UNITED STATES ENVIRONMENTAL PROTECTION AGENCY
WASHINGTON, DC 20460

FEB 27 2018

MEMORANDUM

SUBJECT: Transmittal of Meeting Minutes and Final Report for the FIFRA Scientific Advisory Panel Meeting Held November 28-29, 2017

TO: Stanley Barone, Ph.D.,
Acting Director
Office of Science Coordination and Policy

FROM: Todd Peterson, Ph.D.,
Designated Federal Official
FIFRA Scientific Advisory Panel
Office of Science Coordination and Policy

THRU: Steven M. Knott, M.S.,
Executive Secretary
FIFRA Scientific Advisory Panel
Office of Science Coordination and Policy

Attached, please find the meeting minutes of the FIFRA Scientific Advisory Panel open meeting held in Arlington, Virginia on November 28-29, 2017. This report addresses a set of scientific issues being considered by the Environmental Protection Agency regarding the Continuing Development of Alternative High-Throughput Screens to Determine Endocrine Disruption, Focusing on Androgen Receptor, Steroidogenesis, and Thyroid Pathways.

Attachment

cc:

Nancy Beck
Louise Wise
Charlotte Bertrand
Seema Schappelle
Ronnie J Bever
Scott Lynn
Katie Paul-Friedman
Richard Judson
Rusty Thomas
Richard Keigwin
Anna Lowit
Anita Pease
Wayne Miller
Robert McNally
Marietta Echeverrria
Jackie Mosby
Dana Vogel
Delores Barber
Yu-Ting Guilaran
Mike Goodis
Linda Strauss
OPP Docket

FIFRA Scientific Advisory Panel Members
Dana Boyd Barr, Ph.D.
Marion F. Ehrich, PhD, DABT, ATS
David A. Jett, PhD
James McManaman, PhD
Joseph Shaw, PhD
Sonya K. Sobrian, PhD

FQPA Science Review Board Members
Ionnis Androulakis, PhD
Scott Belcher, PhD
Veronica Berrocal, Ph.D.
Rebecca Clewell, Ph.D.
Kristi Pullen Fedinick, Ph.D.
J. David Furlow, Ph.D.
Susan Nagel, Ph.D.
Michael Pennell, Ph.D.
Edward Perkins, Ph.D.
Thomas Zoeller, Ph.D.

FIFRA Scientific Advisory Panel
Meeting Minutes and Final Report
No. 2018 - 03

A Set of Scientific Issues Being Considered by the Environmental Protection Agency Regarding:

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

November 28-29, 2017
FIFRA Scientific Advisory Panel Meeting,
Held at the EPA Conference Center
One Potomac Yard,
Arlington, Virginia

NOTICE

The Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA), Scientific Advisory Panel (SAP) is a Federal advisory committee operating in accordance with the Federal Advisory Committee Act and established under the provisions of FIFRA as amended by the Food Quality Protection Act (FQPA) of 1996. The FIFRA SAP provides advice, information, and recommendations to the U.S. Environmental Protection Agency (EPA or Agency) Administrator on pesticides and pesticide-related issues regarding the impact of regulatory actions on health and the environment. The SAP serves as a primary scientific peer review mechanism of the EPA, Office of Pesticide Programs (OPP), and is structured to provide balanced expert assessment of pesticide and pesticide-related matters facing the Agency. FQPA Science Review Board members serve the FIFRA SAP on an *ad hoc* basis to assist in reviews conducted by the FIFRA SAP. The meeting minutes and final report are provided as part of the activities of the FIFRA SAP.

The FIFRA SAP carefully considered all information provided and presented by the Agency, as well as information presented by the public. The minutes represent the views and recommendations of the FIFRA SAP and do not necessarily represent the views and policies of the Agency, nor of other agencies in the Executive Branch of the Federal government. Mention of trade names or commercial products does not constitute an endorsement or recommendation for use.

The meeting minutes and final report do not create or confer legal rights or impose any legally binding requirements on the Agency or any party. The meeting minutes and final report of the November 28-29, 2017 FIFRA SAP meeting represent the SAP's consideration and review of scientific issues associated with "Continuing Development of Alternative High-Throughput Screens to Determine Endocrine Disruption, Focusing on Androgen Receptor, Steroidogenesis, and Thyroid Pathways." Steven Knott, M.S., FIFRA SAP Executive Secretary, reviewed the minutes and final report. James McManaman, Ph.D., FIFRA SAP Chair, and Todd Peterson, Ph.D., FIFRA SAP Designated Federal Official, certified the minutes and final report which is publicly available on the SAP website (http://www.epa.gov/sap/) under the heading of "Meetings" and in the public e-docket, Docket No. EPA-HQ-OPP-2017-0214, accessible through the docket portal: http://www.regulations.gov. Further information about FIFRA SAP reports and activities can be obtained from its website at http://www.epa.gov/sap/. Interested persons are invited to contact Todd Peterson, Ph.D., SAP Designated Federal Official, via e-mail at peterson.todd@epa.gov.

CONTENTS

FIFRA Scientific Advisory Panel
Meeting Minutes and Final Report
No. 2018 - 03

A Set of Scientific Issues Being Considered by the Environmental Protection Agency Regarding:

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

November 28-29, 2017
FIFRA Scientific Advisory Panel Meeting,
Held at the EPA Conference Center
One Potomac Yard,
Arlington, Virginia

Jim McManaman, Ph.D.
FIFRA SAP, Chair
FIFRA Scientific Advisory Panel
Date: FEB 27 2018

Todd Peterson, Ph.D.
Designated Federal Official
FIFRA Scientific Advisory Panel
Date: FEB 27 2018

Federal Insecticide, Fungicide, and Rodenticide Act
Scientific Advisory Panel Meeting
November 28-29, 2017

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

PARTICIPANTS

FIFRA SAP, Chair

James McManaman, Ph.D., Professor and Chief, Section of Basic Reproductive Sciences, Department of Obstetrics & Gynecology, Physiology & Biophysics, University of Colorado, Denver, Aurora, CO

Designated Federal Official

Todd Peterson, Ph.D., FIFRA Scientific Advisory Panel Staff, Office of Science Coordination and Policy, EPA

FIFRA Scientific Advisory Panel Members

Dana Barr, Ph.D., Research Professor, Department of Environmental and Occupational Health, Rollins School of Public Health, Emory University, Atlanta, GA

Marion F. Ehrich, Ph.D., Co-director, Laboratory for Neurotoxicity Studies, Professor, Pharmacology and Toxicology, Department of Biomedical Sciences & Pathobiology, Virginia-Maryland College of Veterinary Medicine, Blacksburg, VA

David A. Jett, Ph.D., Director, National Institute of Health Counter ACT Program National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD

Joseph Shaw, Ph.D., Associate Professor, School of Public and Environmental Affairs, Indiana University, Bloomington, IN

Sonya K. Sobrian, Ph.D., Associate Professor, Department of Pharmacology, Howard University College of Medicine, Washington, DC

FQPA Science Review Board Members

Ioannis Androulakis, Ph.D., Professor, Department of Chemical & Biochemical Engineering, School of Engineering, Rutgers, The State University of New Jersey, Piscataway, NJ

Scott M Belcher, Ph.D., Professor, Department of Biological Sciences, North Carolina State University, Raleigh, NC

Veronica J. Berrocal, Ph.D., Associate Professor, Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, MI

Rebecca Clewell, Ph.D., Chief Scientific Officer, Scitovation, Research Triangle Park, NC

J. David Furlow, Ph.D., Professor, , Dept. of Neurobiology, Physiology and Behavior, University of California, Davis, CA

Susan Nagel, Ph.D., Associate Professor, Obstetrics, Gynecology & Women's Health University of Missouri, Columbia, MO

Michael Pennell, Ph.D., Associate Professor, Division of Biostatistics, College of Public Health, The Ohio State University, Columbus, OH

Edward J. Perkins. Ph.D., Environmental Laboratory, U.S. Army Engineer Research and Development Center (ERDC), US Army Corps Engineers (USACE), Vicksburg, MS

Kristi Pullen Fedinick, Ph.D., Staff Scientist, Health and Environment Program, Natural Resources Defense Council, Washington, DC

Grant Weller, Ph.D., Research Statistician, Savvysherpa, Inc., Minneapolis, MN

Thomas Zoeller, Ph.D., Chair, Department of Biology, University of Massachusetts, Amherst, MA

LIST OF ACRONYMS AND ABBREVIATIONS

AC50	Concentration required to elicit a 50% response in an *in vitro* assay
Agency	United States Environmental Protection Agency
AO	Adverse Outcome
AOP	Adverse Outcome Pathway
AP-1	Activator Protein-1
AR	Androgen Receptor
AUC	Area Under the Curve
CASRN	Chemical Abstracts Service Registry Number
DHT	5α-dihydrotestosterone
DIO	Iodothyronine Deiodinase
DMSO	Dimethyl Sulfoxide
DUOX	Dual Oxidase
E2	Estradiol
EDSP	Endocrine Disrupter Screening Program
EDSTAC	Endocrine Disruptors Screening and Testing Advisory Committee
EPA	United States Environmental Protection Agency
ER	Estrogen Receptor
FIFRA	Federal Insecticide, Fungicide, and Rodenticide Act
FQPA	Food Quality Protection Act
HT	High-Throughput
HTS	High-Throughput Screening
IC50	Half-Maximal Activity. The Concentration of an Inhibitor Where the Response (or Binding) Is Reduced by Half
ICCVAM	Interagency Coordinating Committee on the Validation of Alternative Methods
IYD	Iodotyrosine Deiodinase
KE	Key Event
KER	Key Event Relationship
LT	Low-Throughput
MIE	Molecular Initiating Event
mMD	Mean Mahalanobis Distance
maxmMD	Maximum mean Mahalanobis Distance
MTT	Tetrazolium Dye MTT
NAS	National Academies of Sciences
NICEATM	NIH National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods

NIS	Sodium-Iodide Symporter
NR	Nuclear Receptor
OCSPP	U.S. EPA Office of Chemical Safety and Pollution Prevention
OECD	Organisation for Economic Co-operation and Development
ORD	EPA Office of Research and Development
PXR	Pregnane X Receptor
SAP	Scientific Advisory Panel
SARMS	Selective Androgen Receptor Modulators
SMILES	Simplified Molecular Input Line-Entry System
T	Testosterone
T3	3,3′,5-Triiodothyronine
T4	Thyroxine
TDCs	Thyroid Disrupting Chemicals
TH	Thyroid Hormone
ToxCast	EPA's Toxicity Forecaster
Tox21	Toxicology in the 21st Century – the NTP/NCGC/EPA/FDA consortium for chemical hazard HT
TPO	Thyroperoxidase
TR	Thyroid Hormone Receptor
TRH	Thyrotropin Releasing Hormone
TRHR	Thyrotropin Releasing Hormone Receptor
TSH	Thyroid Stimulating Hormone
TSHR	Thyroid Stimulating Hormone Receptor

INTRODUCTION

The Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA), Scientific Advisory Panel (SAP) completed its review of the set of scientific issues being considered by the Environmental Protection Agency (EPA) regarding the Continuing Development of Alternative High-Throughput Screens to Determine Endocrine Disruption, Focusing on Androgen Receptor, Steroidogenesis, and Thyroid Pathways. Advance notice of the meeting was published in the *Federal Register* on June 6, 2017. The review was conducted in an open Panel meeting held in Arlington, Virginia, on November 28-29, 2017. The White Paper, supplemental files, and related documents in support of the SAP meeting are posted in the public e-docket at http://regulations.gov (ID: EPA-HQ-OPP-2017-0214). Dr. James McManaman chaired the meeting. Dr. Todd Peterson served as the Designated Federal Official.

In preparing these meeting minutes and final report, the Panel carefully considered all information provided and presented by the Agency presenters, as well as information presented by public commenters. These meeting minutes and final report address the information provided and presented at the meeting, especially the Panel response to the Agency charge.

During the FIFRA SAP meeting, US EPA personnel provided the following presentations (listed in order of presentation):

Welcome – Stanley Barone, Ph.D., Acting Director, Office of Science Coordination and Policy (OSCP), EPA

Welcome and Opening Remarks – Seema Schappelle, Ph.D., Director, Exposure Assessment Coordination and Policy Division (EACPD, Office of Science Coordination and Policy (OSCP)

Background – Ronnie Joe Bever, Ph.D., DABT, EACPD, OSCP

Androgen Receptor (AR) Pathway Activity – Richard Judson, Ph.D., Office of Research and Development (ORD), and National Center for Computational Toxicology (NCCT)

Discussion of the Second-Generation AR Pathway Model – Ronnie Joe Bever, Ph.D., DABT, EACPD, OSCP

Steroidogenesis Pathway Activity – Katie Paul-Friedman, Ph.D., ORD and NCCT

Discussion of the Steroidogenesis Assay – Ronnie Joe Bever, Ph.D., DABT, EACPD, OSCP, EPA

Description of the Developing Thyroid Conceptual Framework and Challenges
Scott Lynn, Ph.D., DABT, EACPD, OSCP, EPA

PUBLIC COMMENTERS

Oral statements were presented as follows:

Ellen Mihaich, Ph.D., DABT, Environmental and Regulatory Resources, LLC, on behalf of the Endocrine Policy Forum

Christopher Borgert, Ph.D., Applied Pharmacology and Toxicology, on behalf of the Endocrine Policy Forum

Steve Levine, M.S., Ph.D., American Chemistry Council, on behalf of the Endocrine Policy Forum

Brandy Riffle, Ph.D., BASF Corporation, on behalf of the Endocrine Policy Forum

Catherine Willett, Ph.D., on behalf of the Humane Society of the United States

Esther Haugabrooks, Ph.D., on behalf of the Physicians Committee for Responsible Medicine

Written statements were provided as follows:

Ellen Mihaich, Ph.D., DABT, Environmental and Regulatory Resources, LLC, on behalf of the Endocrine Policy Forum

OVERALL SUMMARY

The U.S. Environmental Protection Agency (Agency) 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 is committed to the use of validated HT assays and computational models to: 1) prioritize chemicals for further EDSP screening and testing based on predicted bioactivity; 2) use as alternatives to EDSP Tier 1 assays; and 3) contribute to the weight-of-evidence evaluation of the potential endocrine bioactivity of a chemical. The Panel was charged with advising the Agency on these areas of interest in relation to: an androgen receptor model, steroidogenesis model and a thyroid pathway conceptual framework. The Agency's White Paper as well as Agency presentations at the November 28-29, 2017 SAP meeting discusses these topics:

Androgen Receptor (AR) Activity

The Agency presented an updated approach for determining androgen bioactivity based on a computational model integrating data from 11 HT screening assays.

Steroidogenesis Pathway Activity

The Agency presented an approach that describes the development of a HT H295R steroidogenesis model and a novel statistical approach for this model. Two variations in the analysis of the HT H295R assay results were 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. The second variation uses a novel statistical approach to integrate the measurements of 9 additional steroid hormones from the HT H295R assay.

Thyroid Conceptual Framework

The Agency presented initial work in establishing a framework utilizing thyroid-related molecular initiating events (MIEs) in an adverse outcome pathway context, and the status of developing a set of HT assays for a subset of these thyroid-related MIEs. The ultimate goal of the Agency's framework is to identify potential thyroid disrupting chemicals (TDCs).

Overall, the FIFRA SAP highlighted advancements and progress in all topic areas. With the most work to date on the AR model, the Agency asked the Panel for comments in anticipation of adopting the HT model as an alternative to the LT Tier 1 assay. For the AR model the Panel discussion is in part a retrospective assessment of progress made on the model in light of the prior, 2014, SAP comments. Attention to the 2014 SAP

comments brought new discussion and further recommendations. The Panel indicates further attention to specific points is needed before moving forward to full acceptance of the model.

Of the three topics discussed, the AR model represents the area of greatest effort to date. The work on steroidogenesis is an area of active ongoing development and the thyroid pathway model effort is in the early stages.

Questions asked of the Panel regarding steroidogenesis address the strengths and limitations of multiple hormone responses and related analysis and the statistical integration of the multiple responses assessed by the assay. The Panel encourages further development of the steroidogenesis assay and identified both strengths and limitations leading to recommendations in responses to all three questions.

The Panel likewise encourages further work on the thyroid pathway assay with a number of technical points raised when discussing the complexity of the thyroid biology and multiple endpoints in relation to specific MIEs and key events (KE) which the Agency presented in the White Paper and during presentations at the SAP meeting.

