U.S. Environmental Protection Agency (EPA) Office of Research and Development (ORD)

EPA Transcriptomic Assessment Products (ETAP) Panel

Hybrid Meeting Minutes

July 11–12, 2023

Dates and Times: July 11, 2023, 9:00 a.m. to 5:45 p.m.; July 12, 2023, 9:00 a.m. to 5:15 p.m. Eastern Time

Location: EPA Research Triangle Park Research Facility, 109 TW Alexander Drive, Durham, North Carolina

Meeting Minutes

Provided below is a list of the presentations and discussions that took place during the meeting with hyperlinked page numbers. The minutes follow. The agenda is provided in Appendix A, the participants are listed in Appendix B, and the charge questions are provided in Appendix C.

DRAFT

Tuesday, July 11, 2023

The meeting generally followed the issues and timing as presented in the agenda provided in Appendix A of this meeting summary. However, some minor deviations occurred due to speakers' schedules, which are noted.

Welcome

Maureen Gwinn, Principal Deputy Assistant Administrator for Research and Development, Office of Research and Development (ORD)

Dr. Maureen Gwinn thanked the members for their contributions. She discussed the mission of the EPA Transcriptomic Assessment Product (ETAP) Panel and the importance of utilizing transcriptomics for human risk assessment. She thanked the ORD authors for developing the reports and regulatory partners for providing feedback.

Introduction to the Panel

Tom Tracy, Designated Federal Official (DFO), Office of Science Advisor, Policy, and Engagement, ORD

Mr. Tom Tracy, Designated Federal Official (DFO) for the Board of Scientific Counselors (BOSC), introduced ETAP Panel members Katherine von Stackelberg, Craig Rowlands, Tom Burke, Harvey Clewell, Stephen Edwards, Rebecca Fry, Gladys Liehr, Gloria Post, Justin Teeguarden, Chris Vulpe, Fred Wright, and Lauren Zeise. He then introduced Samantha Jones to provide an overview of how ETAP could be used to evaluate risk assessment of data poor chemicals.

EPA ORD Portfolio and Approach to Where ETAP Fits

Samantha Jones, National Program Director, ORD

Dr. Samantha Jones stated that the aim of ORD is to provide the scientific foundation for EPA to execute its mandate to protect human health and the environment. Dr. Jones stated that ORD provides research to support the Agency in identifying and solving long-term environmental challenges. Dr. Jones emphasized that ORD can provide valuable information on the hazard identification and dose-response in chemical assessment.

Dr. Jones stated that the ORD aims to provide science assessments that focus on high quality, transparent, consistent, and scientifically defensible assessments. Dr. Jones described examples of science assessments including Integrated Science Assessment (ISAs), Integrated Risk Information System (IRIS) Assessments, Provisional Peer Reviewed Toxicity Values (PPRTVs), and Provisional Advisory Levels (PALs).

Dr. Jones emphasized that it has been a challenge for the environmental health community to provide chemical assessments in a timely manner, limited by the amount of chemical data and the amount time to assess chemical toxicity. Dr. Jones proposed the need to conduct toxicity testing in a shorter period of time and matching the needs of the assessment to the products assessed ("fit for purpose").

Dr. Jones provided an overview of the assessment products provided by ORD including IRIS, PPRTVs, PALs, and IRAs, outlining the trade-offs between products that are data-rich but

require longer development times and those that are less data rich but require shorter development times. Assessments products such as IRIS require several years to generate but can be very data rich. PPRTVs and PALs have fewer data than IRIS, but can be assessed in less than two years depending on the chemical. However, a major challenge for IRIS, PPRTVs, and PALs is the availability of data and time. Dr. Jones emphasized the need for IRAs to fill in data gaps for chemicals with limited data and in a shorter period of time.

