection Agency). (2009e). Endocrine Disruptor Screening Program test guidelines: OPPTS 890.1250: Estrogen receptor binding assay using rat uterine cytosol (ER-RUC) [EPA Report]. (EPA 740-C-09-005). Washington, D.C.: U.S. EPA; Office of Chemical Safety and Pollution Prevention. https://www.regulations.gov/document?D=EPA-HQ-OPPT-2009-0576-0005
U.S. EPA (U.S. Environmental Protection Agency). (2009f). Endocrine Disruptor Screening Program test guidelines: OPPTS 890.1300: Estrogen receptor transcriptional activation (human cell line (HeLa-9903)) [EPA Report]. (EPA 740-C-09-006). Washington, D.C.: U.S. EPA, Office of Chemical Safety and Pollution Prevention. https://www.regulations.gov/document?D=EPA-HQ-OPPT-2009-0576-0006
U.S. EPA (U.S. Environmental Protection Agency). (2009g). Endocrine Disruptor Screening Program test guidelines: OPPTS 890.1350: Fish short-term reproduction assay [EPA Report]. (EPA 740-C-09-007). Washington, D.C.: U.S. EPA, Office of Chemical Safety and Pollution Prevention. https://www.regulations.gov/document?D=EPA-HQ-OPPT-2009-0576-0007
U.S. EPA (U.S. Environmental Protection Agency). (2009h). Endocrine Disruptor Screening Program test guidelines: OPPTS 890.1400 Hershberger bioassay [EPA Report]. (OCSPP 890.1400). Washington, DC. https://www.regulations.gov/document?D=EPA-HQ-OPPT-2009-0576-0008

U.S. EPA (U.S. Environmental Protection Agency). (2009i). Endocrine disruptor screening program test guidelines: OPPTS 890.1450. Pubertal development and thyroid function in intact juvenile/peripubertal female rats [EPA Report]. (EPA 740-C-09-009). Washington, DC. https://www.regulations.gov/document?D=EPA-HQ-OPPT-2009-0576-0009

U.S. EPA (U.S. Environmental Protection Agency). (2009j). Endocrine disruptor screening program test guidelines: OPPTS 890.1500. Pubertal development and thyroid function in intact juvenile/peripubertal male rats [EPA Report]. (EPA 740-C-09-012). Washington, DC. https://www.regulations.gov/document?D=EPA-HQ-OPPT-2009-0576-0010

U.S. EPA (U.S. Environmental Protection Agency). (2009k). Final list of initial pesticide active ingredients and pesticide inert ingredients to be screened under the Federal Food, Drug, and Cosmetic Act [EPA Report] (pp. 17579-17585). (74 FR 17579). https://www.federalregister.gov/documents/2009/04/15/E9-8709/final-list-of-initial-pesticide-active-ingredients-and-pesticide-inert-ingredients-to-be-screened

U.S. EPA (U.S. Environmental Protection Agency). (2011a). Hershberger assay, OCSPP guideline 890.1400: Standard evaluation procedure (SEP) [EPA Report]. Washington, DC. https://www.epa.gov/sites/production/files/2015-07/documents/final_890.1600_hershberger_assay_sep_10.6.11.pdf

U.S. EPA (U.S. Environmental Protection Agency). (2011b). Weight-of-Evidence: Evaluating Results of EDSP Tier 1 Screening to Identify the Need for Tier 2 Testing. http://dx.doi.org/EPA-HQ-OPPT-2010-0877-0021

U.S. EPA (U.S. Environmental Protection Agency). (2012). Endocrine disruptor screening program universe of chemicals and general validation principles [EPA Report]. https://www.epa.gov/endocrine-disruption/endocrine-disruptor-screening-program-edsp-universe-chemicals-and-general

U.S. EPA (U.S. Environmental Protection Agency). (2013a). Endocrine Disruptor Screening Program: Final second list of chemicals and substances for tier 1 screening. Fed Reg 78: 35922-35928.