EXECUTIVE SUMMARY

TOPIC: Androgen Receptor (AR) Pathway Activity

The U.S. Environmental Protection Agency's (Agency) AR pathway model is a potential alternative for the existing Endocrine Disruptor Screening Program (EDSP) Tier 1 AR binding assay. The 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. Critical to understanding the predictive ability of the model is the performance of the model with reference chemicals, and systematic curation of data sources to define this set of reference chemicals.

The mammalian AR signaling pathway was probed using a set of 11 biochemical and cell-based *in vitro*, high-throughput (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 activated across technologies; 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.

The computational approach to combine information from multiple AR assays is very similar to the approach previously used to predict estrogen receptor (ER) activity. This pathway approach attempts to minimize the incidence of false negatives by using a consensus result based on the understanding of where a chemical may act in the AR pathway. Computational and pathway models were discussed in the White Paper (see Section 1.6).

The White Paper presents 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. The Panel in 2017 made an assessment based on the following single charge question.

Question 1: Please comment on the Agency's efforts to address the suggestions of the previous SAP, thus confirming the suitability of the current HT AR pathway model to be used as an alternative to the low-throughput (LT) Tier 1 AR binding assay (OCSPP 890.1150).

Summary

The Scientific Advisory Panel (SAP, Panel) finds that the Agency has made a great effort to address the comments raised by the previous SAP, particularly with respect to accounting for uncertainty, assay interference, cytotoxicity, expansion of the assay battery, and extension of the method to a larger number of reference chemicals, in addition to transparency with data, methods, and results. This new model addresses many concerns raised by the previous SAP for improving the scientific basis of the pathway model. While use of this model to prioritize chemicals for testing under the EDSP is reasonable, there are remaining issues to address before the model is suitable for use as an alternative for the LT Tier 1 AR binding assay.

The Agency's efforts to distinguish between cell toxicity/cell stress, assay interference and authentic AR antagonism using a z-score based on confirmatory *in vitro* antagonist assay data, and cell stress/cytotoxicity information, are considered valuable and appropriate to address the SAP comments. However, panelists meeting in 2017 suggested that the effort could be improved by adding assays that probe non-classical mechanisms of protein regulation and that confidence scoring needs to optimized.

The Panel considered the Agency response to the 2014 SAP comment on optimizing assessment of activities, particularly antagonism, is satisfactory. Overall, the panel feels that the addition of confirmatory assays is a clever and effective way to confirm that the action of a particular chemical is specific to the AR pathway. However, some panelists noted that the relatively few chemicals tested due to technical limitations of the ToxCast dataset (e.g. use of DMSO solvent) weakened confidence in the model.

The Panel found that the Agency response to the 2014 SAP comment to build upon the battery of AR bioactivity assays appears to be adequate. However, providing a biological argument that no key assays have been missed would strengthen this response.

The Panel considers the Agency response adequately addresses the 2014 SAP comment to address the narrow area under the curve (AUC) value range, to include a wider range of chemicals among different structural classes, and to inform future studies using these methodologies. The Agency addresses this suggestion by analyzing 1855 different chemicals and using a robust systematic review process to identify 65 chemical standards that had a range of potencies. However, one panelist indicated the Agency should provide

greater coverage of the EDSP universe by selecting additional reference chemicals representing different clustering groups such as those groups identified by Jarvis Patrick clustering. Another panelist notes that the *in vitro* assays used by the AR model should also be examined using ethanol or water as solvents for test chemicals to ensure that responses are similar to those determined by the current AR binding assays.

The Panel considers the Agency's response to the 2014 SAP Comment: "Measures should be taken to demonstrate that results from the model are reproducible" inadequate. While the incorporation of uncertainty estimates via a bootstrap resampling approach is particularly commendable, more details are needed to understand whether the confidence intervals constructed using bootstrap resampling correctly account for all different types of uncertainties. Data fitting functions as used in the model may have resulted in model overfitting. As a result, the Panel recommends examining the performance of the model on a set of chemicals that are not in the set of chemicals to which functions were fitted to truly determine the performance of the AR model. A more suitable validation approach would be to provide to an independent group the following information for assessment: a description of the mathematical functions needed for construction of the model R-code, *in vitro* assay test data to which Rj values were fit in the Agency model, and independent testing of model reproducibility using data to which Rj values were not fit.

The Panel felt that the Agency's response to the 2014 SAP comment, recommending that attention should be given to alternative, non-classical pathway AR-related assays, metabolism of chemicals, and potential off-target effects, is appropriate. Since the Agency is currently focused on assessing whether or not the AR model is a suitable alternative to the AR binding assay, the Agency did not evaluate non-competitive mechanisms of antagonism.

The Panel considers that the Agency's response to the 2014 SAP comment that details of the methods can be improved and further results must be available to increase transparency. Overall, the Agency has made details and results of the model available to increase transparency. While providing code and data is a step in the right direction, a detailed description of the algorithm used would be appreciated, particularly for those who may not be able to interpret R code.

TOPIC: Steroidogenesis Pathway Activity

The Agency next presented a second area of consideration with a set of objectives for the screening methodology for the steroidogenesis pathway, including:

1. A comparison of 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. Introduction of a novel statistical approach that integrates the measurements of E2, T, and 9 additional steroid hormones from the HT H295R assay to quantify the overall impact of the substance on the steroidogenesis pathway.

3. Providing a regulatory perspective on potential future use of the HT H295R assay.

4. The Panel was charged with providing responses for the following three charge questions.

Question 2: Based on the comparison of the performance of the HT H295R assay with the LT H295R assay, and the effects of reference chemicals on the synthesis of T and E2 levels only, please comment on the suitability of the HT H295R assay as an alternative to the LT H295R assay. See Sections 3.3 and 3.4.

Summary

In considering performance, reference chemicals, and the suitability of the high-throughput (HT) as an alternative to the low-throughput (LT) assay, the Panel agrees overall that the performance of the HT H295R assay, in its current form, presents some clear benefits. Additional points made by the Panel concern additional performance optimization along with transparent demonstration of assay reproducibility, reliability, and portability are needed before the HT is deemed a suitable alternative for the LT H295R assay.

Advantages incorporated into the HT assay include the use of 96-well cell culture format and the 48-hour stimulation by a forskolin pretreatment. The Panel however cautions that sensitivity of the assays may be decreased by this approach and specific investigation of the impact on the dynamic range of the assay and possible optimization is needed.

In assessing the status of the HT assay, the Panel recommends the Agency provide additional quantitative data for a comparison of the HT and LT assays. A quantitative comparison of the relative potencies of the positive controls in each assay is needed, along with evidence that demonstrates sensitivity of the HT assay in comparison to the LT assay. At present, for the HT assay, there is no analysis of relative potency of positive

controls as only the maximum concentration tested is listed, thus limiting the ability to compare assays.

The Panel considers inclusion of the cell viability assessment as a strength for the HT assay, but with some concerns about a reduction in the maximum allowable loss in viability, from 80 to 70%, from the guideline viability standard used for the LT assay. Further justification for the lower standard is needed in light of potential biological importance of a greater than 20% loss in viability and subsequent impacts and any negative impacts on assay selectivity, performance, and interpretation of results. Further, a generous viability cutoff would potentially inflate "hit calls" in the assay due to off target toxicity. The panel also noted that the specific measure used here for decreased viability is related to alteration of mitochondrial function, which is particularly important to steroidogenesis. The Panel suggested further evaluating the appropriateness of a 70% viability cutoff by: comparing assay performance at 70% versus greater than 80% viability, considering the incorporation of an appropriate cytotoxicity z-score (similar to AR model) into the analysis rather than the ATP assay, and investigation of uncoupling the cell viability assessment from mitochondrial function (i.e. another measure of cytotoxicity).

Regarding reference chemicals used in the inter-lab analysis of the OECD guideline H295R steroidogenesis assay (Hecker et al, 2011), the HT H295R assay performed with relatively less sensitivity. One Panel member stated that failure of the HT H295R assay to accurately identify reference chemicals disrupting E2 and T production renders the current assay inadequate for protecting public health. The Panel noted that for the set of reference chemicals used in the inter-lab analysis of the OECD guideline LT steroidogenesis assay, the HT assay appears to be performing with relatively less sensitivity. This suggests performance of the HT assay presently does not meet the requirements for assay detection of endocrine disrupting chemicals as set forth in the final report of the Endocrine Disruptor Screening and Testing Advisory Committee.

Additional concerns indicated by the Panel include replication or reliability of the HT assay and the approach used for comparative analysis. In contrast to the OECD evaluation of replications from 11 laboratories from around the world, the Agency presents HT assay data from a single laboratory which indicated an apparent difficulty in replication across different assay blocks. The Panel recommends establishing the reliability of the assay/analysis from day-to-day (across blocks). This concern extended to the ability to replicate assay results for future testing and in different labs.

The Agency did not indicate how many times individual reference chemicals were analyzed. The Panel noted that specific information on the number of biological replicates is needed to compare the reproducibility of the results for the reference chemicals and to assess whether the statistical approach used for comparison is

appropriate. Additional concerns expressed by the Panel include an inability to fully assess the appropriateness of the pre-screening approach. The Panel notes the goal of a screen is to cast a wide net with an eye on setting priorities and that the Agency needs further justification that these tests are better than the current method. That is, screening assays should be fit-for purpose, high quality, rigorous and with reproducible methodology, yet with a good match with available resources. However, while the prescreening approach allows more chemicals to be tested quickly, which is important for ToxCast, using only the limit (highest non-cytotoxic) dose could result in reduced ability to identify compounds with complex dose-response curves, or compounds with borderline cytotoxicity. The Agency should also demonstrate that the HT approach does not undermine the purpose of the multi-concentration approach to capture dose-response. Some panel members recommended omitting the prescreen and using full dose-response evaluations for chemicals of interest.

Question 3: Please comment on the strengths and limitations of integrating multiple hormone responses beyond T and E2 (*i.e.* 11 hormones vs 2 hormones) in a pathway-based analysis of the HT H295R assay. Please comment on the suitability of this HT H295R pathway model (using 11 hormones) to serve as an alternative to the LT H295R assay. See Section 3.7.2.

Summary

The Panel moved from a discussion on the suitability of the HT assay to considering the strengths and limitations of integrating multiple hormone responses beyond T and E2 in a pathway-based analysis of the HT-H295R assay. In light of the strengths and limitations, the Panel was again asked to address use of the HT assay as an alternative to the LT assay. Overall, the HT assay provides more information from the measurement of multiple hormones, the use of the Mahalanobis distance metric, and improved sensitivity. These all contribute to the assay's future use for prioritization as an alternative to the LT assay.

Strengths as noted by the Panel include:

1. A comparison of assays indicates potential increased accuracy for the HT over that for the LT assay.

2. The HT assay monitors an integrated response for multiple pathway components as opposed to isolated, individual elements, and offers higher sensitivity and additional, mechanistic, information.

3. The 11 measured hormones represent 4 distinct classes, adding diversity to the assay.

4. The HT assay uses the same cell lines as the LT assay allowing for comparisons.

5. The “revised” confusion matrix elements indicates a strong correlation in performance in characterizing E2 and T compared to the LT assay.

6. The use of a modified Mahalanobis distance metric is a creative solution that enables integration of multiple features into a single metric.

Limitations or areas needing further attention noted by the Panel, include:

1. HT method lacks validation across multiple laboratories and fewer technical and biological replicates were tested for the HT assay compared to the LT Assay (3 in LT, only 1 in HT).

2. There appear to be no analyses of relative potency of positive controls as only the maximum concentration tested is presented for the HT assay—this is needed to allow for a quantitative comparison between the two assays.

3. Use of the aggregate Mahalanobis score for the complex hormone release patterns compared to the confusion matrices, based on analysis of individual hormones, needs validation and further clarification to allow clearer interpretation of the Mahalanobis score for the HT assay results.

4. Determination of whether significantly more chemicals are identified when additional hormones are measured by the HT assay is needed.

5. The Mahalanobis metric needs to be assessed for the weakly active chemicals that hit only 1 or 2 hormones.

6. Use of the HT assay for prioritization purposes is likely appropriate, however, classification of “progestogen disruptor” or “corticosteroid disruptor” based on an assay with no positive or negative controls for these pathways is questionable.

During the discussion some of the Panel members continued to express concern, as was the case for second charge question, for using the 70% versus 80% cell viability standard. These Panel members believe that, although 70% viability is the statistical limitation of the assay, biologically, 30% loss of viability is high and likely affects results. These members advise providing additional justification of this limitation and assess how the results change if the viability cutoff were 80% as in the original assay.

Even with the cited limitations and need for additional work, the Panel generally expressed that the HT-H259R is a scientifically sound potential alternative to the LT H295R. The Panel recommends additional analyses to support assay conditions and methods before implementation of the HT assay in the Endocrine Disruption Screening Program (EDSP). Furthermore, the assay should be validated against chemicals affecting corticosteroid/progestogen pathways where there is no positive/negative control data.

Question 4: The work herein presents a novel statistical integration of multiple hormone responses indicative of steroid biosynthesis in the HT H295R assay. A summary statistical metric, the maximum mean Mahalanobis distance (maxmMd), has been suggested as a tool for use in prioritization of chemicals. In addition to the use of the maxmMd to indicate the magnitude of potential effects on the steroid biosynthesis pathway expressed in H295R cells, an examination of the hormone responses that contribute to the maxmMd may provide valuable biological information to inform the weight-of-evidence evaluations performed for chemicals subjected to EDSP Tier 1 evaluation. Please comment on the strengths and limitations of using the maxmMd and the pattern of steroid hormone responses in the HT H295R assay for chemical prioritization and weight-of-evidence applications. See Sections 3.2.4, 3.3.2, and 3.7.2.

Summary

The Panel's review of the proposed maximum mean Mahalanobis distance approach, as a tool for chemical prioritization, identifies both strengths and limitations.

Strengths as noted by the Panel, include:

1. The mean Mahalonobis distance (mMD) is the multi-dimensional equivalent of the z-score for univariate normally-distributed observations that:
    a. can be used to flag outliers; and
    b. allows the combination of multiple hormone responses measurements into a single summary measure, while accounting for the variability of each individual hormone response measurements.

2. The proposed framework for prioritization of chemicals based on the maxmMD computed over multiple concentrations is a conservative approach for flagging a chemical as an outlier with respect to controls.

Limitations cited and recommendations offered by the Panel, include:

1. There is difficulty in identifying what type of effect a chemical must impose on the steroid biosynthesis pathway in order to be flagged.

2. The Panel sees a need for further clarification to assess whether the mMD approach:
    a. tends to flag chemicals that deviate from the expected relationships between hormone responses.
    b. allows to prioritize chemicals that display absolute differences from controls when the sampling distribution of the residuals is not normal.

3. The Panel was concerned with the critical values used and the Type I error rate. The Panel recommends:
    a. Conducting simulation experiments that evaluate the Type I error rate of the proposed method using the data in the White Paper.
    b. Cite the simulation studies performed when describing this methodology.
    c. Provide a rationale for the use of a 1% Type I error rate instead of a more conventional 5% Type I error rate.

4. The Panel was concerned with the appropriateness of the estimated covariance matrix used to derive the mMD. The Panel suggests that a more thorough investigation of the behavior and appropriateness of the estimated covariance matrix be carried out as incorrectly estimating the sample covariance matrix, might overestimate the variability and thus lead to an inflation of the Type II error rate.

5. There are minor concerns regarding values that fall below the limit of detection, use of terms, labels on Figure 3-10, and other details provided below in the discussion of the charge.

6. The Panel recommends that a distance metric such as Tukey's half space depth be investigated due to its appealing characteristic of being a nonparametric method to rank-order multivariate observations.

7. The Panel advises additional methods for comparison of multidimensional vectors that represent biological pathways or networks.

The Panel concluded that although the maximum mean Mahalanobis distance might not be the optimal statistical approach to integrate multiple hormone responses due to some limitations or due to the fact that the approach does not take into account biological pathways, the Agency is moving in the right direction in the effort to develop a framework to assess chemicals' potential for effect on steroidogenesis.

TOPIC: Thyroid Conceptual Framework

The third area for the SAP's consideration included assessing the current work in the Agency's effort to developing an EDSP strategy for a thyroid conceptual framework to identify potential thyroid disrupting chemicals (TDCs). The White Paper outlines known thyroid-related pathways, reviews thyroid-related molecular initiating events (MIEs) in an adverse outcome pathway (AOP) context, and presents the status of a developing set of high throughput (HT) assays for a subset of these thyroid-related MIEs.