Dr. Jones outlined the use of Human Health Toxicity Assessments (HHTAs) as an example of a "fit for purpose" approach that could be applied across all products. Dr. Jones utilized Perfluoropropanoic Acid (PFPrA) as an example of a recently developed HHTA. HHTAs, modeled after PPRTVs, can be context-specific to assess site-specific issues. The process includes an internal peer review by ORD scientists, a contract-led, independent letter external peer review and no public comment period.

Dr. Jones discussed how ETAP could be used to fill in the gap of data poor chemicals and do so in a shorter period. ETAP could enable the development and release of chemical assessment in as little as 9 months from chemical procurement to assessment release. Dr. Jones emphasized that the ETAP assessment is a major advancement in toxicity assessment, using a science-driven approach to assess novel data.

Dr. Jones outlined the process involved in the selection of assessment products. The first step involves problem formulation, which assesses what decision needs to be made and in what amount of time. Following problem formulation, data gathering of all streams of evidence, including *in vivo* and *in vitro,* is conducted. If there are sufficient data to assess the formulated problem question, a complete human health assessment sufficient for the data available and time needed is conducted. If there are not sufficient data, a read-across approach using a suitable chemical analogue may be used. However, if there are neither sufficient data nor a suitable analogue, appropriate toxicity testing specific for the problem is conducted and used for human health assessment.

Dr. Jones summarized that ORD aims to conduct science to inform decision-makers at all levels using high quality, transparent, consistent, and scientifically defensible assessment products. ORD is implementing a portfolio approach to chemical assessments where the assessment products match the needs of the decisions being made. Newer assessments products like ETAP allow ORD to produce assessments that contain actionable science sooner.

Dr. Harvey Clewell asked for the definitions of the acronyms ISA and PPRTV. Dr. Jones answered that ISA is an Integrated Science Assessment that is like an IRIS assessment but is focused on six particular pollutants that fall under the National Ambient Air Quality Standard. PPRTV is a Provisional Peer Reviewed Toxicity Value, which deals with data specifically related to superfund sites and informs which chemicals we assess as a concern for human health. Dr. Jones elaborated that the chemicals from these superfund sites can often be data poor and ETAP could help resolve these data gaps. Dr. Gloria Post noted that PPRTVs are similar to HHTAs, but HHTAs are not used for superfund sites.

Day 1 Agenda, Introduction of ETAP Team, and Charge to Panel

Rusty Thomas, Director, Center for Computational Toxicology and Exposure (CCTE)

Dr. Rusty Thomas outlined the goals and objectives of the ETAP team, which is to develop and operationalize a new EPA human health assessment product for data poor chemicals that can be completed from chemical procurement to publications of assessment in less than 9 months. The objectives of the project were to (1) review relevant literature, (2) refine dose response analysis methods for standardized study design, (3) compare error in concordance with variability in toxicity studies, (4) develop standardized method for ETAP, (5) compare transcriptomic reference values with traditional reference doses (RfD), (6) develop example ETAP for data poor PFAS, and (7) conduct socioeconomic case study on the human health and economic value of ETAP. The goals and objectives were provided in three EPA reports. Objectives #1-3 were addressed in "Scientific support for developing and applying transcriptomic points-of-departure," objectives #4-6 were addressed in "The standardized methods for running the short-term *in vivo* transcriptomic studies and developing the ETAP," and objective #7 was addressed in "Socioeconomic case study on the human health and economic value of the ETAP."

Dr. Thomas addressed that the socioeconomic case study (Objective #7) was developed from the National Academy of Sciences assessment that time was significant factor in the "nature and quality" of risk assessment. Conducting more studies or improvements in the assessment can reduce uncertainty but may require additional resources and delay. Dr. Thomas recommended that EPA adopt a value of information (VOI) analysis to provide a more objective decision framework in assessing the trade-offs of time, uncertainty, and cost. VOI is a method for quantifying the expected gain in economic terms for reducing uncertainty through the collection of additional data or information. VOI has been applied in chemical risk assessment but to date has not considered the impact of time.