U.S. EPA (U.S. Environmental Protection Agency). (2013b). EPAs State of the Science Evaluation: Nonmonotonic Dose Responses as They Apply to Estrogen, Androgen, and Thyroid Pathways and EPA Testing and Assessment Procedures (Draft) [EPA Report]. https://www.epa.gov/chemical-research/nonmonotonic-dose-responses-they-apply-estrogen-androgen-and-thyroid-pathways-and

U.S. EPA (U.S. Environmental Protection Agency). (2013c). January 2013 FIFRA SAP Meeting - Prioritizing the Universe of Endocrine Disruptor Screening Program (EDSP) Chemicals using Computational Toxicology Tools. https://www.regulations.gov/docket?D=EPA-HQ-OPP-2012-0818

U.S. EPA (U.S. Environmental Protection Agency). (2013d). June 2013 FIFRA SAP Meeting - Endocrine Disruptor Screening Program (EDSP) Tier 2 Ecotoxicity Tests [EPA Report]. (EPA-HQ-OPP-2013-0182). https://www.regulations.gov/docket?D=EPA-HQ-OPP-2013-0182

U.S. EPA (U.S. Environmental Protection Agency). (2014a). December 2014 FIFRA SAP Meeting - Endocrine Activity and Exposure-based Prioritization and Screening [EPA Report]. https://www.regulations.gov/docket?D=EPA-HQ-OPP-2014-0614

U.S. EPA (U.S. Environmental Protection Agency). (2014b). Integrated bioactivity and exposure ranking: A computational approach for the prioritization and screening of chemicals in

the endocrine disruptor screening program [EPA Report]. (EPA-HQ-OPP-2014-0614-0003). https://www.regulations.gov/document?D=EPA-HQ-OPP-2014-0614-0003

U.S. EPA (U.S. Environmental Protection Agency). (2014c). July/August 2014 FIFRA SAP Meeting - Scientific Issues Associated with New High Throughput Methods to Estimate Chemical Exposure [EPA Report]. https://www.regulations.gov/docket?D=EPA-HQ-OPP-2014-0331

U.S. EPA (U.S. Environmental Protection Agency). (2015a). Endocrine Disruptor Screening Program test guidelines: OPPTS 890.2100: Avian two-generation toxicity test in the Japanese quail [EPA Report]. (EPA 740-C-15-003). Washington, D.C.: U.S. EPA, Office of Chemical Safety and Pollution Prevention. https://www.regulations.gov/document?D=EPA-HQ-OPPT-2009-0576-0017

U.S. EPA (U.S. Environmental Protection Agency). (2015b). Endocrine Disruptor Screening Program test guidelines: OPPTS 890.2200: Medaka extended one generation reproduction test (MEOGRT) [EPA Report]. (EPA 740-C-15-002). Washington, D.C.: U.S. EPA, Office of Chemical Safety and Pollution Prevention. https://www.regulations.gov/document?D=EPA-HQ-OPPT-2009-0576-0019

U.S. EPA (U.S. Environmental Protection Agency). (2015c). Endocrine Disruptor Screening Program test guidelines: OPPTS 890.2300: Larval amphibian growth and development assay (LAGDA) [EPA Report]. (EPA 740-C-15-001). Washington, D.C.: U.S. EPA, Office of Chemical Safety and Pollution Prevention. https://www.regulations.gov/document?D=EPA-HQ-OPPT-2009-0576-0018

U.S. EPA (U.S. Environmental Protection Agency). (2015d). Endocrine disruptor screening program tier 1 screening determinations and associated data evaluation records. Available online at https://www.epa.gov/endocrine-disruption/endocrine-disruptor-screening-program-tier-1-screening-determinations-and (accessed July 20, 2017).

U.S. EPA (U.S. Environmental Protection Agency). (2015e). ToxCast and Tox21 Data from invitrodb_v2. Washington, DC. Retrieved from http://www2.epa.gov/chemical-research/toxicity-forecaster-toxcasttm-data

U.S. EPA (U.S. Environmental Protection Agency). (2015f). 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) [EPA Report] (pp. 35350-35355). (EPA-HQ-OPPT-2015-0305). Washington, D.C.: Federal Register. https://www.federalregister.gov/articles/2015/06/19/2015-15182/use-of-high-throughput-assays-and-computational-tools-endocrine-disruptor-screening-program-notice

U.S. EPA (U.S. Environmental Protection Agency). (2015g). Use of high throughput assays and computational tools; endocrine disruptor screening program; notice of availability and opportunity for comment. Fed Reg 80: 35350-35355.

U.S. EPA (U.S. Environmental Protection Agency). (2016a). Endocrine disruptor screening program tier 1 screening determinations and associated data evaluation records. Available online at https://www.epa.gov/endocrine-disruption/endocrine-disruptor-screening-program-tier-1-screening-determinations-and (accessed July 18, 2017).