The Panel was given the following two charge questions to assess the Agency's strategy in its early stages:

Question 5: Please refer to White Paper Section 4.2. EPA has identified AOPs for thyroid hormone disruption related to potential xenobiotic-induced alterations of thyroid homeostasis. Please comment on the completeness of the MIEs (Table 4-1), KEs, and adverse outcomes within the thyroid AOP network (Figure 4-1). Also, please provide information on any missing pathways, adverse outcomes, or other AOP-related information (*e.g.* MIEs or KEs) critical for capturing the complexity of systems biology controlled by thyroid hormones.

Summary

The Panel acknowledges that the Agency includes a 'largely complete' set of molecular initiating events (MIEs) and key events (KEs) in the White Paper. The Panel then turned its attention to Table 4-1 (i.e. Potential MIEs for Thyroid-Based AOPs), to make a set of requests to add information to the White Paper presentation, including, but not limited to:

1. Adding a new column to Table 4-1 to include adverse outcomes that would be predicted to result from interference with the MIE identified in that row.

2. Use a single row for each MIE (i.e., protein target) rather than lump them into classes.

3. Adopt language, including the use of the term 'distributor protein,' to be consistent with and cognizant of the most recent developments.

4. For the Hepatic Nuclear Receptors, identify the specific receptors that are related to serum T_4 and T_3 clearance (each would be a separate MIE).

5. The regulation of thyrotropin-releasing hormone (TRH) synthesis or neuronal activity may be important and could be separately identified in Table 4-1.

The White Paper Table 4-2 describes the Tier 1 and Tier 2 assays. The Panel's recommendations for this table include:

1. For Tier 1, thyroid-specific endpoints of serum T_4 and thyroid-stimulating hormone (TSH), thyroid weight and thyroid histopathology are known to be separable and as the Panel's detailed response indicates, the HT assays need to adopt strategies for addressing distinctions for these end points to achieve a reasonable balanced accuracy for the HT assays.

2. The Agency identifies a 10% reduction in serum T_4 as an adverse outcome, but growth and body weight may not be affected until the most severe of circumstances. The Panel recommends the Agency stipulate that many adverse outcomes will occur while growth and body weight remain normal.

3. For tier 2, thyroid-specific neurohistopathologic changes should be identified.

While the White Paper Figure 4-1 complements Table 4-1 well, the figure lacks the level of detail to support the Agency's use as a tool. The Panel's detailed response provides comments to reinforce information important to the presentation in Table 4-1.

Question 6: Please refer to White Paper Section 4.3. EPA has summarized currently available assays and test guidelines informative of thyroid AOPs and is developing HT assays for a number of MIEs. Please comment on the ranked importance of MIEs (Table 4-3) and on whether assays for environmentally important MIEs are missing, and include information on both the biological and environmental relevance of these MIEs. In addition, please comment on other assays that would supplement or be orthogonal to the assays currently identified in Table 4-3 or for other KEs or AOs in the thyroid AOP framework (Figure 4-2).

Summary

The panel appreciated the overall construct of the Adverse Outcome Pathway (AOP) as the best way to organize the conceptual framework that will guide ongoing and future screening efforts for how environmental chemicals may impact the thyroid hormone endocrine system. The Panel recommended that the Agency should provide a clearer definition of what high, medium, and low ranking means in terms of priority for action and proposed timelines (Table 4-3: HT assay status and prioritization ranking of MIEs). As such, the Panel response outlines a suggested high, medium, and low ranking for specified MIEs. Supporting information for each specified ranking is detailed further in the next section of this report.

High:

A. The sodium/iodide symporter (NIS)
B. Thyroperoxidase (TPO)
C and D. Hepatic TH metabolism and PXR (pregnane X receptor)
E. The iodothyronine deiodinases (DIO)(Types I, II and III (D1, D2 and D3, respectively))

Medium:

A. Thyroid hormone regulated transcription (initiated at the TRs):
B. Serum TH transport proteins (also known as distributor proteins)
C. Membrane Transporters
D and E. TRH receptor (TRHR) and TSH receptor (TSHR) assays

Low:

A. Thyroid hormone receptor binding (*in vitro* assays)
B. For other steps of TH synthesis beyond NIS and TPO (e.g. pendrin, DUOX, IYD)

A detailed discussion and rationale for ranking each MIE and further considerations for any missing assays or MIEs is provided in the detailed response below.

Recommendations:

The Panel suggested a set of orthogonal (mostly transcriptomic-based) and gap – filling (RXR, biotransformation) assays to support the emerging direction of the TH disruption program.

The Agency should clarify what is meant by high, medium, and low ranking of MIEs as a means for future Panels to evaluate, in real terms, subsequent decision-making processes.

Lessons learned from estrogen and androgen disruptor programs could inform MIE assay development for the thyroid AOP context.

A clear understanding of how many orthogonal assays for each MIE are required for high-level confidence in sensitivity and specificity would be very useful.

A need for systems modeling across MIEs, cell types, species and life stages to fully integrate and validate the high throughput screening program is recognized and the Agency is encouraged to pursue this modeling.

DETAILED PANEL DISCUSSION AND RECOMMENDATIONS

The United States Environmental Protection Agency's (Agency) Endocrine Disruptor Screening Program (EDSP) must use validated assays to screen and test for endocrine disrupting chemicals. Since the issuance of the June 19, 2015 Federal Register Notice (US EPA 2015), the Agency has continued the development of high throughput assays and computational tools for the detection of the potential to disrupt the endocrine system. The SAP is asked to provide review and comment on the Agency's: (1) proposed high-throughput computational model of androgen receptor binding as an alternative to the current Tier 1 androgen receptor assay (OCSPP 890.1150: Androgen Receptor Binding [Rat Prostate Cytosol]); (2) development of high-throughput computational model of steroidogenesis to be used as an alternative to the current Tier 1 steroidogenesis assay (OCSPP 890.1550: Steroidogenesis [Human Cell Line – H295R]); and (3) proposed thyroid toxicity pathway framework.

Please provide comment and advice on the following questions. In addressing these questions consider the completeness of the data sets evaluated.

TOPIC: Androgen Receptor (AR) Pathway Activity

In December 2014, the Agency and the NIH National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) introduced an AR pathway model during the Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA) Scientific Advisory Panel (SAP). At the time, the model integrated 9 assays and was evaluated using 23 reference chemicals. In accordance with the SAP's suggestions, the model was expanded and now includes 11 assays and has been evaluated using 65 reference chemicals of varying potencies. The SAP also asked that cytotoxicity and cell stress be monitored and confirmatory tests be employed. In the current model, cell stress and cytotoxicity are assessed using a statistical measure called a z-score and a second confirmatory assay for AR antagonists was performed and integrated into the model. For a summary of the SAP's comments and the Agency's responses, please see Section 2.5.2 of the White Paper. For a full description of the AR model, see Section 2.

Question 1: Please comment on the Agency's efforts to address the suggestions of the previous SAP, thus confirming the suitability of the current HT AR pathway model to be used as an alternative to the low-throughput (LT) Tier 1 AR binding assay (OCSPP 890.1150).

Response

The following sections address each of the SAP comments made in December 2014, the Agency response, and the current SAP observations and recommendations made during the November 2017 meeting of the SAP.

December 2014 SAP 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 as detailed in Section 2.2.5 (of the White Paper) 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 (of the White Paper).

November 2017 Comment on Agency Response

Overall, the Panel feels that the Agency has done well in adding a Caution Flag or Cytotoxicity Filter, based on Cytotoxic and Cell Stress flags, to the model to address concerns about cytotoxicity interfering with true AR responses. The Agency tried to incorporate cytotoxicity and cell stress in the proposed framework while also accounting for the additional source of uncertainty that cytotoxicity and cell stress introduce in the assay data. While using a z-score approach to flag AC50 values considerably below the median AC50 for cytotoxicity is somewhat informal, it does effectively compare the toxicity identified in the assays to expected cytotoxic effects, and can flag any response well outside the expected range for cytotoxicity. One panelist indicates that the approach undertaken for confidence scoring is not yet optimal and still requires some work. In particular, Figure 2-9 in the White Paper showed a rather large spread of AUC values within each confidence score class. Ideally, it would be better to have a greater separation between the different confidence score classes. More formalization of this use of the cytotoxicity metric is needed. A Panel member asked: How will this metric be applied to new chemicals that are not tested in the entire ToxCast/Tox21 battery of assays? Another panel member also commented that there is an error in the document regarding the direction of subtraction in the z-score; as currently presented a highly negative z-score should flag a chemical with non-cytotoxic activity and not as a highly positive score.

The two additional assays probing antagonist behavior added some additional value, however the assays are limited due to their ability to only probe competitive mechanisms of antagonism. Antagonism of the androgen receptor can be initiated via non-ligand binding mechanisms (Jones 2009). The Panel noted that additional assays that probe non-classical mechanisms of protein regulation are essential to ensuring that biological functions are not missed in prioritization and screening tests.

The current model, particularly the use of the confidence score, is a major improvement over the ER model. While the confidence score could be a useful addition to the method, particularly compared to the ER Model, it is unclear how this scoring metric compares to results from the Tier 1 assay. The Panel found that the chart on slide 71 of the Agency's presentation (see figure below) appears to show the comparison results of the AUC scores of the AR Model versus Tier 1 List 1 AR binding assay – not a comparison of the Tier 1 List 1 results against the AR Model confidence score. If this is in fact the case, it is not straightforward to assess the ability of the confidence score to properly assign positive or negative status to standards or other chemicals in the EDSP universe.

Results vs Tier 1 AR Binding Assay

Source	ICCVAM	ICCVAM	EPA List 1	EPA List 1	EPA List 1
Class	Active	Inactive	Active	Inconclusive	Inactive
Number	24	31	9	7	31
AR Agonist Positive	8	2 (B)	0	0	0
AR Antagonist Positive	14	7 (B)	2	0	3 (D)
AR Model Inconclusive	1 (A)	3 (B)	0	1	1 (D)
AR Model Negative	1 (A)	19	7 (C)	6	24

Slide 71 of 171

Careful assessment of the general properties of solvent and test chemicals in *in vitro* assays should be considered. These factors are critical for the AR bioactivity assays due to the prevalence of chemicals that predominantly express antagonist activity rather than agonist activity. Methyltrienolone (R1881) is used as the reference androgen in all AR binding assays. Since 5α-dihydrotestosterone (DHT) is metabolized by animal tissue cytosolic preparations and also by many cell lines, R1881 is the reference androgen of choice for binding assays and *in vitro* AR TA assays. However, some substances, when dissolved in DMSO, appear to bind with lower affinity to the receptor. Therefore, final concentrations greater than 0.1% are not used. Currently all chemicals tested are DMSO soluble.

The current LT Tier 1 AR binding assay allows for testing chemicals that are water-soluble. This is a drawback of the HTS. However, during the presentations, the Agency said that testing on water soluble chemicals in the HT assays has begun and will continue but at a low priority, although no data was presented to demonstrate this ability.

December 2014 SAP 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.

November 2017 Comment on Agency Response

Overall, the panel felt that the addition of confirmatory assays is a clever and effective way to confirm that the action of a particular chemical is specific to the AR pathway. For chemical screening, the final report of the Endocrine Disruptor Screening and Testing Advisory Committee (EDSTAC) recommends that Tier 1 assays "be more 'sensitive' than they are 'specific,' meaning that they should have as their primary objective the minimization of false negative or Type II errors, while permitting an as-of-yet undetermined, but acceptable, level of false positive or Type I errors" (EPA 1999). The Agency response to questions about false negatives being allowed due to the prioritization aspirations of these assays is misleading, in that these tests will be used for *both prioritization and screening*. In fact, this charge question specifically asks for the ability of the AR Model to serve as an alternative to a Tier 1 screening test. The inability to evaluate the performance of the model for chemicals that reside outside of the limited chemical standards tested (due, in part, to technical limitations of ToxCast) in the AR Model limit the confidence in this particular method.

One panelist expressed that the Agency should better address the compression of AUC scores. The current AUC value range is narrow and lacks significant magnitude/range for discriminating between AR bioactivity values/scores that assigned to specific chemicals. Further, the Endocrine Policy Forum presents cogent arguments regarding the need to eliminate compression of AUC scores. The Agency should explore methods to eliminate compression of AUC scores.

One panelist felt that the assay data sets suffered from limitations due to the fact that chemicals were only tested using DMSO as a solvent in assay media. However, the panel recognizes that DMSO is used commonly as a solvent in which to dissolve test chemicals

for most, if not all, high throughput assays. DMSO is a recommended solvent in the Tier 1 AR binding assay although ethanol or water are preferred (see EDSP 2011). The Agency indicated work continues on water solubility.

December 2014 SAP 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.

November 2017 Comment on Agency Response

In the current AR pathway model, two additional assays are added by the Agency. *In vitro* assays used by the AR model now include 3 biochemical radioligand AR binding assays, one transactivation assay measuring reporter RNA transcript levels, three transactivation assays measuring reporter protein level readouts, and two transactivation antagonist assays. The Panel noted that addition of the two competitive binding assays seems helpful for increasing the ability of the model to detect antagonists. The Agency argues that more assays are probably not necessary due the excellent predictive capability of the model. While this is understandable from a statistical standpoint, the Agency should provide a biological argument that no key assays were missed. In addition to assays that extend the technical capabilities of the assays, it would be beneficial for the Agency to explore the use of higher maximum concentrations in order to reduce the false negative rate found during the comparison of the Tier 1 List 1 results to the AR Model. Overall, the panel felt that the Agency's response to the comment is adequate. Providing a biological argument that no key assays have been missed would strengthen this response as well as demonstrating that technical limitations in the assay do not prohibit testing of a wider range of chemicals within the EDSP universe.

December 2014 SAP 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 are 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 agonists 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, (of the White Paper) respectively.

November 2017 Comment on Agency Response

One Panel member notes the Agency adequately addresses a wider range of chemicals by analyzing 1855 different chemicals and using a robust systematic review process to identify 65 chemical standards that cover a range of potencies.

Another member indicated the Agency should provide greater coverage of the EDSP universe by selecting additional reference chemicals representing different clustering groups. For example, Jarvis Patrick clustering (Kmin = 5; K = 10) identifies 2,797 clusters across 6,447 chemicals including 6,425 chemicals in the EDSP universe with available chemical CASRN/SMILES. In comparison, the systematic review selected standards covered only 36 of the clusters identified.

Further, while the ToxCast system focuses on a wide range of DMSO-soluble chemicals, the use of DMSO as a solvent may lead to different results than when ethanol or water is used as a carrier solvent. The current Androgen Receptor Binding Assay (OCSPP 890.1150) allows for the use of ethanol, water, or DMSO as solvents for chemical solubility. Therefore, the Agency should demonstrate that equivalent results can be obtained using water or ethanol as solvents prior to acceptance of the HT AR assay as an alternative to the AR binding assay.

December 2014 SAP Comment

Measures should be taken to demonstrate that results from the model are reproducible.

Agency Response

Results of uncertainty analysis run for the model (see Section 2.2.7), are reported by Kleinstreuer et al. (2017): Figure S7 (see tx6b00347_si_001.pdf in Kleinstreuer et al., 2017) 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., 2017). 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., 2017). These results demonstrate adequate reproducibility for the model. (for referenced PDF files go to: http://pubs.acs.org/doi/suppl/10.1021/acs.chemrestox.6b00347)

November 2017 Comment on Agency Response

The Panel observed that the incorporation of uncertainty estimates via a bootstrap resampling approach is particularly commendable. The fact that the analysis incorporated several assays does support the reproducibility of the results in that it wasn’t influenced by the sensitivity of one particular assay. However, more details are needed to understand whether the confidence intervals constructed using bootstrap resampling correctly account for all different types of uncertainties. From the description of the bootstrap resampling procedure, it is unclear how the resampling is done and whether the entire workflow procedure (e.g. model fitting to estimate the R values, curve fitting, etc) was applied. In particular, were the data relative to a chemical resampled within assay and concentrations each time, or was the data relative to a chemical resampled without doing the resampling within assay-concentration pair?