Dr. Thomas utilized a publication from Hagiwara et. al. (2022) as a case study of how to utilize VOI in chemical risk assessment. The authors of the study used VOI to look at chemical characteristics (i.e., exposure level, population variability to exposure, population size, health effects, population variability to toxicity) and toxicity testing characteristics (i.e., uncertainty of effect level, timeliness, cost) to generate VOI metrics where you can examine both toxicity tests A and B by comparing the costs of testing and the benefit of testing for each test.

Dr. Thomas discussed how the framework from this study could be adapted to assess data poor chemicals, specifically how the ETAP approach compares with traditional toxicity testing. To do this, the group constructed 306 data driven scenarios comparing ETAP with traditional toxicity testing. The health costs were based on a range of exposure estimates and population variability, different population sizes, range of controls costs, range of health endpoints and associated costs, uncertainty assumptions comparing ETAP and chronic bioassay, and target risk versus benefit risk decision context. Comparison of the two approaches generates a bounded range of VOI metrics for all possible scenarios.

Dr. Thomas summarized the results of the ETAP socioeconomic case study. The analysis of the trade-offs in the timeliness, uncertainties, and costs demonstrated that the ETAP approach was favored over traditional toxicity testing and human health assessment in most of the scenarios examined. For benefit-risk decisions, ETAP was favored 81% of the time, while the remaining 19% favored neither ETAP or traditional process. For target-risk decisions, ETAP was favored between 89 – 99% of the time, while 9% favored neither ETAP or traditional process. Across all scenarios evaluated, the median difference between ETAP and the traditional process in one VOI

metric, the Expected Net Benefit of Sampling (ENBS), was $47 billion for benefit-risk decisions and $81 billion for target-risk decisions. Negative values for ENBS were more frequently observed for the traditional process in the benefit-risk decision context, suggesting that the delay and costs associated with testing and decision-making were greater than the eventual benefit in some scenarios.

Dr. Thomas then summarized the agenda for Day 1. He thanked all those who contributed to the project, including the ETAP team, the ICF team, and the reviewers for ORD.

Dr. Thomas also discussed the structure of responses to charge questions. The response categories are:

- Tier 1: Recommendations – responses necessary to adequately support scientific basis of the ETAP and implementation as a new ORD assessment product or to improve clarity of the presentation.
- Tier 2: Suggestions – responses for EPA to consider to strengthen the scientific basis of the ETAP and implementation as a new ORD assessment product or to improve clarity of the presentation.
- Tier 3: Future Considerations – Advice you may have for scientific exploration or research to inform future work.

Dr. Thomas stated that the charge questions would be broken out by day.

Dr. Justin Teeguarden asked for clarification on whether the benefits in the study were only economic or also included human health benefits. Dr. Thomas clarified that the human health benefits were evaluated in economic terms such as loss of productivity and health care costs. Dr. Tom Burke thanked Dr. Thomas for the presentation and commended the group for accounting for the economic costs of time in the assessment of data poor chemicals.

Science Support Introduction/Background

Alison Harrill, Associate Director, CCTE

Dr. Harrill introduced the current worldwide and domestic chemical and toxicity testing and human health assessment landscape and the need for an additional ORD assessment product. Dr. Harrill discussed that greater than 350,000 chemicals or mixtures have been registered in one or more chemical inventories globally and, in the United States, the Toxic Substances Control Act (TSCA) inventory contains over 86,000 chemicals with 42,000 of those chemicals commercially available.

Dr. Harrill emphasized that understanding the human health impacts of chemical exposure requires toxicity testing and access to toxicity data which often entails guideline and special animal studies. She pointed out that insufficient data on many of these chemicals coupled with the required time and resources to study them has led to fewer chemicals with reference values for regulatory applications.

Dr. Harrill introduced ETAP as a way to develop transcriptomic reference values (TRVs) to meet the need for toxicity testing and human health assessment of chemicals. ETAP can be used for data-poor chemical substances with insufficient existing or publicly accessible, repeated-dose

experimental toxicity testing or suitable human epidemiological evidence. The target for completion is 9 months from chemical procurement to ETAP issuance.