U.S. EPA (U.S. Environmental Protection Agency). (2016b). iCSS ToxCast Dashboard. Available online at https://actor.epa.gov/dashboard/

U.S. EPA (U.S. Environmental Protection Agency). (2017a). EDSP Universe on the Chemistry Dashboard. Available online at http://comptox.epa.gov/dashboard/chemical_lists/edspu

U.S. EPA (U.S. Environmental Protection Agency). (2017b). Endocrine disruptor screening program (EDSP) overview. Available online at https://www.epa.gov/endocrine-disruption/endocrine-disruptor-screening-program-edsp-overview (accessed August 1, 2017).
Verhaeghe, E; Buisson, D; Zekri, E; Leblanc, C; Potin, P; Ambroise, Y. (2008). A colorimetric assay for steady-state analyses of iodo- and bromoperoxidase activities. Anal Biochem 379: 60-65. http://dx.doi.org/10.1016/j.ab.2008.04.041
Vickers, AE; Heale, J; Sinclair, JR; Morris, S; Rowe, JM; Fisher, RL. (2012). Thyroid organotypic rat and human cultures used to investigate drug effects on thyroid function, hormone synthesis and release pathways. Toxicol Appl Pharmacol 260: 81-88. http://dx.doi.org/10.1016/j.taap.2012.01.029
Vigone, MC; Fugazzola, L; Zamproni, I; Passoni, A; Di Candia, S; Chiumello, G; Persani, L; Weber, G. (2005). Persistent mild hypothyroidism associated with novel sequence variants of the DUOX2 gene in two siblings. Hum Mutat 26: 395. http://dx.doi.org/10.1002/humu.9372
Villeneuve, DL; Crump, D; Garcia-Reyero, N; Hecker, M; Hutchinson, TH; Lalone, CA; Landesmann, B; Lettieri, T; Munn, S; Nepelska, M; Ottinger, MA; Vergauwen, L; Whelan, M. (2014). Adverse outcome pathway (AOP) development I: Strategies and principles. Toxicol Sci 142: 312-320. http://dx.doi.org/10.1093/toxsci/kfu199
Visser, TJ. (1996). Pathways of thyroid hormone metabolism [Review]. Acta Medica Austriaca 23: 10-16.
Visser, TJ; Kaptein, E; van Toor, H; van Raaij, JA; van den Berg, KJ; Joe, CT; van Engelen, JG; Brouwer, A. (1993). Glucuronidation of thyroid hormone in rat liver: effects of in vivo treatment with microsomal enzyme inducers and in vitro assay conditions. Endocrinology 133: 2177-2186. http://dx.doi.org/10.1210/endo.133.5.8404669
Visser, WE; Friesema, ECH; Visser, TJ. (2011). Minireview: Thyroid hormone transporters: The knowns and the unknowns. Mol Endocrinol 25: 1-14. http://dx.doi.org/10.1210/me.2010-0095
Waltz, F; Pillette, L; Ambroise, Y. (2010). A nonradioactive iodide uptake assay for sodium iodide symporter function. Anal Biochem 396: 91-95. http://dx.doi.org/10.1016/j.ab.2009.08.038
Wambaugh, JF; Wang, A; Dionisio, KL; Frame, A; Egeghy, P; Judson, R; Setzer, RW. (2014). High Throughput Heuristics for Prioritizing Human Exposure to Environmental Chemicals. Environ Sci Technol 48: 12760-12767. http://dx.doi.org/10.1021/es503583j
Wambaugh, JF; Wetmore, BA; Pearce, R; Strope, C; Goldsmith, R; Sluka, JP; Sedykh, A; Tropsha, A; Bosgra, S; Shah, I; Judson, R; Thomas, RS; Setzer, RW. (2015). Toxicokinetic Triage for Environmental Chemicals. Toxicol Sci 147: 55-67. http://dx.doi.org/10.1093/toxsci/kfv118
Wang, DL; Stapleton, HM. (2010). Analysis of thyroid hormones in serum by liquid chromatography-tandem mass spectrometry. Anal Bioanal Chem 397: 1831-1839. http://dx.doi.org/10.1007/s00216-010-3705-9
Wang, LQ; James, MO. (2006). Inhibition of sulfotransferases by xenobiotics [Review]. Curr Drug Metab 7: 83-104. http://dx.doi.org/10.2174/138920006774832596
Watt, E. (2016). eapath: Computational Models for Estrogen and Androgen Receptor Activity. Retrieved from https://github.com/ericwatt/eapath