Although the AR pathway model results for the reference chemicals are quite impressive, the comparison with the results obtained by the Tier I binding assay indicate disagreement between the Tier I binding assay and the proposed model. The Agency has investigated the reasons for the discordance in results, and while the justification that the Agency has provided is reasonable, it raises the question as to whether this is a result of inadequate model validation. The AR pathway model is in some sense “trained” using the reference chemicals in mind (see below), and thus the impressive performance of the model on the reference chemicals could be considered a sort of in-sample validation or lack of independent test samples, while the application and results obtained on the additional set of chemicals with Tier 1 AR binding assay data and ICCVAM data can be considered as an out-of-sample validation.

The Agency indicated that the only model fitting in their approach is the initial concentration-response fitting to a constant, Hill equation, or constrained gain-loss model. This is then followed by integrating AUC scores across all assays and filtering out those that might be due to cytotoxicity or other interference. An overall AUC score for a chemical is considered significant if it’s over 0.1. However, the Agency does appear to fit parameters for R_j values that minimize the difference between the predicted assay values and the measured values (see White Paper section 2.2.3 Mathematical Representation of the Pathway Model: "The model seeks a set of Rj 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"). If a model "seeks values" then a model is generally being fit to something. The Agency appears to find these values using least squares minimization or a variation on a linear regression. Because they are fitting the Rj values to the data (A_i^{meas}), the Panel noted that there appears to be potential for model overfitting. Since no data are held out during model fitting for proper validation, the model could be over-fitted, resulting in bias toward the data set analyzed. A proper validation would require something similar to a cross-validation using a training dataset and a separate testing dataset that was not used to estimate the Rj values to. As a result, the model does appear to be trained on the chemical data set used.
The Panel noted that to truly determine the performance of the AR model, there is a need to examine performance on a set of chemicals that are not in the "training set."

When determining AR activity of different chemicals, the Agency used performance based criteria in demonstrating the reproducibility of the AR model. This is based on the idea that the assays and model are too sophisticated to be run in a naïve laboratory thereby precluding the testing of the model by independent groups. Since the reproducibility of the model is the principle question of concern, the use of performance based criteria may not be justified in this case. Validation of the *in vitro* assays were not the question asked but rather validation of the model that integrates and provides a measure of AR activation. Clearly, all data, procedures and processes for the mathematical model are available to even naïve labs. Therefore, the Panel noted that a more suitable validation approach would be to provide an independent group a description of the mathematical functions needed for construction of the model R-code, *in vitro* assay test data to which R_j values were fit in the Agency model, and independent testing of model reproducibility using data to which R_j values were not fit.

The Panel observed that one statistical concern with the proposed model is the number of preprocessing steps involved in the analysis pipeline, which makes an inference procedure more prone to error and uncertainty, and may result in varying performance due solely to modeling decisions made throughout the pipeline. Future iterations of the analysis approach may incorporate other approaches, such as the deep learning approach offered by Burgoon (2017). It is noteworthy that the development of this approach was made possible by Agency transparency in making assay data publicly available. The Panel recommended that the Agency should continue to strive for transparency in documenting all steps of the analysis pipeline, and describing in detail the modeling choices made at each step.

December 2014 SAP 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.

November 2017 Comment on Agency Response

The Panel feels that the Agency's response regarding alternative AR-related assays, non-classical mechanisms of activation, and metabolism of chemicals is appropriate. Since the Agency is currently focused on assessing whether or not the AR model is a suitable alternative to the AR binding assay, the Agency did not evaluate non-competitive mechanisms of antagonism. Both the AR binding assay and the proposed AR model do not take into account: bioactivation, mechanisms that do not follow the classical genomic/nuclear receptor pathway, or cause off-target effects. The absence of assays to measure non-competitive mechanisms could render the model less useful and could significantly impact the ability of the model to correctly identify chemicals that act in non-classical ways. The Agency is currently developing assays to replicate known *in vivo* activity with *in vitro* assays and the investigation of non-classical/non-genomic mechanisms of AR pathway activation are scheduled for future studies.

The Panel noted that to better predict effects of chemicals, the AR bioactivity battery should include methods to assess the potential effects of chemicals, as well as their metabolites formed by enzymatic conversion in biological systems. *In vitro* assays may not always predict *in vivo* outcomes due to their limited coverage of metabolic processes present in a whole organism. This is especially important for compounds that undergo bioactivation, as these chemicals can produce false negatives when tested in assays without metabolic activity. This limitation of the Tier 1 binding assays should not be incorporated into the HT models. The Agency recognized in the White Paper the importance of metabolically active cell lines and is considering *in silico* approaches plus additional assays with metabolic competency to address these issues.

The Panel noted that other potential areas that the Agency can investigate for non-classical/non-genomic mechanisms of AR pathway activation include: Activation of 2nd messenger pathways including ERK, Akt and MAPK that are identified in a number of cell lines (e.g., osteoblasts and osteocytes). Indirect gene trans-repression can also occur, by the AR binding and sequestering transcription factors such as activator protein-1 (AP-1) that are normally required to upregulate target gene expression (e.g. Ngfr (Kallio et al 1995) and Mmp-13 (Schneikert et al 1996)), in the absence of the AR binding to DNA.

One panel member noted that the ability of the AR model to identify chemicals that exert action outside of the canonical AR-binding AOP is essential for the Agency's future goals of the EDSP focus on replacing the *in vivo* Hershberger assay. This would be facilitated by expanding the chemical library to include non-genomic androgen antagonists.

December 2014 SAP 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., 2017). The R-code for the analysis is supplied (Watt, 2016). Extensive efforts were made in the White Paper to be comprehensive in supplying information in order to be completely transparent. (Supplemental documents are located in the public e-docket, Docket No. EPA-HQ-OPP-2017-0214, accessible through the docket portal: http://www.regulations.gov)

November 2017 Comment on Agency Response

The Panel found that overall, the Agency made details and results of the model available to increase transparency. The Agency made significant effort to publish their work in the peer-reviewed literature, as illustrated by the citations presented in the White Paper. The Agency made all raw and processed data as well as computer codes publicly available (http://epa.gov/ncct/toxcast/data.html). Assay descriptions, data and analysis files including R code are available as supplementary materials to the White Paper. While providing code and data is a step in the right direction, the Panel recommends presenting a detailed description of the algorithm used, particularly for those who may not be able to interpret R code.

TOPIC: Steroidogenesis Pathway Activity

A number of environmental chemicals are shown to interfere with the biosynthesis of estrogens (*e.g.*, estradiol) and androgens (*e.g.*, testosterone), and the EDSP Tier 1 screening battery includes several *in vitro* and *in vivo* assays designed to detect compounds that may affect steroid synthesis. One *in vitro* assay in the Tier 1 EDSP battery, the Steroidogenesis Assay (H295R cell-based steroidogenesis assay, OCSPP 890.1550/ OECD TG 456) utilizes human adrenocortical carcinoma cells as a model of adrenal, ovarian, and testicular steroidogenic function and is used currently to screen for potential perturbations in the steroid synthesis of estrogens and androgens. Testosterone (T) and estradiol (E2) levels are measured in the cell culture medium of chemically-exposed H295R cells, and hormone concentrations in the medium serve as indicators of steroidogenesis disruption.

The Agency developed a high-throughput (HT) H295R cell-based assay (Karmaus, *et al.*, 2016) that uses high-performance liquid chromatography followed by tandem mass spectrometry. A comparison of the low-throughput (LT) and HT H295R assays for detecting the disruption of synthesis of T and E2 is presented. This comparison enabled evaluation of the utility of the HT H295R assay as an alternative to the LT Tier 1 H295R assay.

As an expanded component of the HT H295R assay, data from 9 additional steroid hormones (including progestagens, glucocorticoids, androgens, and estrogens) were collected (see Section 3 of the White Paper). The data for all 11 hormones were integrated using a novel statistical approach to quantify the overall impact of the chemical on the steroidogenesis pathway. In consideration of both the comparison of the LT and HT H295R assays and the new statistical approach to assess the impact on the steroidogenesis pathway, please address the following three charge questions:

Question 2: Based on the comparison of the performance of the HT H295R assay with the LT H295R assay, and the effects of reference chemicals on the synthesis of T and E2 levels only, please comment on the suitability of the HT H295R assay as an alternative to the LT H295R assay. See Sections 3.3 and 3.4.

Response:

The Panel observed that the HT H295R steroidogenesis assay, for the measurement of E2 and T only, is based on generally well-conceived modifications of the existing and validated H295R cell based steroidogenesis assay (OCSPP 890.1550/OECD TG 456). Conceptually, the modifications of the LT H295R assay, to facilitate analysis in a 96-well cell culture format, are logical and scientifically sound. The HT H295R steroidogenesis assay benefits from a number of strengths. For example, the incorporation of a forskolin

pretreatment to increase baseline steroid production in the assay is generally considered a positive modification. As indicated in Table 3-1 of the White Paper, in comparing OECD TG 456 versus HT H295R the latter assay differed in that there is a 48-hour pre-stimulation with forskolin. However, it is possible that sensitivity of the assays might be decreased by pre-stimulation with forskolin. The Panel believes that specific demonstration of the impacts of pre-stimulation with forskolin on the dynamic range, sensitivity and overall assay performance are necessary. The question is whether or not forskolin-induced upregulation of basal steroid biosynthesis and the resulting increase of steroid concentration in media affect sensitivity, dynamic range, and the resulting ability of the HT assay to detect changes in activity (especially for chemicals that stimulate rather than inhibit). The specific impacts of pre-stimulation with forskolin require further evaluation and optimization. The Agency is likely aware of this given the White Paper statement on page 104: "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." The Panel recommends additional efforts in evaluating the effects pre-stimulation with forskolin, assay validation and optimization.

The Panel held specific concerns related to a lack of demonstrated sensitivity and reproducibility that limit the suitability of the HT H295R steroidogenesis assay as an alternative or replacement for the LT H295R cell-based assay. In some cases, there is a lack of quantitative data available in the Agency's White Paper or in the presentations given to the Panel to allow assessment for suitability of the HT assay as an alternative to the LT assay. The Panel believes that a quantitative comparison of the relative potencies of the positive controls in each assay is needed. Several Panel members expressed substantial concern that the HT assay is less sensitive than the LT assay. A direct comparison between the low and high throughput steroidogenesis assays that determines the concentration of E2 and T generated and the relative potency for positive control chemicals is needed to assess the value of using the HT assay as an alternative to the LT assay. Supplemental Table 10 gives some measure of potency/sensitivity for the OECD LT assay (i.e. LOEC). However, for the HT assay, there is no analysis of relative potency of positive controls as only the maximum concentration tested is listed. This does not allow for any quantitative comparison between the assays. The Panel recommended that the Agency calculate the IC50s and AC50s for positive control chemicals for each of the 11 hormones in the HT assay for comparison. Additionally, the lowest IC50 of the 11 hormones (the most sensitive endpoint) should be assessed for each control compound and compared with the LT assay.

Retaining an assessment of cell viability as part of the HT H295R assay is also considered a strength. However, the reduction of the cell viability cutoff (70 vs 80%) raises significant concern with the Panel. The Panel appreciates that 70% viability is presented as the statistical limitation of cell viability as used in conjunction with the HT

steroidogenic assay. Nevertheless, some Panel members consider the deviation from the guideline standard of 80% viability is poorly justified and problematic for interpretation of assay results. These members believe that a 30% loss of viability would be biologically impactful, and would result in negative impacts on assay performance and that those effects would obfuscate some interpretations. In cell-based inhibition assays, reduced viability will artificially inflate the number of chemicals flagged as "hits." Specifically, the alteration of mitochondrial functions resulting from decreased viability, rather than direct impacts on steroid biosynthetic enzymes, could result in significant alterations in steroid levels detected, such effects are expected to increase Type 1 error. The Panel believes that additional justification for the appropriateness of the 70% viability cut-off is necessary before this approach can be broadly applied to chemical screening. Examples of additional evidence necessary could include: 1) Evaluating the impact on findings if the viability cutoff were set to 80% as in the LT H295R steroidogenesis assay, and 2) investigating the utility of incorporating an appropriate cytotoxicity z-score into analysis. Additionally, the use of alternative cell viability assays—those that avoid the use of mitochondrial reductase function—that are less variable than the MTT viability assay is recommended. The Panel suggests that further investigation of uncoupling the cell viability assessment from mitochondrial function is necessary. While the MTT and related assays are reliable to a degree, that assay was considered especially problematic for use with the steroidogenesis assay because many of the key (initial) steps in steroid biosynthesis occur in the mitochondria and require an intact mitochondrial membrane potential. The Panel stresses that even small decreases of ATP levels have large impacts on steroid biosynthesis, and that those impacts are independent of the steroidogenic enzymes being evaluated by the H295R steroidogenesis assay.

For the set of reference chemicals used in the inter-lab analysis of the OECD guideline H295R steroidogenesis assay (Hecker et al, 2011), the HT H295R assay performed with less sensitivity. For detection of T related endpoints, one Panel member noted the reported sensitivities of 0.55, 0.67 or 0.75 (Fig 3.8 of the White Paper) are unacceptable from a public health protection standpoint. From this point of view, the failure of the HT H295R assay to accurately identify the E2 and T production disrupting reference chemicals rendered the assay in its current form inadequate for protecting the health of populations. As a result, in its current state, the data and the information presented in the White Paper indicate that the performance of the HT H295R assay does not meet the requirements of assays as set forth in the final report of the Endocrine Disruptor Screening and Testing Advisory Committee. The report specifies Tier 1 assays must "be more 'sensitive' than they are 'specific,' meaning that they should have as their primary objective the minimization of false negative or (Type II) errors, while permitting an as-of-yet undetermined, but acceptable, level of false positive or (Type I) errors."

There are some additional concerns voiced by the Panel related to reproducibility and reliability of the HT H295R assay and the approach used for comparative analysis. It is not readily apparent if the comparison of the HT results to the performance of the LT H295R assay, in an intra-laboratory performance assessment across seven different laboratories world-wide, is the most appropriate metric for evaluating the performance of the HT Assay. The findings of performance for the OECD guideline H295R steroidogenesis assay presented in Hecker et al (2011) is an evaluation of replication of results across 7 different international laboratories, the information presented for the Agency's HT assay is data from a single laboratory, which indicates an apparent difficulty in replication across different assay blocks (Karmaus et al, 2016). Additional studies demonstrating transportability and replication of the HT H295R assay and results across biological replicates and across different laboratories is needed. Overall, the Panel believes that it is not possible to interpret the reliability of the assay from run-to-run without more information about the consistency of the results across replicates (majority of chemicals were run with only 1 biological replicate), and no rigorous evaluation was performed to test assay reproducibility. The Panel recommends establishing the reliability of the assay/analysis from day-to-day (across blocks). Concern of assay reliability extended to the ability to replicate assay results for future testing and in different labs. It would have been useful for example, for the Agency to report the independent retesting of chemicals tested in the Karmaus et al (2016) to assess replicability across time.

Regarding the presented comparative analysis, while it is indicated that 16% of the screened chemicals were analyzed in more than 1 "plate-block," the Agency does not indicate how many times the individual reference chemicals were analyzed. It is important that the Agency demonstrate that the assay was performed on more than a single biological replicate and that the presented analysis was robust and meaningful. Most test chemicals analyzed in the HT R295R assay were examined only once as duplicate technical replicates in a single block, but one is left to assume that this is not the case for each of the reference chemicals. Because the reference chemicals were analyzed by ANOVA and Dunnett's test for comparison with the LT assay results, one is left to assume that more than 1 biological replicate was analyzed using these statistical methods, however the supporting data for this is not provided. The lack of specific information on the number of biological replicates makes it difficult to compare the reproducibility of the results for the reference chemicals or whether the statistical approach used for comparison is appropriate.

Additional concerns expressed by the Panel include an inability to fully assess the appropriateness of the pre-screening approach. Karmaus et al (2016) reported that over 50% of the samples pulled randomly from the non-concentration response selected batches produced an effect on at least one hormone. This was considered by the Panel to be potential evidence that the pre-screening approach might be missing endocrine active chemicals and was resulting in an unacceptable level for Type II (false negative)—even

from an EDC prescreening perspective. The Panel recognized that alternative screening assays should be fit-for purpose, that is, be high quality, and have rigorous and reproducible methodology, but also be a good match with available resources. The importance of recognizing that the goal of a screen was to cast a wide net with an eye on setting priorities, not exoneration by a lack of testing, was emphasized. Though resource constraints are noted, the Panel recommended that the Agency enhance justification that tests discussed are better than those currently used.