Dr. Harrill described the development of ETAP. Transcriptomics, the large-scale measurement of gene expression changes, can be applied in toxicology to enable the broad characterization of biological processes impacted following chemical exposure. Additionally, a substantial body of research has demonstrated that doses causing transcriptional changes are concordant with doses causing adverse apical effects in traditional toxicology studies, underscoring the potential applications for regulatory decision making.

Dr. Harrill described the Agency's longstanding commitment to utilizing emerging technologies to enhance testing paradigms and improve the utility and predictability of risk assessment methods and how transcriptomic data can be utilized to accomplish these goals. In 2002, the EPA released an interim policy that advocated using transcriptomics data on a case-by-case basis in a weight of evidence approach. In 2004, the Agency released its first report on potential applications of genomics and emphasized the applications, prioritization, monitoring, reporting provisions, mode of action, identifying sensitive populations, and addressing mixtures. The report also identified gaps including the need for adequate technical infrastructure, technical framework for data analysis and standardization criteria for acceptance of transcriptomic data.

The Agency used transcriptomics information in a mode-of-action, weight-of-evidence cancer risk assessment of acetochlor in 2004. The study utilized time course transcriptomics data derived from rat olfactory mucosa at a single dose and found that early gene expression changes interpreted to be consistent with oxidative damage to DNA followed by cell proliferation, while late gene expression changes interpreted to be consistent with tumorigenic progression.

In 2007, EPA released an interim guidance for microarray data submissions, quality, and analysis in 2007 that provided recommendations on performance approaches for quality assessment parameters, data analysis, Agency data submissions, and data management and provided a template for genomics data evaluation.

EPA released a case study for applications of transcriptomic data to human health risk assessment of dibutyl phthalate in 2009. It outlined a systematic and flexible approach to accommodate different health and risk assessment practices, focusing primarily on informing mode-of-action as part of a weight-of-evidence approach. The case study also provided some recommendations on best practices and highlighted current limitations.

Dr. Harrill emphasized the reproducibility of transcriptomic data has bolstered the confidence in its use. MicroArray Quality Control (MAQC), led by FDA, in cross-site, cross-platform studies demonstrated strong inter- and intra-platform reproducibility using RNA reference samples. MicroArray/Sequencing Quality Control (SEQC) evaluated next-generation sequencing (NGS) technology, establishing best practices for RNA-seq methods, characterizing inter-platform reproducibility of RNA-seq protocols and technologies, and evaluating bioinformatic tools. SEQC2 brought 300 scientists from 150 industry, academic, government organizations to evaluate a variety of NGS technologies, including RNA-seq for germline variant detection, cancer genomics, biomarker discovery, precision medicine. This provided a scientific foundation that enabled regulatory approvals of NGS-based liquid biopsy tests for solid tumors as

companion diagnostics and increased confidence in use of transcriptomic-based technologies for regulatory applications.

Dr. Harrill outlined the efforts being made to build an infrastructure to analyze transcriptomic data. BMDExpress2 was developed via intergovernmental partnership to evaluate transcriptomic dose response. BMDExpress2 employs validated continuous parametric models and fits curves computed for each transcript, followed by functional classification analysis that assigns transcriptomic features into predefined gene sets and determines gene set potency estimates as benchmark dose/lower limits. The Organization for Economic Cooperation and Development (OECD) also introduced an internationally accepted transcriptomic reporting framework in 2021, which provided guidance on reporting of transcriptomic information that fosters transparency and reproducibility. The template is awaiting formal approval at the OECD and – once adopted – will be deployed for Agency usage and ETAP.

Dr. Harrill summarized the presentation by emphasizing the need to address prior data gaps, for personnel trained in utilizing transcriptomics data for human health assessment, for reliable and reproducible wet lab and data analysis workflows, and continued collaboration between domestic and international partners to develop robust reporting frameworks for regulatory review.