Watt, E. (2017). toxboot: Bootstrap Methods for ToxCast High Throughput Screening Data. Retrieved from https://github.com/ericwatt/toxboot
Wegner, S; Browne, P; Dix, D. (2016). Identifying reference chemicals for thyroid bioactivity screening. Reprod Toxicol 65: 402-413. http://dx.doi.org/10.1016/j.reprotox.2016.08.016
Westholm, DE; Stenehjem, DD; Rumbley, JN; Drewes, LR; Anderson, GW. (2009). Competitive Inhibition of Organic Anion Transporting Polypeptide 1c1-Mediated Thyroxine Transport by the Fenamate Class of Nonsteroidal Antiinflammatory Drugs. Endocrinology 150: 1025-1032. http://dx.doi.org/10.1210/en.2008-0188
Wetmore, BA; Wambaugh, JF; Allen, B; Ferguson, SS; Sochaski, MA; Setzer, RW; Houck, KA; Strope, 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: 121-136. http://dx.doi.org/10.1093/toxsci/kfv171
Wilson, VS; Bobseine, K; Lambright, CR; Gray, LE. (2002). A novel cell line, MDA-kb2, that stably expresses an androgen- and glucocorticoid-responsive reporter for the detection of hormone receptor agonists and antagonists. Toxicol Sci 66: 69-81. http://dx.doi.org/10.1093/toxsci/66.1.69
Wittwehr, C; Aladjov, H; Ankley, G; Byrne, HJ; de Knecht, J; Heinzle, E; Klambauer, G; Landesmann, B; Luijten, M; Mackay, C; Maxwell, G; Meek, ME; Paini, A; Perkins, E; Sobanski, T; Villeneuve, D; Waters, KM; Whelan, M. (2017). How adverse outcome pathways can aid the development and use of computational prediction models for regulatory toxicology [Review]. Toxicol Sci 155: 326-336. http://dx.doi.org/10.1093/toxsci/kfw207
Wolff, J. (1998). Perchlorate and the thyroid gland [Review]. Pharmacol Rev 50: 89-105.
Wu, SY; Green, WL; Huang, WS; Hays, MT; Chopra, IJ. (2005). Alternate pathways of thyroid hormone metabolism. Thyroid 15: 943-958. http://dx.doi.org/10.1089/thy.2005.15.943
Xu, M; Iwasaki, T; Shimokawal, N; Sajdel-Sukowska, EM; Koibuchi, N. (2013). The effect of low dose lipopolysaccharide on thyroid hormone-regulated actin cytoskeleton modulation and type 2 iodothyronine deiodinase activity in astrocytes. Endocr J 60: 1221-1230.
Yanes, LL; Romero, DG. (2009). Dihydrotestosterone stimulates aldosterone secretion by H295R human adrenocortical cells. Mol Cell Endocrinol 303: 50-56. http://dx.doi.org/10.1016/j.mce.2008.12.020
You, SH; Gauger, KJ; Bansal, R; Zoeller, RT. (2006). 4-Hydroxy-PCB106 acts as a direct thyroid hormone receptor agonist in rat GH3 cells. Mol Cell Endocrinol 257-258: 26-34. http://dx.doi.org/10.1016/j.mce.2006.06.009
Yu, WG; Liu, W; Jin, YH. (2009). Effects of perfluorooctane sulfonate on rat thyroid hormone biosynthesis and metabolism. Environ Toxicol Chem 28: 990-996. http://dx.doi.org/10.1897/08-345.1
Zavacki, AM; Ying, H; Christoffolete, MA; Aerts, G; So, E; Harney, JW; Cheng, SY; Larsen, PR; Bianco, AC. (2005). Type 1 iodothyronine deiodinase is a sensitive marker of peripheral thyroid status in the mouse. Endocrinology 146: 1568-1575. http://dx.doi.org/10.1210/en.2004-1392
Zhang, JH; Chung, TDY; Oldenburg, KR. (1999). A simple statistical parameter for use in evaluation and validation of high throughput screening assays. J Biomol Screen 4: 67-73. http://dx.doi.org/Doi 10.1177/108705719900400206

Zhang, X; Chang, H; Wiseman, S; He, Y; Higley, E; Jones, P; Wong, CK; Al-Khedhairy, A; Giesy, JP; Hecker, M. (2011). Bisphenol A disrupts steroidogenesis in human H295R cells. Toxicol Sci 121: 320-327. http://dx.doi.org/10.1093/toxsci/kfr061

Zoeller, RT. (2007). Environmental chemicals impacting the thyroid: Targets and consequences [Review]. Thyroid 17: 811-817. http://dx.doi.org/10.1089/thy.2007.0107

Zoeller, RT; Crofton, KM. (2005). Mode of action: developmental thyroid hormone insufficiency--neurological abnormalities resulting from exposure to propylthiouracil [Review]. Crit Rev Toxicol 35: 771-781. http://dx.doi.org/10.1080/10408440591007313

Zoeller, RT; Tan, SW. (2007). Implications of research on assays to characterize thyroid toxicants [Review]. Crit Rev Toxicol 37: 195-210. http://dx.doi.org/10.1080/10408440601123578