A Panel member noted the White Paper includes additional ADME studies, aromatase assays, and other studies in the prioritization process and asked if these would be more appropriate for follow-up analysis after completing initial screening.

Analysis limitations exist in that only effects observed for a given hormone—when two consecutive concentrations demonstrated significant effects—were considered meaningful. Panel members pointed out that the two concentrations may either be too broadly spaced or too closely spaced for this to be meaningful. For example, one might observe only two very high concentrations showing activity, which would result in analysis bias at high concentrations. Additionally, this approach will be limited for detection of non-monotonic concentration responses as it may result in an effective concentration interval that is too far apart to detect non-monotonic effects. The Agency should be confident and demonstrate that this approach does not undermine the purpose of the multi-concentration approach to capture dose-response.

One Panel member indicated that the inadequacy of the HT assay is demonstrated by its inability to adequately characterize the known effects of phthalates – chemicals known to interrupt the steroidogenesis pathway. A recent report by the National Academies of Sciences (NAS) that looks at the application of systematic review for evaluating low-dose toxicity from endocrine active chemicals used these effects of phthalates as one of the case studies (National Academies of Sciences, 2017). The report performed systematic reviews for a number of phthalates—chemicals that at least in part, act via disruption of testosterone synthesis. The NAS committee found that the current HT assays that rely on a human adrenal cell line (e.g., the H295R assay) are not sufficient for identifying phthalates like DEHP (a chemical the committee found evidence to support calling the chemical a presumed reproductive toxicant in humans) because adrenal steroidogenesis *in vivo* is not affected by phthalate exposure via the same mechanism. However, it was noted by one panel member that the phthalates are problematic as all evidence of hormone effects are seen in the rat (but not mouse) and substantial evidence exists to indicate that humans may not have the same hormone effects (Spade et al., 2014). Nonetheless, the use of an adrenal cell to evaluate hormones that are produced in the ovary or testes *in vivo* requires additional validation of the biological relevance to the intact human.

In summary, it was the general feeling of the Panel that the performance of the HT H295R assay in its current form has some clear benefits, but additional performance optimization along with transparent demonstration of assay reproducibility, reliability, and portability are necessary before it is a suitable alternative for the LT H295R assay.

Question 3: Please comment on the strengths and limitations of integrating multiple hormone responses beyond T and E2 (*i.e.* 11 hormones vs 2 hormones) in a pathway-based analysis of the HT H295R assay. Please comment on the suitability of this HT H295R pathway model (using 11 hormones) to serve as an alternative to the LT H295R assay. See Section 3.7.2.

Response:

In general, the Agency clearly describes the high-throughput (HT) assay. Overall, Panel members believe that the HT assay offers significant advantages compared to the low throughput (LT) assay. The multiple hormones measured, in conjunction with the statistical metric (mean Mahalanobis Distance; mMD) enables the incorporation of additional information. The comparison between LT and HT assays indicates a correlation in the accuracy of the assays. The Panel found that the HT assay provides improved accuracy, additional information, and improved sensitivity. As such it has the potential to be effectively used for prioritization.

The HT assay monitors activity of several hormones encompassing a simplifying network of cross-regulated elements of the steroidogenesis pathway. As such it enables monitoring of an integrated response as opposed to isolated, individual elements. The analytical system offers high sensitivity and because multiple components of a pathway are monitored at once, the ability to measure coordinated responses is expected to increase sensitivity. The 11 measured hormones represent 4 distinct classes, adding significant diversity to the assay measurements. The ability to measure multiple elements has the potential to not only improve accuracy of predictions, but also may provide additional, mechanistic, information. The ability to measure multiple hormones, and the complex patterns of hormone concentration that emerge in response to exposure to a chemical, demonstrates a more complex picture than one using two hormones, and improves characterization. The two assays (high and low throughput) use the same cell lines, The HT assay includes both hormones (T and E2) measured in the LT assay plus additional components. This allows direct comparison with the LT assay. The HT assay performed comparably to the LT assay in terms of quantifying E2 and T effects, confirming the efficacy of the measurements. The “revised” confusion matrix elements (Figure 3-8 in the White Paper) indicate improved performance in characterizing E2 and T using the HT assay compared to the LT assay. The development of the modified Mahalanobis distance metric enabled the integration of multiple features into a single metric. In the long run, the ability to monitor responses at a pathway level will provide critical information

towards the development of dynamic, quantitative systems toxicology models.

The Panel members identified a number of issues that require further examination. To some panel members, the HT assay, as implemented, appears to lose some of the advantages of the LT system, such as validation across multiple laboratories and fewer technical and biological replicates (3 in LT, only 1 in HT). To assess the suitability of the HT assay as an alternative to the LT assay, a quantitative comparison of the relative potencies of the positive controls of each assay must be conducted.

Several Panel members expressed concerns that the HT assay is less sensitive than the LT assay. A direct comparison of the relative potency of positive control chemicals on the concentration of E2 and T between the low and high throughput steroidogenesis assays is needed to assess the value of using the HT assay as an alternative to the LT assay. Supplemental Table 10 gives some measure of potency/sensitivity for the OECD LT assay (i.e. LOEC), however, for the HT assay, there is no analysis of relative potency of positive controls as only the maximum concentration tested is listed, so this does not allow for a quantitative comparison between the assays.

Panel members suggested that IC50s and AC50s be calculated for positive control chemicals for their effects on each of the 11 hormones in the HT assay and use the lowest IC50-most sensitive endpoint for comparison with the LT assay. The confusion matrices indicated strong correlation between the LT and HT assessment. However, the confusion matrices are based on analysis of individual hormones. It was unclear to some Panel members whether the same trends will hold true with the aggregate Mahalanobis score incorporating the lack of response in other hormones for estrogen/androgen specific chemicals. Even though the Venn diagrams (See White Paper Figure 3-4) point to complex hormone release patterns, the interpretation of these results was not clear to all Panel members.

The Agency notes that at least 400 chemicals impacted only 1 or 2 hormones, and about 300 chemicals hit 3- 5 hormones. That 307 chemicals hit all 4 pathways (based on the Venn diagram results) is in line with 300 chemicals that hit 6 or more hormones. These findings could be used to suggest that less than a third of the “positive” chemicals are promiscuous. While measuring 11 hormones, in the context of a pathway, panel members suggest that it is valuable to provide stronger support about the clear need for added hormone measurements. This includes a comparison of how many chemicals would be called a “hit” on androgen/estrogens alone versus those called a “hit” based on the combined pathway score. In other words, will significantly more chemicals identified when the additional hormones are measured?

While the Agency suggested that most chemicals affect all 4-hormone classes, this

circumvents the fact that at least 1/3 of the tested chemicals only affected 1 or 2 hormones. The Agency used a cutoff criterion that a chemical would only be considered active if it affected 3 or more hormones.

Panel members suggested that the application of the cutoff of at least 3 hormones being changed may not be protective or conservative in risk assessment. Panel members appreciate that the decision was made based on R&D resources during assay development. However, the lack of analysis, or interpretation, of the 1-2 hormone data is concerning when used in a chemical prioritization context and some panel members voiced strong opposition to continuing this practice in an EDSP implementation.

Some panel members expressed concern regarding the use of the mean Mahalanobis distance (mMD) metric for identifying potential endocrine disruptors. While the approach is a creative an intriguing approach to deal with a multi-factorial biological problem, additional evaluations are needed to show that the metric would be appropriately sensitive to chemicals that are not "promiscuous" (i.e., affect steroidogenesis broadly), but rather affect only 1-2 hormones.

The Panel suggested that the Mahalanobis metric should also be assessed for the weakly active chemicals that hit only 1 or 2 hormones. Without looking into these data, the analysis could be biased for compounds that work on the upstream nodes in the pathway and against compounds that affect the terminal nodes. It could be that the Mahalanobis score would work as well for the chemicals that hit only 1 or 2 hormones. If these analyses have been done, they should be added to the record. If they have not been done, the Panel suggested that they should be conducted before implementing a path forward with this assay and combined pathway metric for EDSP.

The Panel noted that, even though multiple hormones are measured, there is a lack of reference for the additional elements of the steroidogenesis pathway measured by the HT assay. For prioritization purposes, the pathway scores are likely appropriate. However, classification of "progestogen disruptor" or "corticosteroid disruptor" based on an assay that has no positive or negative controls for these pathways could be questionable. Broader limitations of the HT H295R assay include the inability to measure metabolic effects and not DMSO-soluble chemicals. However, the Panel realizes that this is a broadly applicable issue for all *in vitro* systems, including the LT H295R assay.

The Panel noted that assay cell viability requirement in HT was reduced (from 80% in LT to 70% in HT). Although the White Paper indicated that anything above 70% would be difficult to discern statistically with the MTT viability assay, several members of the Panel felt that this is an issue worthy of further evaluation. This is because as little as 10% loss of ATP can be directly correlated with a concomitant drop in hormone production even for negative controls as much of the steroid metabolism occurs in the

mitochondria and ATP measures mitochondrial health. These Panel members believe that, although 70% loss is the statistical limitation of the assay, biologically, 30% loss of viability is high and likely affects results. These members advise providing additional justification of this limitation and assess how the results change if the viability cutoff were 80% as in the original assay. Alternatively, Panel members asked if a different (less noisy) measure of viability could be used? What if the developed z-score was used–how would that affect hit calls? Finally, some Panel members were concerned that increasing the dimensionality of the feature vector (11 instead of 2) increases the information content, yet makes interpretation likely more complicated. The White Paper noted that the set of chemicals was reduced to focus only on chemicals inducing changes in 3 hormones or more. However, the development of the mMD metric greatly reduced the burden of representing and interpreting the data.

In summary, Panel members believe the HT-H259R is a scientifically sound potential alternative to the LT H295R. However, additional analyses to support assay conditions (viability cutoff), and analysis methods (multiple hormone effect cutoff, Mahalanobis score for chemicals that weakly affect 1-2 hormones) before implementation in EDSP would be advisable. Furthermore, recommendations should be developed for chemicals affecting corticosteroid/progestogen pathways where there is no positive/negative control data. Panel members felt that this would be important in the context of prioritization vs. hazard identification risk communication with the public.

Question 4: The work herein presents a novel statistical integration of multiple hormone responses indicative of steroid biosynthesis in the HT H295R assay. A summary statistical metric, the maximum mean Mahalanobis distance (maxmMd), has been suggested as a tool for use in prioritization of chemicals. In addition to the use of the maxmMd to indicate the magnitude of potential effects on the steroid biosynthesis pathway expressed in H295R cells, an examination of the hormone responses that contribute to the maxmMd may provide valuable biological information to inform the weight-of-evidence evaluations performed for chemicals subjected to EDSP Tier 1 evaluation.

Please comment on the strengths and limitations of using the maxmMd and the pattern of steroid hormone responses in the HT H295R assay for chemical prioritization and weight-of-evidence applications. See Sections 3.2.4, 3.3.2, and 3.7.2.

Response:

The Panel commends the effort of the Agency to consider multiple hormone responses simultaneously to obtain an integrated and comprehensive indication of the magnitude of the potential effect of a chemical on the steroid biosynthesis pathway. In reviewing the proposed maximum mean Mahalanobis distance approach as a tool for chemicals

prioritization, the Panel has identified the following strengths and limitations.

Strengths:

The proposed approach for assessing steroid biosynthesis generates multi-dimensional data (precisely, data on 11 hormone responses) for each chemical at various concentrations. The Panel recognizes that the maximum mean Mahalanobis distance is a way to summarize these multi-dimensional data into a single scalar quantity using a metric that has close ties to quantities typically used in statistics, such as the Hotelling T^2 test statistic used to test whether there are significant differences between two groups with respect to multidimensional data (see also White Paper page 78). In a nutshell, the mean Mahalonobis distance of a multidimensional observation from the center of a multivariate normal distribution is the multi-dimensional equivalent of the z-score of a univariate observation that is normally-distributed. Thus, like the z-score, the mean Mahalanobis distance can be used to flag outliers. The Panel recognizes that an advantage of using the mean Mahalanobis distance is that it allows combining measurements on multiple hormone responses into a single summary measure, while accounting for the second moment of the sampling distribution, that is, while accounting for the variability of each individual hormone response measurements as well as the correlation among the various measurements. The Panel believed that working with such a summary metric would allow controlling for highly variable hormone responses and avoids incurring problems related to multiple testing.

The Panel highlighted that while the mean Mahalanobis distance might be most appropriate for multivariate normal data, this does not constitute a major concern within the considered application for two main reasons: (i) the mean Mahalanobis distance is applied on the log hormone response measurements, which are more likely to not display characteristics such as skeweness and long-right tailed distribution; and (ii) analogous assays that measure hormone responses have been shown to generate data that, when transformed via a log transformation, appear to be approximately normally distributed (see Zhang, Chung and Oldenburg (1999)) .

The Panel further believed that the proposed framework for prioritization of chemicals based on the maximum mean Mahalanobis distance computed over multiple concentrations, is a conservative approach for flagging a chemical as an outlier with respect to controls.

Limitations:

The Panel found difficulty in understanding what type of effect a chemical should have on the steroid biosynthesis pathway to be flagged by the proposed maximum mean Mahalanobis distance approach.

More specifically, the Panel believed that the maximum mean Mahalanobis distance metric may result in prioritizing a chemical that has relatively small absolute differences from the control with respect to any single hormone measurement, but unusual combinations of hormone responses with respect to the sampling distribution of the residuals. An example could be a chemical for which a hormone response measurement is above the mean and a second hormone response is below the mean when the two hormone responses are instead expected to be positively correlated. Under the proposed framework, the Panel believed that this chemical would be flagged whereas a chemical which displays very large absolute deviations from the control but small deviations when adjusted for the "typical" correlation structure would not.

In summary, the Panel hypothesized that the proposed approach would tend to: a) flag mostly chemicals that deviate from the expected relationships between hormone responses; and b) not allow prioritizing chemicals that display absolute differences from controls, regardless of the sampling distribution of the residuals. As there is not much clarity around these two points, the Panel believed that it is important to clarify those issues, in particular determining whether the hypotheses of the Panel are indeed correct.

The Panel is concerned that the critical values used, and the Type I error rate controlled for, might not be appropriate. Specifically, the White Paper indicates that the proposed approach to flag chemicals uses critical values that were determined following the method developed by Nakamura and Imada (2005). The latter requires equal sample sizes across comparisons and known covariance matrix. Neither of these conditions are satisfied in the analyses presented in the White Paper. Thus, as also mentioned in the White Paper, the Panel believed that nominal Type I error rates will not be achieved. The White Paper states that the Type I error rate would be "approximate" under the proposed approach, however, without any numerical result to support this statement, the Panel found it hard to believe the accuracy and appropriateness of the approximation. Hence the Panel suggested that simulation experiments be carried out. Specifically, the Panel recommended to:

1. Perform extensive simulation studies that evaluate the Type I error rate of the proposed method using the data in the White Paper as a guideline for the simulation settings.

2. Cite (in the White Paper and future documentation) any simulation studies that have already been performed, as such studies will be vital if the proposed approach is going to be a standard methodology going forward.

3. Provide a rationale for the use of a 1% Type I error rate instead of a more conventional 5% Type I error rate; in particular, clarify whether such choice was dictated more by a concern that the Type I error rate might be inflated. Simulation studies can help determine whether this is an adequate correction or if it is too conservative.

The Panel raised some concerns regarding the appropriateness of the estimated covariance matrix used to derive the mean Mahalanobis distance(s). As mentioned in previous points, the covariance matrix plays an important role in the derivation of the mean Mahalanobis distance. For example, an inflated estimate of the covariance matrix will tend to produce mean Mahalanobis distance values that are smaller than they should be, with consequent inflation of the Type II error rate.

The Panel suggested that a more thorough investigation of the behavior and appropriateness of the estimated covariance matrix be carried out. From the description on page 78 of the White paper, it appears that all the hormone response measurements that were not flagged or removed were used to estimate the sample covariance matrix employed in the mean Mahalanobis distance, regardless of: a) whether the hormone response measurements refer to a control chemical or not, b) the mode of action of a chemical, and c) the concentration of the chemical. The Panel believes that it might be plausible, from a biological point of view, that correlation and variability in the 11 hormone response measurements are different depending on the type of chemical (control vs chemical tested), and the concentration level.