Dr. Teeguarden asked where the transcriptomic studies will be performed. Dr. Harrill stated that some of the services will be contracted but EPA has pipelines in place. Dr. Thomas further noted that ETAP will initially be completed at EPA or NTP but longer term may be made into guidelines as demands dictates. Dr. Teeguarden asked how the OECD templates will be incorporated into current EPA transcriptomic data analysis. Dr. Harrill clarified that the templates will supplement ETAP methods.

Dr. Gladys Liehr asked if people outside of the EPA would be able to run ETAP analysis and how that would be evaluated. Dr. Thomas stated that while the initial studies will start at EPA or NTP, the aim is to fill the gap to make them scalable and allow industry to be able to do ETAP themselves. Dr. Post asked if ETAP was going to be incorporated into EPA risk assessment guidance for regulatory standards. Dr. Harrill stated that the EPA has no plans to incorporate them into risk assessment. Dr. Thomas further noted that while IRIS provides a RfD for a particular chemical, ETAP provides a best estimate of a reference value along with uncertainties of the assessment. Dr. Jones stated that the Agency is considering how to incorporate ETAP into other products. However, there is a need to utilize restrictive language to describe these values.

Dr. Teeguarden asked if there are any other elements of the EPA infrastructure that are not transcriptomics that can support ETAP. Dr. Harrill mentioned that there are people in other divisions that work on chemical analysis but, overall, there is a large infrastructure to deal with multiple types of data.

Dr. Katherine von Stackelberg asked if the transcriptomic data will be incorporated into existing databases such as ToxCast and CompTox. Dr. Harrill specified that ETAP is for data poor chemicals; however, once there are more data, ETAP may be incorporated. Dr. Post asked if there are a sufficient number of data poor chemicals in the environment to perform a 5-day, repeated dose study. Dr. Mike Devito stated that one limitation is acquiring a sufficient amount

of chemicals. Dr. Devito stated that the advantage of 5-day study is that you need only one-tenth of the chemical to run the a 90-day study.

Transcriptomics in Dose Response Assessment Literature Review

Leah Wehmas, Genomics Scientist, CCTE

Dr. Leah Wehmas introduced the rationale for applying dose response assessment to transcriptomics data. Reasons to apply dose response assessment include that gene expression changes occur with chemical exposure, changes precede apical adverse effects, changes can be measured and are dose dependent, and changes in gene response can indicate chemical potency.

Dr. Wehmas outlined a 2007 study in which benchmark dose (BMD) modeling methods were applied to transcriptomic data from an acute study of formaldehyde in rats. The study demonstrated alignment of gene set-based transcriptomic and apical effect BMD values from chronic toxicity study. Additionally, a growing number of studies began using and adapting the approach to compare transcriptomic BMD values from short-term studies with apical responses in traditional toxicity studies.

Dr. Wehmas described using a literature search to demonstrate the utility of dose-response approach to transcriptomics. An abstract sifter was developed at EPA that used 54 expert curated papers as seed studies, which were used to find similar studies in PubMed that generated 874 unique articles. After review by two people, there were 214 potentially relevant articles. Further review using the inclusion criteria (*in vivo* studies, mouse or rat, chemical exposures, at least 3 dose levels, BMD on gene expression data, apical endpoint data) resulted in 81 reports and four additional reports from NTP that were not available in PubMed. Of the 85 articles, 32 were primary research, while others were reanalysis, methods, or review papers.

Dr. Wehmas described that once the data were collected, the researchers evaluated important considerations for application of dose-response assessment to transcriptomics, including study duration, chemical modes of action, chemical properties, route of exposure, tissue selection, and gene expression platform.