The Panel also raised the following minor comments regarding the White Paper:

1. It is unclear how values below the limit of detection were handled.

    a. Two hormones were excluded from the analyses described in the White paper because of this issue. Although the Panel believed that values below the detection limit might have been identified using something standard, like ½ the Detection Limit, the Panel believed that for clarity, the White paper should clearly state how values such as these are handled.

2. 'Critical value' and 'critical limit' seemed to be used interchangeably. This is confusing; a more homogeneous nomenclature (possibly, critical value) should be used.

3. Figure 3-10 appears to be a box plot of maxmMd, not adjusted mMd, since all values are positive. Open symbols represent negative *adjusted* maxMd values.

4. On page 105, it is not clear how the "NA" yielded adjusted maxmMd.

5. On page 111, the confidence interval formula and example calculation should be provided for clarity and completeness.

In summary, although the maximum mean Mahalanobis distance might not be the optimal statistical approach to integrate multiple hormone responses due to some of the limitations mentioned above or due to the fact that it does not take into account the biological pathways, the Panel recognized that this a step in the right direction in the effort of developing a framework to assess chemicals' potential to effect steroidgenesis.

The Panel also recommended that a distance metric such as Tukey's halfspace depth (see Tukey 1975 for the conceptual overview of the metric, Struyf and Rousseeuw (2000), and the R package 'depth' (Genest et al., 2017) for computational implementations), be investigated due to its appealing characteristic as a nonparametric method to rank-ordering multivariate observations. In addition, the Panel recommended that methods such as the one proposed in the paper by Ovacik and Androulakis (2013), be considered for comparison of multidimensional vectors that represent biological pathways or networks. More specifically, the Panel recommended that efforts be placed into revising the maximum mean Mahalanobis distance approach to take into account the biological pathway, thus developing a metric that measures distance between networks rather than simply distance between multidimensional vectors.

TOPIC: Thyroid Conceptual Framework

Over the last several years, the Agency significantly expanded research efforts on thyroid related HT assays, and the design of EDSP's framework for screening of potential thyroid hormone disruptors is in its early stages. Unlike screening for modulators of estrogen and androgen receptors, which captures much of the estrogenic and androgenic bioactivities of xenobiotics; chemicals that perturb thyroid homeostasis may act via one or more heterogeneous targets in the thyroid adverse outcome pathway (AOP) network (see Figure 4-1 in the White Paper). Thus, a larger set of assay targets, beyond just hormone receptors/signaling, should be considered to screen for potential disruption of thyroid hormone-related bioactivity. Currently, a number of assays are available, with several more in development; however, assays do not yet exist to interrogate every molecular initiating event (MIE) in the thyroid AOP network. Also, in contrast to the estrogen and androgen receptor pathway models, it is unlikely that multiple orthogonal assays for each target (*i.e.*, MIE or key event (KE)) will be available in the near future.

Section 4 of the White Paper outlines a thyroid AOP network (Section 4.2) and presents the current status for high-throughput assays (Section 4.3). The thyroid AOP network aims to serve as a foundation for a future EDSP strategy or framework to identify and prioritize potential thyroid-disrupting chemicals. The Agency seeks insights from the SAP on the direction of its proposed approach.

Question 5: Please refer to White Paper Section 4.2. EPA has identified AOPs for thyroid hormone disruption related to potential xenobiotic-induced alterations of thyroid homeostasis. Please comment on the completeness of the MIEs (Table 4-1), KEs, and adverse outcomes within the thyroid AOP network (Figure 4-1). Also, please provide information on any missing pathways, adverse outcomes, or other AOP-related information (*e.g.* MIEs or KEs) critical for capturing the complexity of systems biology controlled by thyroid hormones.

Response:

Overall, the Agency presented a largely complete set of molecular initiating events (MIEs) and key events (KEs) in Table 4-1 of the White Paper. However, the Panel recommended adding a new column to Table 4-1 to include adverse outcomes because those listed in Figure 4-1 are not sufficiently specific. It is important that the concept that a "…comprehensive pathway-based approach, that incorporates screening for potential interaction with multiple MIEs, is needed to effectively screen for TDCs" is central to the Agency's strategy for the thyroid. Considering this, success is dependent in part on the Agency's approach and in part on the biology of the thyroid system. For example, the Agency uses the example of thyroid homone receptor (TR) to illustrate their point. The observation is that TR activity *in vitro* fails to predict the vast majority of thyroid hormone related findings in *in vivo* studies and the interpretation is that the ligand binding domain of the TR is too restricted. But the *in vivo* findings include: a) serum T_4, b) serum thyrotropin (TSH), c) thyroid weight and histopathology. Although there is ample evidence to support the conclusion that the thyroid hormone receptor ligand binding domain is more restricted than that of the ER, it is also true that TRβ2 selectively mediates negative feedback on the hypothalamus and pituitary. In contrast, TRα1 does not affect serum T_4 in rodents or in humans. Therefore, only chemicals that interact with TRβ2 would be expected to influence serum T_4 *in vivo.* In contrast a TRα1-binding chemical would not influence serum T_4 in rodents or humans. Because serum T_4 is the primary *in vivo* endpoint to which ToxCast/Tox21 data are being compared, it is important to align the molecular initiating events (MIEs) and key events (KEs) with adverse outcomes that are consistent with the pathway.

The point here is that the adverse outcome pathway (AOP) needs to link the specific MIE to known adverse effects and through known KEs and KE relationships (KERs) that will be used to identify thyroid disrupting chemicals. In many cases, these may not be known and this represents a significant challenge for the Agency. This discussion bears directly on the design of Table 4-1 and Figure 4-1. Specifically, the MIE's in Table 4-1 need to be more specific, with a separate row for each and creating a new column with "Adverse Outcome" linked to that MIE. Thus, the Panel recommended Table 4-1 should be revised as follows:

1. Add a final column that includes the adverse outcome that would be predicted to result from interference with the MIE identified in that row. While this is somewhat covered in Figure 4-1, to highlight this here would provide an opportunity to reference the scientific evidence for these adverse effects, and would also highlight what we know and what we don't know. This is a complex system and the Agency has made great strides in organizing their work effectively. Articulating what is known/unknown in terms of the adverse effect resulting from specific MIEs would provide a roadmap for the Agency in future work.

2. Use a single row for each MIE (i.e., protein target) rather than lump them into classes. The example of TRs is useful, because the different TRs mediate different actions and therefore different adverse outcomes.

3. Since the recognition of cellular transport proteins by Grueters and others (e.g., MCT8 (Friesema et al. 2004)), the serum binding proteins have been called "distributor" proteins. The reason for this is that in early work, ^{125}I-labeled T_3 and T_4 were shown to be distributed throughout perfused tissues only if the binding proteins were present. To adopt this language might be useful so that this effort appears consistent with and cognizant of the most recent developments.

4. For the Hepatic Nuclear Receptors, identify the specific receptors that are related to serum T_4 and T_3 clearance (each would be a separate MIE). There are two reasons for this. First, some chemicals may activate a rat nuclear receptor (NR) but not a human NR and this could be evaluated here. Second, identifying specific NRs in Table 4-1 would help provide a place for the scientific evidence underlying these. In addition, activation of some NRs can bioactivate chemicals that then interfere with some other thyroid MIE, thereby building links between AOPs.

5. For Sulfation and Glucuronidation, the same issues hold as for NRs.

6. The regulation of thyrotropin-releasing hormone (TRH) synthesis or neuronal activity may be important and could be separately identified in Table 4-1. There are several known pathways that can lead to a change in TRH neuronal activity and this may be reflected in TRH mRNA or peptide.

7. "TH Transcription" in Table 4-1 and Figure 4-1 should be "TH-regulated Transcription." This is a very large field and it might be useful to expand on some of these pathways that are better known to be related to an AOP.

In the discussion of the EDSP's Tier1 and Tier2 in the Agency's White Paper, it should be clear to state that Tier 1 is hazard identification and Tier 2 is hazard characterization. Table 4-2 describes the Tier 1 and Tier 2 assays. Recommendations for this table are:

1. For Tier 1, thyroid-specific endpoints of serum T_4 and TSH, Thyroid weight and Thyroid histopathology are known to be separable. That is, some chemicals cause a reduction in serum T4 (both total and free) but do not cause an increase in serum TSH. In the absence of increased TSH, thyroid weight and histopathology are not altered (Bansal et al. 2014; Hood et al. 1999). This means first that thyroid weight and histopathology are endpoints related to TSH, not T_4 directly. Second, this means that in the absence of a clear AOP that can discriminate between those chemicals that affect T_4 and TSH in an "idealized" way compared to those that do not, the validation of these HTS assays will continue to be very problematic. This should be made clear with strategies for ways of addressing that to achieve a reasonable balanced accuracy for the HT assays.

2. The Agency identified a 10% reduction in serum T_4 as an adverse outcome. This level of T_4 reduction (in fact, even an 80% reduction in serum T_4) would not affect growth or body weight. Thus, it would be prudent for the Agency to stipulate that growth and body weight can be affected by low T_4, but only in the most severe circumstances and that many adverse outcomes will occur while growth and body weight remain normal.

3. Thyroid endpoints captured in Tier 2 assays are largely the same as those captured in Tier 1, with the possible inclusion of neurohistopathology, neurobehavioral tests and brain weight. The Agency should reflect whether or not the neurohistopathology measures captured in this are specific for "thyroid related" effects and these should be identified. This also holds for the neurobehavioral tests, since not all behavioral performance measures are sensitive to thyroid hormone. Finally, the Agency should explicitly state the degree of sensitivity of brain weight as a measure of thyroid disruption.

The White Paper Figure 4-1 complements Table 4-1 well, providing a visual diagram of the various thyroid-related AOPs. However, it is difficult to populate this figure with the granularity required for the Agency to employ as a tool. A few Panel comments that reinforce comments regarding Table 4-1 include:

1. Negative feedback in the pituitary and hypothalamus is mediated by TRβ(2) specifically (Dupre et al. 2004; Wondisford 2003) and this should be specified. This is important because TR alpha-null mice have normal serum T_4 and TSH (e.g., (Suzuki and Cheng 2003)). In case reports of a TR alpha mutation in humans, there is more variability. In one case, serum T_4 was low normal, but

TSH was normal (Bochukova et al. 2012). The specific serum and clinical profile is related to the specific mutation in the TR alpha (Demir et al. 2016; Moran and Chatterjee 2016). This indicates that, if there are chemicals that interact with the TR alpha, the effect will not be seen in current guideline endpoints, but adverse outcomes could occur that would not be attributed to thyroid. A significant number of chemicals appear to interact with THR alpha1 in the ToxCast database, with sometimes very low AC50's.

2. The "Delta T_3 in cells and tissues" need to point to "TR binding/transactivation." The endpoints for AOPs need to be more granular.

Question 6: Please refer to White Paper Section 4.3. EPA has summarized currently available assays and test guidelines informative of thyroid AOPs and is developing HT assays for a number of MIEs. Please comment on the ranked importance of MIEs (Table 4-3) and on whether assays for environmentally important MIEs are missing, and include information on both the biological and environmental relevance of these MIEs. In addition, please comment on other assays that would supplement or be orthogonal to the assays currently identified in Table 4-3 or for other KEs or AOs in the thyroid AOP framework (Figure 4-2).

Response:

The Panel's discussion on this charge follows closely and logically from the discussion based on Charge question 5. The Panel reviewed Section 4.3, and discussed the proposed MIE targets for expanded screening efforts, and their ranking in terms of proposed priority for the Agency.

On the issue of the proposed ranking of the MIEs, coverage of identified molecular initiating events (MIEs) for the thyroid hormone endocrine system is deemed to be quite comprehensive, as outlined in Table 4-1. The panel appreciated the overall construct of the Adverse Outcome Pathway (AOP) as the best way to organize the conceptual framework that will guide ongoing and future screening efforts for environmental chemicals that may impact the thyroid hormone endocrine system. The definitions used to describe the status of assays for each MIE, including suitability for adaptation for high throughput assays, are reasonable.

The Panel asked however, that the Agency should provide a clearer definition of what high, medium and low ranking means in terms of priority for action and proposed timelines (Table 4-3). For instance, might "medium" mean placing a hold on new assay development since good enough assays are already in hand, or does "medium" mean some assays exist, but a few more orthogonal ones still need to be developed? Does "low" mean the Agency would not develop assays until there is a possible hit on that MIE

from the literature, or if an effect of a chemical on thyroid hormone (TH) synthesis for example is not explained by existing assays, such as thyroid peroxidase (TPO) or sodium-iodide symporter (NIS) inhibition? Lastly, the Panel discussed whether highest ranking should be placed on MIEs that are most likely to cause a reduction in serum T4, since this endpoint has been focused on in the Tier 1 pubertal rat assay. For the purposes of this discussion, "ranking" is considered in terms of expedited timelines and resource commitment by the Agency in assay development, validation, and refinement for high throughput screening.

With this caveat in mind, each group of MIEs by suggested ranking is addressed below, including the Panel's suggestions for supplemental or orthogonal assays where they may be available for consideration.

The "high" ranking MIEs:

A. The sodium/iodide symporter (NIS): The Panel agreed the relevance of this MIE is agreed to be high, and the presence of developed assays also support this as an important MIE. Other than measuring enhanced radiolabeled iodine uptake in cultured cells as in current use, it is difficult to imagine alternative assays for NIS activity. Expression of NIS in *Xenopus* oocytes as a model (Dai, Levy et al. 1996) or standardizing transient transfection of NIS expression vectors in continuous cell lines, while likely lower throughput approaches, could provide more flexibility to examine different species' NIS chemical sensitivity (Dayem, Basquin et al. 2008), human polymorphisms in NIS (Pohlenz, Rosenthal et al. 1998), and splicing variants versus creating new stable cell lines each time a particular NIS variant is to be screened. Interestingly, NIS knockout mice can take up iodide in the absence of NIS expression (provided that the free iodide is very high); this suggests a secondary route of uptake may also exist (Ferrandino, Kaspari et al. 2017).

B. Thyroperoxidase (TPO): As a rate limiting, key step in TH synthesis, and one that already has well established reference chemicals, the Panel agreed that TPO presents a highly relevant and high priority MIE for screening. Two assays are currently under consideration by the Agency, and their utility based on Tox21 library generated data is currently being evaluated (Agency White Paper, Paul, Hedge et al. 2014). One concern noted by the Agency relates to the reliance on loss of signal as the output in these assays, which may yet be an issue, but appears to be adequately understood by the Agency with a series of controls run in parallel.

C. (and D). Hepatic TH metabolism: The Agency also proposes hepatic TH metabolism via nuclear receptor mediated pathways (e.g. CAR (constitutive androstane receptor) and PXR (pregnane X receptor)) as a high ranking MIE for their role in induction of Phase I and Phase II xenobiotic detoxification enzymes and drug transport genes, as well as a

focus on TH sulfation and glucuronidation via cognate Phase II enzymes (sulfotransferases (SULT family) and UDP-glucuronyltransferases (UGT family)). Metabolism via Type I iodothyronine deiodinase is discussed below. These MIEs are ranked highly based on a well-documented concept that several known chemicals reduce serum T4 levels via enhanced activation of these pathways. While well argued, and supported, they are also not necessarily specific for TH metabolism e.g SULT1E1, the so-called estrogen sulfotransferase, is highly active toward T4 as well (Kester, van Dijk et al. 1999). The Panel does not consider this a major drawback, per se, yet it will important to emphasize up front that the investment in screening these MIEs would likely have broader significance for other endocrine systems of high priority (e.g. steroid hormones, amines).

Adequate coverage of this MIE will also require close attention to the species of interest. CAR and PXR show significant variation in ligand specificities across species, even among mammals (Krasowski, Yasuda et al. 2005). In *Xenopus laevis*, PXR has been designated as BXR (benzoate X receptor) because of its preferred binding to benzoate and related compounds (Krasowski, Ni et al. 2011). The suitability of existing assays under consideration is not discussed in any detail in the White Paper, although multiple assays are listed as in existence. The high priority ranking for these MIEs might also be framed as potentially most relevant when compensatory negative feedback loops are not fully established during development, particularly in the critical window for TH effects on brain development. Even so, it is interesting to note that certain chemicals have been identified that increase T4 clearance and decreased serum T4 levels, yet do not lead to appreciably increased thyroid stimulating hormone (TSH) in adults (Miller, Crofton et al. 2009, Bansal, Tighe et al. 2014), as would be expected due to decreased negative feedback and a fully adaptive compensatory response. The mechanisms underlying this phenomenon are not well understood, and would be highly relevant to uncover since TSH levels are commonly used clinically as a marker of adequate circulating serum T4 levels.

E. The iodothyronine deiodinases (DIO)(Types I, II and III (D1, D2 and D3, respectively)): For these enzymes, the physiological relevance is quite clear, with both pharmacological approaches and genetic models as supporting evidence (Gereben, McAninch et al. 2015). D2 and D3 are key players in intracellular T3 concentrations in target cells (although liver D3 also plays a role in regulating systemic TH levels as well), and D1 is a key enzyme in hepatic iodine recycling but its role in contributing to systemic T3 levels varies by species. While suitable assays and a full suite of reference chemicals are not fully developed, DIO roles in TH signaling and homeostasis are such that the Panel agreed these MIEs are highly relevant for consideration. Like the hepatic Phase I and Phase II enzymes, expression levels as well as intrinsic enzymatic activity may be influenced by environmental chemical exposure.

The “medium” ranking MIEs:

A. Thyroid hormone regulated transcription (initiated at the TRs): As opposed to a strong focus on receptor binding and activation-based screening in the estrogen and androgen disruption programs, the Agency proposes that endocrine disruption via direct thyroid hormone receptor (TR) interaction is a lower (hence, medium ranked) priority at this point, relative to other targets in the TH AOP framework. This proposal is based on two main observations: one, that many chemicals have been discovered that alter circulating TH levels yet do not discernably interact with the TRs, and in high throughput screens using existing transcriptional activity and receptor binding assays, a relatively small number of chemicals have been identified as reliable positive hits. One point of clarification from the White Paper: six “agonist” and four “antagonist” candidates were identified from an initial screen of 1280 chemicals (Freitas, Miller et al. 2014); however, the larger Tox21 screening (8500 chemicals) results have not been fully validated and published to date. The Panel concurs that the TR ligand-binding pocket is fairly selective and may not be affected by a large number of chemicals (particularly without prior biotransformation); yet, those that do bind may affect the pathway at the closest step to the biology. Beyond this particular caveat, the Panel emphasized a few additional considerations.

Thyroid hormone transcriptional activation endpoints will also need to carefully consider differential TR subtype specificity; for example, TRβ, in particular TRβ2, is most responsible for regulating serum T4 levels via its role in negative feedback, whereas TRα1 plays a less prominent role (Flamant and Gauthier 2013). There is some evidence to suggest that chemicals might differentially bind to the highly conserved binding pockets (T4 and T3 do not appear to discriminate in terms of binding affinity to the receptor itself), but in target gene chromatin in target cells, assembly of regulatory complexes may differ, and thus modulate natural and environmental ligand potency differentially by subtype (Flamant and Gauthier 2013, Schroeder and Privalsky 2014). Ideally, future assay development should move beyond the use of over-expressed receptors, including gal4 fusions, for these reasons. In addition, a relatively large number of antagonists (relative to potential agonists) are typically observed in these screens; this may be valid but assay interference also could be the result of an unintended artifact of the specific methodology, e.g. affecting reporter gene enzyme activity or stability, or variety of reasons that are not specific to inhibition of TH induced TR transactivation (Hsieh, Sedykh et al. 2015).

Thus, orthogonal assays for TR mediated transcriptional responses should be considered, and the field has progressed such that the Agency might move to either specific downstream target genes if identified as a key event (for example, Klf9 is a broadly relevant target gene across species (Denver and Williamson 2009, Furlow and Kanamori

2002)) or to newer high throughput transcriptomics approaches (Brockmeier, Hodges et al. 2017). Such endogenous gene responses also have the advantage of interrogating down-regulated gene expression, which is less understood mechanistically and generally ignored by conventional reporter assays (Vella and Hollenberg 2017), despite the fact that 50% of regulated genes are typically down-regulated in most nuclear receptor regulated gene expression programs. TH mediated down regulation also critically includes negative feedback on TSH expression in thyrotropes. Thus, the Agency should consider incorporation of targeted high throughput RNA sequencing in amenable cells or tractable model organisms for identification of activated pathways related to thyroid function and disruption, to replace or extend reporter gene based approaches.

Unfortunately, very few cell lines other than the pituitary derived GH-3 cell line retain strong TH responsiveness (Freitas, Cano et al. 2011). Additional resources such as primary or induced stem cell lines can be explored, particularly if they represent key target cell types (neurons, hepatocytes, etc). Another option is to extend animal model based assays, if they can be adapted to at least medium throughput approaches. For example, *Xenopus laevis* tadpoles are TH responsive at very young post-hatching stages (Mengeling, Wei et al. 2017), and the animals may be adapted to medium throughput assays. The allotetraploid *X. laevis* genome is now complete (Session, Uno et al. 2016), as is the genome for its diploid relative *X. tropicalis* (Karimi, Fortriede et al. 2017) with a large number of known TH target genes for validation available, such as the aforementioned Klf9. Zebrafish also have potential as a medium throughput animal model for TH disruption, if thyroid endocrine physiology is more fully explored to the extent that very specific TH related outcomes (and downstream key target genes) can be reliably measured. So far, changes in pigmentation and swim bladder show the most promise (McMenamin, Bain et al. 2014, Stinckens, Vergauwen et al. 2016). The use of relevant cell and animal models also present an emerging opportunity to use genome editing (e.g. CRISPR/Cas) (Tandon, Conlon et al. 2017, Li, Zhao et al. 2016) to link MIEs to candidate KEs, potentially filling in critical gaps between MIEs and particular AOs. Use of intact cell based or organismal assays may also allow interrogation of chemical effects on multiple MIEs at once such as deiodinases, transporters, and receptor subtype activation, for example, where they are or will need to be well characterized.

B. Serum TH transport proteins (also known as distributor proteins): These MIEs have also been investigated over the past several years for their potential roles as targets of TH disruption, many studies focused on their interaction with brominated flame retardants. Medium to high ranking is warranted given their roles in carrying T4 (primarily) through the bloodstream and balancing access of T4 to target tissues. Serum albumin also serves as a relatively non-specific but significant distributor protein. Again, species differences are important to consider here, given the relative importance of TTR in many vertebrates (indeed, many lack thyroxine-binding globulin (TBG)), and differences in T4 vs. T3 binding affinities in amphibians and fish (Schreiber 2002). The Panel noted the quite high

positive hit rate in the current assays that raised concerns about specificity, however, and should be investigated further by the Agency.

C. Membrane Transporters: Genetic evidence clearly links MCT8 to adequate T4 and T3 uptake across the endothelial cells of the blood brain barrier and neurons, among other cell types, with severe psychomotor deficits in humans lacking MCT8. MCT8 remains the most efficient and specific TH transporter identified to date (Visser, Friesema et al. 2011), although the related MCT10 and the organic anion transporting OATP1 subfamily also play roles as well. While homologs of these transporters are expressed in nonmammalian organisms, and cell specific expression can vary even among mammals (Vancamp and Darras 2017), less is known about any differences in substrate specificity or interactions with reference chemicals. In *Xenopus laevis* at least, MCT8 and OATP1C1 behave similarly to their human counterparts (Mughal, Leemans et al. 2017). Other transporters may still remain to be discovered (Visser, Friesema et al. 2011), however, an MCT8 assay is in development, and is deemed a good place to start in this particular MIE.

D (and E). TRH receptor (TRHR) and TSH receptor (TSHR) assays: These key steps in the hypothalamic-pituitary-thyroid (HPT) axis are also deemed reasonable to include, and the currently existing assay for TSHR has been used to extensively screen chemical libraries for preclinical research. Reliance on cAMP as the read out may result in positives that affect downstream events that are not specific to the TSHR per se (e.g. phosphodiesterase inhibitors perhaps), which may require a secondary screen to rule out. Species differences are also important here; in some amphibians, it has been established that corticotropin releasing hormone (CRH), acting via the CRH receptor, drives metamorphosis rather than TRH and its cognate receptor (Denver 1997, Watanabe, Grommen et al. 2016).

The “low” ranking MIEs:

A. Thyroid hormone receptor binding (*in vitro* assays): The Panel agrees that this MIE is less informative than transcriptional regulation assays. If specific transcriptional read outs of TR activity can be adapted for high throughput, such approaches are more fruitful at this point, although direct TR binding may still be of interest as a secondary assay.

B. For other steps of TH synthesis beyond NIS and TPO (e.g. pendrin, DUOX, IYD), the Panel concluded that for the time being it is fair to rank these as a lower priority. If assays are not readily available or the literature does not clearly indicate potential involvement of these MIEs, the Agency can first examine whether NIS and TPO cover a broad spectrum of chemicals of concern affecting TH synthesis. However, keeping these MIEs in “reserve,” versus elimination as points of concern, is warranted.

For the portion of the charge question regarding any missing assays/MIEs, the following considerations were discussed by the Panel:

For any assay above, the ability to link biotransformation using liver microsomes or other means will continue to be an important consideration. A key hydroxylation step in specific flame retardants for example is necessary to allow not only interaction with TRs and TTRs, but also possibly deiodinases, sulfatases or TH specific transporters as well, versus the parent compounds (Meerts, van Zanden et al. 2000, Macaulay, Chen et al. 2015).

The potential role for the retinoid – X receptor ligands (of pharmaceutical or environmental origin) should be examined in more detail. RXRs form heterodimers with TRs on their response elements, and RXR ligands modulate TR activity in a gene specific and cell specific manner, contrary to prevailing dogma in the nuclear receptor field (Mengeling and Furlow 2015). That high affinity RXR ligands (such as the drug bexarotene) strongly suppress TSH, to the extent of inducing overt hypothyroidism in humans, has been known for some time (Haugen, Brown et al. 1997, Sherman, Gopal et al. 1999). Others have shown environmental and pharmaceutical RXR ligand effects on TH action in specific cell types or during specific developmental stages in amphibians and rodents as well (Mengeling, Murk et al. 2016, Santos-Silva, Andrade et al. 2018). The current Tox21 assays for RXR are in isolation as gal4 fusions, which may be limiting in utility for this use as discussed above for TRs.

Still other MIEs of importance may emerge, and sometimes from unexpected sources. Lithium for example, still used widely to treat bipolar disorders, leads to hypothyroidism in a significant number of patients (Lazarus 2009). While the mode of action is not entirely resolved and may occur at multiple points, evidence suggests that it is linked to impaired TH release from thyrocytes downstream of TSH signaling (Mori, Tajima et al. 1989), and thus representing another potential MIE of interest.

Lastly, continued investigation of a range of organisms will be useful, both as surrogates with potential advantages for higher throughput screening and as sentinel threatened ecological species, as a clearer understanding of the extent of conservation of various points of the thyroid AOP, including endogenous THs, become clearer (Holzer and Laudet 2013, Taylor and Heyland 2017).

General comments:

In summary, the Panel found that the MIEs that were identified and the discussion of the state of associated, existing assays were comprehensive and provide a useful reference for future assay development, also outlined in recent reviews (Murk, Rijntjes et al. 2013).

The Panel suggested a set of orthogonal (mostly transcriptomic- based) and gap – filling (RXR, biotransformation) assays to support the emerging direction of the TH disruption program. However, more clarity is needed from the Agency as to what high, medium, and low ranking of MIEs means for future Panels to evaluate, in real terms regarding future subsequent decision-making processes.

Given that this set of charge questions was paired with a discussion of the performance of AR transactivation and steroidogenesis assays, lessons learned from estrogen and androgen disruptor programs could inform MIE assay development for the thyroid AOP context. A clear understanding of how many orthogonal assays for each MIE are required for high-level confidence in sensitivity and specificity would be very useful (which may be inherent to each assay's performance). The need for systems modeling across MIEs, cell types, species and life stages to fully integrate and validate the high throughput screening program was recognized and encouraged. It was also noted that the Agency recognizes that greater understanding of quantitative interactions leading to and from key events affecting AOs is also key to this approach. Ultimately, clearly linking MIEs to KEs to AOPs, including where they overlap and intersect, may require a reiterative process between MIE assay development and basic research in cells and systems to identify quantifiable KEs downstream of the targeted MIEs, extending beyond the TH program discussed here.

Other comments on the White Paper:

Table 4-1 and 4-3 should read "TH regulated transcription"
Table 4-3 wrong Tox21 TH transcription assay ((TOX21_TSHR_Antagonist_ratio 2X (TOX21_TR_LUC_GH3_Agonist, TOX21_TR_LUC_GH3_Antagonist), but this was fixed in the slides presented by the Agency at the meeting.
TR should also be shown in the pituitary in AOP Figure 4-2

REFERENCES

Bansal, R., D. Tighe, A. Danai, D. F. Rawn, D. W. Gaertner, D. L. Arnold, M. E. Gilbert and R. T. Zoeller. (2014). Polybrominated diphenyl ether (DE-71) interferes with thyroid hormone action independent of effects on circulating levels of thyroid hormone in male rats. Endocrinology 155(10): 4104-4112.

Bochukova E, Schoenmakers N, Agostini M, Schoenmakers E, Rajanayagam O, Keogh JM, Henning E, Reinemund J, Gevers E, Sarri M, Downes K, Offiah A, Albanese A, Halsall D, Schwabe JW, Bain M, Lindley K, Muntoni F, Vargha-Khadem F, Dattani M, Farooqi IS, Gurnell M, and K. Chatterjee. (2012). A mutation in the thyroid hormone receptor alpha gene. The New England journal of medicine 366(3): 243-249.

Burgoon, Lyle D. (2017). Autoencoder Predicting Estrogenic Chemical Substances (APECS): An improved approach for screening potentially estrogenic chemicals using *in vitro* assays and deep learning, In Computational Toxicology, Volume 2, 2017, Pages 45-49, ISSN 2468-1113, https://doi.org/10.1016/j.comtox.2017.03.002.

Brockmeier, E. K., G. Hodges, T. H. Hutchinson, E. Butler, M. Hecker, K. E. Tollefsen, N. Garcia-Reyero, P. Kille, D. Becker, K. Chipman, J. Colbourne, T. W. Collette, A. Cossins, M. Cronin, P. Graystock, S. Gutsell, D. Knapen, I. Katsiadaki, A. Lange, S. Marshall, S. F. Owen, E. J. Perkins, S. Plaistow, A. Schroeder, D. Taylor, M. Viant, G. Ankley and F. Falciani. (2017). The Role of Omics in the Application of Adverse Outcome Pathways for Chemical Risk Assessment. Toxicol Sci 158(2): 252-262.

Dai, G., O. Levy and N. Carrasco. (1996). Cloning and characterization of the thyroid iodide transporter. Nature 379(6564): 458-460.

Dayem, M., C. Basquin, V. Navarro, P. Carrier, R. Marsault, P. Chang, S. Huc, E. Darrouzet, S. Lindenthal and T. Pourcher. (2008). Comparison of expressed human and mouse sodium/iodide symporters reveals differences in transport properties and subcellular localization. J Endocrinol 197(1): 95-109.

Demir K, van Gucht AL, Buyukinan M, Catli G, Ayhan Y, Bas VN, Dündar B, Özkan B, Meima ME, Visser WE, Peeters RP, and T.J. Visser. (2016). Diverse Genotypes and Phenotypes of Three Novel Thyroid Hormone Receptor-alpha Mutations. J Clin Endocrinol Metab 101(8): 2945-2954

Denver, R. J. (1997). Environmental stress as a developmental cue: corticotropin-releasing hormone is a proximate mediator of adaptive phenotypic plasticity in amphibian metamorphosis. Horm Behav 31(2): 169-179.

Denver, R. J. and K. E. Williamson. (2009). Identification of a thyroid hormone response element in the mouse Kruppel-like factor 9 gene to explain its postnatal expression in the brain. Endocrinology 150(8): 3935-3943.

Dupre SM, Guissouma H, Flamant F, Seugnet I, Scanlan TS, Baxter JD, Samarut J, Demeneix BA, and N. Becker. (2004). Both thyroid hormone receptor (TR)beta 1 and TR beta 2 isoforms contribute to the regulation of hypothalamic thyrotropin-releasing hormone. Endocrinology 145(5): 2337-2345.

Ferrandino, G., R. R. Kaspari, A. Reyna-Neyra, N. E. Boutagy, A. J. Sinusas and N. Carrasco. (2017). An extremely high dietary iodide supply forestalls severe hypothyroidism in Na(+)/I(-) symporter (NIS) knockout mice. Sci Rep 7(1): 5329.

Flamant, F. and K. Gauthier. (2013). Thyroid hormone receptors: the challenge of elucidating isotype-specific functions and cell-specific response. Biochim Biophys Acta 1830(7): 3900-3907.

Freitas, J., P. Cano, C. Craig-Veit, M. L. Goodson, J. D. Furlow and A. J. Murk. (2011). Detection of thyroid hormone receptor disruptors by a novel stable in vitro reporter gene assay. Toxicol In Vitro 25(1): 257-266.

Freitas, J., N. Miller, B. J. Mengeling, M. Xia, R. Huang, K. Houck, I. M. Rietjens, J. D. Furlow and A. J. Murk. (2014). Identification of thyroid hormone receptor active compounds using a quantitative high-throughput screening platform. Curr Chem Genom Transl Med 8: 36-46.

Friesema EC, Grueters A, Biebermann H, Krude H, von Moers A, Reeser M, Barrett TG, Mancilla EE, Svensson J, Kester MH, Kuiper GG, Balkassmi S, Uitterlinden AG, Koehrle J, Rodien P, Halestrap AP, and T.J.Visser. (2004). Association between mutations in a thyroid hormone transporter and severe X-linked psychomotor retardation. Lancet 364(9443): 1435-1437

Furlow, J. D. and A. Kanamori. (2002). The transcription factor basic transcription element-binding protein 1 is a direct thyroid hormone response gene in the frog Xenopus laevis. Endocrinology 143(9): 3295-3305.

Genest, M., Masse, JC, Plante, JF. (2017). depth: Nonparametric Depth Functions for Multivariate Analysis. R package version 2.1.1. https://cran.r-project.org/web/packages/depth/depth.pdf.

Gereben, B., E. A. McAninch, M. O. Ribeiro and A. C. Bianco. (2015). Scope and limitations of iodothyronine deiodinases in hypothyroidism. Nat Rev Endocrinol 11(11): 642-652.

Hallinger, DR; Murr, AS; Buckalew, AR; Simmons, SO; Stoker, TE; and SC Laws. (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.

Haugen, B. R., N. S. Brown, W. M. Wood, D. F. Gordon and E. C. Ridgway. (1997). The thyrotrope-restricted isoform of the retinoid-X receptor-gamma1 mediates 9-cis-retinoic acid suppression of thyrotropin-beta promoter activity. Mol Endocrinol 11(4): 481-489.

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; and G. Timm. (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.

Holzer, G. and V. Laudet. (2013). Thyroid hormones and postembryonic development in amniotes. Curr Top Dev Biol 103: 397-425.

Hood A, Hashmi R, and CD Klaassen. (1999). Effects of microsomal enzyme inducers on thyroid-follicular cell proliferation, hyperplasia, and hypertrophy. Toxicol Appl Pharmacol 160(2): 163-170.

Hsieh, J. H., A. Sedykh, R. Huang, M. Xia and R. R. Tice. (2015). A Data Analysis Pipeline Accounting for Artifacts in Tox21 Quantitative High-Throughput Screening Assays. J Biomol Screen 20(7): 887-897.

Jones, J.O., Bolton, E.C., Huang, Y., Feau, C., Guy, R. K., Yamamoto, K.R., Hann, B., Diamond, M.I. (2009). Non-competitive androgen receptor inhibition *in vitro* and *in vivo*. Proc Natl Acad Sci U S A. ;106(17):7233-7238. doi:10.1073/pnas.0807282106

Kallio PJ, Poukka H, Moilanen A, Jänne OA, and JJ Palvimo. (1995). Androgen receptor-mediated transcriptional regulation in the absence of direct interaction with a specific DNA element. Mol Endocrinol. 9:1017–28.

Karimi, K., J. D. Fortriede, V. S. Lotay, K. A. Burns, D. Z. Wang, M. E. Fisher, T. J. Pells, C. James-Zorn, Y. Wang, V. G. Ponferrada, S. Chu, P. Chaturvedi, A. M. Zorn and P. D. Vize. (2017). Xenbase: a genomic, epigenomic and transcriptomic model organism database. Nucleic Acids Res.

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.

Kester, M. H., C. H. van Dijk, D. Tibboel, A. M. Hood, N. J. Rose, W. Meinl, U. Pabel, H. Glatt, C. N. Falany, M. W. Coughtrie and T. J. Visser. (1999). Sulfation of thyroid hormone by estrogen sulfotransferase. J Clin Endocrinol Metab 84(7): 2577-2580.

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; and R. Judson. (2017). Supporting information: Development and validation of a computational model for androgen receptor activity [Supplemental Data]. Chem Res Toxicol 30: 946-964.

Krasowski, M. D., A. Ni, L. R. Hagey and S. Ekins. (2011). Evolution of promiscuous nuclear hormone receptors: LXR, FXR, VDR, PXR, and CAR. Mol Cell Endocrinol 334(1-2): 39-48.

Krasowski, M. D., K. Yasuda, L. R. Hagey and E. G. Schuetz. (2005). Evolutionary selection across the nuclear hormone receptor superfamily with a focus on the NR1I subfamily (vitamin D, pregnane X, and constitutive androstane receptors). Nucl Recept 3: 2.

Lazarus, J. H. (2009). Lithium and thyroid. Best Pract Res Clin Endocrinol Metab 23(6): 723-733.

Li, M., L. Zhao, P. S. Page-McCaw and W. Chen. (2016). Zebrafish Genome Engineering Using the CRISPR-Cas9 System. Trends Genet 32(12): 815-827.

Macaulay, L. J., A. Chen, K. D. Rock, L. V. Dishaw, W. Dong, D. E. Hinton and H. M. Stapleton. (2015). Developmental toxicity of the PBDE metabolite 6-OH-BDE-47 in zebrafish and the potential role of thyroid receptor beta. Aquat Toxicol 168: 38-47.

McMenamin, S. K., E. J. Bain, A. E. McCann, L. B. Patterson, D. S. Eom, Z. P. Waller, J. C. Hamill, J. A. Kuhlman, J. S. Eisen and D. M. Parichy. (2014). Thyroid hormone-dependent adult pigment cell lineage and pattern in zebrafish. Science 345(6202): 1358-1361.

Meerts, I. A., J. J. van Zanden, E. A. Luijks, I. van Leeuwen-Bol, G. Marsh, E. Jakobsson, A. Bergman and A. Brouwer. (2000). Potent competitive interactions of some brominated flame retardants and related compounds with human transthyretin in vitro. Toxicol Sci 56(1): 95-104.

Mengeling, B. J. and J. D. Furlow. (2015). Pituitary specific retinoid-X receptor ligand interactions with thyroid hormone receptor signaling revealed by high throughput reporter and endogenous gene responses. Toxicol In Vitro 29(7): 1609-1618.

Mengeling, B. J., A. J. Murk and J. D. Furlow. (2016). Trialkyltin Rexinoid-X Receptor Agonists Selectively Potentiate Thyroid Hormone Induced Programs of Xenopus laevis Metamorphosis. Endocrinology 157(7): 2712-2723.

Mengeling, B. J., Y. Wei, L. N. Dobrawa, M. Streekstra, J. Louisse, V. Singh, L. Singh, P. J. Lein, H. Wulff, A. J. Murk and J. D. Furlow. (2017). A multi-tiered, in vivo, quantitative assay suite for environmental disruptors of thyroid hormone signaling. Aquat Toxicol 190: 1-10.

Miller, M. D., K. M. Crofton, D. C. Rice and R. T. Zoeller. (2009). Thyroid-disrupting chemicals: interpreting upstream biomarkers of adverse outcomes. Environ Health Perspect 117(7): 1033-1041.

Moran C, Chatterjee K. 2016. Resistance to Thyroid Hormone alpha-Emerging Definition of a Disorder of Thyroid Hormone Action. J Clin Endocrinol Metab 101(7): 2636-2639.

Mori, M., K. Tajima, Y. Oda, I. Matsui, K. Mashita and S. Tarui (1989). Inhibitory effect of lithium on the release of thyroid hormones from thyrotropin-stimulated mouse thyroids in a perifusion system. Endocrinology 124(3): 1365-1369.

Mughal, B. B., M. Leemans, E. C. Lima de Souza, S. le Mevel, P. Spirhanzlova, T. J. Visser, J. B. Fini and B. A. Demeneix. (2017). Functional Characterization of Xenopus Thyroid Hormone Transporters mct8 and oatp1c1. Endocrinology 158(8): 2694-2705.

Murk, A. J., E. Rijntjes, B. J. Blaauboer, R. Clewell, K. M. Crofton, M. M. Dingemans, J. D. Furlow, R. Kavlock, J. Kohrle, R. Opitz, T. Traas, T. J. Visser, M. Xia and A. C. Gutleb. (2013). Mechanism-based testing strategy using in vitro approaches for identification of thyroid hormone disrupting chemicals. Toxicol In Vitro 27(4): 1320-1346.

Nakamura, T., and Imada, T. (2005). Multiple comparison procedure of Dunnett's type for multivariate normal means. Journal of the Japanese Society of Computational Statistics, 18, 21-32.

National Academies of Sciences, Engineering, and Medicine. (2017). Application of Systematic Review Methods in an Overall Strategy for Evaluating Low-Dose Toxicity from Endocrine Active Chemicals. Washington, DC: The National Academies Press.

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.

Ovacik, M. A., and I. P. Androulakis. (2013). Enzyme sequence similarity improves the reaction alignment method for cross-species pathway comparison. Toxicology and Applied Pharmacology, 27(1), 363-371.

Paul, K. B., J. M. Hedge, D. M. Rotroff, M. W. Hornung, K. M. Crofton and S. O. Simmons. (2014). Development of a thyroperoxidase inhibition assay for high-throughput screening. Chem Res Toxicol 27(3): 387-399.

Pohlenz, J., I. M. Rosenthal, R. E. Weiss, S. M. Jhiang, C. Burant and S. Refetoff. (1998). Congenital hypothyroidism due to mutations in the sodium/iodide symporter. Identification of a nonsense mutation producing a downstream cryptic 3' splice site. J Clin Invest 101(5): 1028-1035.

Santos-Silva, A. P., M. N. Andrade, P. Pereira-Rodrigues, F. D. Paiva-Melo, P. Soares, J. B. Graceli, G. R. M. Dias, A. C. F. Ferreira, D. P. de Carvalho and L. Miranda-Alves. (2018). Frontiers in endocrine disruption: Impacts of organotin on the hypothalamus-pituitary-thyroid axis. Mol Cell Endocrinol 460: 246-257.

Schreiber, G. (2002). The evolutionary and integrative roles of transthyretin in thyroid hormone homeostasis. J Endocrinol 175(1): 61-73.

Schneikert J, Peterziel H, Defossez PA, Klocker H, de Launoit Y, and A.C.Cato. (1996). Androgen receptor-Ets protein interaction is a novel mechanism for steroid hormone-mediated down-modulation of matrix metalloproteinase expression. J Biol Chem. 271:23907–13.

Schroeder, A. C. and M. L. Privalsky. (2014). Thyroid hormones, t3 and t4, in the brain. Front Endocrinol (Lausanne) 5: 40.

Session, A. M., Y. Uno, T. Kwon, J. A. Chapman, A. Toyoda, S. Takahashi, A. Fukui, A. Hikosaka, A. Suzuki, M. Kondo, S. J. van Heeringen, I. Quigley, S. Heinz, H. Ogino, H. Ochi, U. Hellsten, J. B. Lyons, O. Simakov, N. Putnam, J. Stites, Y. Kuroki, T. Tanaka, T. Michiue, M. Watanabe, O. Bogdanovic, R. Lister, G. Georgiou, S. S. Paranjpe, I. van Kruijsbergen, S. Shu, J. Carlson, T. Kinoshita, Y. Ohta, S. Mawaribuchi, J. Jenkins, J. Grimwood, J. Schmutz, T. Mitros, S. V. Mozaffari, Y. Suzuki, Y. Haramoto, T. S. Yamamoto, C. Takagi, R. Heald, K. Miller, C. Haudenschild, J. Kitzman, T. Nakayama, Y. Izutsu, J. Robert, J. Fortriede, K. Burns, V. Lotay, K. Karimi, Y. Yasuoka, D. S. Dichmann, M. F. Flajnik, D. W. Houston, J. Shendure, L. DuPasquier, P. D. Vize, A. M.

Zorn, M. Ito, E. M. Marcotte, J. B. Wallingford, Y. Ito, M. Asashima, N. Ueno, Y. Matsuda, G. J. Veenstra, A. Fujiyama, R. M. Harland, M. Taira and D. S. Rokhsar. (2016). Genome evolution in the allotetraploid frog Xenopus laevis. Nature 538(7625): 336-343.

Sherman, S. I., J. Gopal, B. R. Haugen, A. C. Chiu, K. Whaley, P. Nowlakha and M. Duvic. (1999). Central hypothyroidism associated with retinoid X receptor-selective ligands. N Engl J Med 340(14): 1075-1079.

Spade DJ, Hall SJ, Saffarini CM, Huse SM, McDonnell EV, and K. Boekelheide. (2014) Differential response to abiraterone acetate and di-n-butyl phthalate in an androgen-sensitive human fetal testis xenograft bioassay. Toxicol Sci. 2014 Mar;138(1):148-60.

Stinckens, E., L. Vergauwen, A. L. Schroeder, W. Maho, B. R. Blackwell, H. Witters, R. Blust, G. T. Ankley, A. Covaci, D. L. Villeneuve and D. Knapen. (2016). Impaired anterior swim bladder inflation following exposure to the thyroid peroxidase inhibitor 2-mercaptobenzothiazole part II: Zebrafish. Aquat Toxicol 173: 204-217.

Struyf, A., and Rousseeuw, P. J. (2000). High-dimensional computation of the deepest location. Computational Statistics & Data Analysis, 34(4), 415-426.

Suzuki H, and SY Cheng. (2003). Compensatory role of thyroid hormone receptor (TR) alpha 1 in resistance to thyroid hormone: study in mice with a targeted mutation in the TR beta gene and deficient in TR alpha 1. Mol Endocrinol 17(8): 1647-1655.

Tandon, P., F. Conlon, J. D. Furlow and M. E. Horb. (2017). Expanding the genetic toolkit in Xenopus: Approaches and opportunities for human disease modeling. Dev Biol 426 (2): 325-335.

Taylor, E. and A. Heyland. (2017). Evolution of thyroid hormone signaling in animals: Non-genomic and genomic modes of action. Mol Cell Endocrinol 459: 14-20.

Tukey, J. W. (1975). Mathematics and the picturing of data. In Proceedings of the International Congress of Mathematicians Vol. 2, pp. 523-531.

U.S.EPA (U.S. Environmental Protection Agency). (1999) Endocrine Disruptor Screening and Testing Advisory Committee (EDSTAC) Final Report; at chapter 7, pg. 2-3.

US EPA, EDSP, Endocrine Disruptor Screening Program (2011). Androgen Receptor Binding (Rat Ventral Prostate Cytosol). Standard Evaluation Procedure. OCSPP 890.1150 U.S. Environmental Protection Agency Washington, DC 20460 October 2011

U.S. EPA. (2014) December 2014 FIFRA SAP Meeting - Endocrine Activity and Exposure-based Prioritization and Screening [EPA Report].

U.S. EPA. (2015). 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.

Vancamp, P. and V. M. Darras. (2017). From zebrafish to human: A comparative approach to elucidate the role of the thyroid hormone transporter MCT8 during brain development. Gen Comp Endocrinol.

Vella, K. R. and A. N. Hollenberg. (2017). The actions of thyroid hormone signaling in the nucleus. Mol Cell Endocrinol 458: 127-135.

Visser, W. E., E. C. Friesema and T. J. Visser. (2011). Minireview: thyroid hormone transporters: the knowns and the unknowns. Mol Endocrinol 25(1): 1-14.

Watanabe, Y., S. V. H. Grommen and B. De Groef. (2016). Corticotropin-releasing hormone: Mediator of vertebrate life stage transitions? Gen Comp Endocrinol 228: 60-68.

Watt, E. (2016). eapath: Computational Models for Estrogen and Androgen Receptor Activity. Retrieved from https://github.com/ericwatt/eapath

Wondisford FE. (2003). Thyroid hormone action: insight from transgenic mouse models. J Investig Med 51(4): 215-220.

Zhang, JH; Chung, TDY; and KR Oldenburg. (1999). A simple statistical parameter for use in evaluation and validation of high throughput screening assays. J Biomol Screen 4: 67-73.