Dr. Wehmas explained the role of study duration in the transcriptomics data. There was a wide range of studies employed, from 1 to 90 days. For some studies, only the BMD lower confidence limit (BMD_L) was listed, while for others both BMD and BMD_L were listed. For the apical comparisons, they were either concurrent or up to 2 years later. Few studies investigated chronic adverse effects. One study that looked at time systematically was Thomas et. Al. (2013), which examined gene set at 5, 15, 30, or 90 days and showed that the BMD was consistent with the 2-year apical BMD across time. A different study that looked at time consistently examined 79 chemicals sampled over time and demonstrated consistent BMD_L across time points.

From the datasets from the literature search, Dr. Wehmas used 14 studies and a total of 38 chemicals to examine how consistent gene set BMD was with the 2-year apical BMD across time points (5, 15, 30, or 90 days) with high correlations between gene set BMD and 2-year apical BMD. The chemicals in these studies range over toxicity domains, types, and modes of action and include neuroactive chemicals, anticancer agents, endocrine active chemicals, receptor mediated effects and genotoxic and carcinogenic chemicals.

Dr. Wehmas examined the physiochemical properties of the dataset from the literature search to the TSCA active inventory. The properties of the dataset (vapor pressure, molecular mass, water solubility, logKow, Henry's law) were consistent with the median and 10th and 90th percentile with the TSCA active inventory. Dr. Wehmas also examined the impact of the route of exposure on gene set BMD. The route of exposure did not greatly affect the 1-90 day gene set BMD concordance. When examining the impact of tissue type, the target tissue gene set BMD performed better than the 2-year apical BMD in terms of the root mean square difference (RMSD). Dr. Wehmas also investigated the impact of different gene expression platforms (microarray, TempO-seq, RNA-seq) and observed consistency in gene set BMD across platforms.

Dr. Wehmas summarized the findings of the literature search:

- The combined set of 1-90-day gene set BMDs are generally concordant with 2-year apical BMDs
- Gene set BMDs following 5-day exposure showed similar concordance with 2-year apical BMDs as other time points
- Concordance of gene set BMDs with 2-year apical BMDs was robust across route of exposure, physicochemical properties, mode-of-action, and measurement platform
- Gene set BMDs from target tissues (although there is a bias toward liver and kidney) were more concordant with 2-year apical BMDs than surrogate/sentinel tissues supporting the collection and analysis of multiple tissues in an ETAP.

Dr. Fred Wright asked if the studies used tissues with low RMSD values and how that impacted gene set BMD. Dr. Wehmas answered that more studies with those tissue types are needed before making a conclusion about tissue-specific gene set BMD for those tissue types. Dr. Rebecca Fry asked whether EPA can provide more details about how the gene sets were selected. Dr. Wehmas answered affirmatively. Dr. Xiong Gao asked how EPA determined that there were enough papers in the literature search. Dr. Wehmas clarified that the group chose this particular search strategy since it was more focused than other strategies. Dr. Stephen Edwards asked how duplicate articles were identified. Dr. Wehmas answered that they used a global simulated score which identifies papers that have similar terms. Dr. Craig Rowlands asked how the 2-year apical BMD was chosen. Dr. Wehmas stated that it depended on the paper, but generally the group used the most sensitive endpoint reported.

Overview of the NTP Approach to Genomic Dose-Response (GDR) Modelling Report

Scott Auerbach, Division of Translational Toxicology, National Institute of Environmental Health Sciences (NIEHS)

Dr. Scott Auerbach presented an overview on the National Toxicology Program (NTP) approach to genomic dose-response (GDR) modelling that was subject to an expert panel review. Dr. Auerbach first described background on the role of NTP and the Division of Translational Toxicology (DTT) at NIEHS (National Institute of Environmental Health Sciences). NTP is an inter-agency program founded in 1978 that is run by the United States Department of Health and Human Services and is headquartered at NIEHS. The role of the NTP is to coordinate testing across the federal government and is responsible for generating, interpreting, and sharing

toxicological information about potentially hazardous substances. In recent years NTP has become more prominent in the evaluation of new test methods, such as the GDR studies.

Both the NTP and the DTT produce several toxicological assessment products that range from systematic review reports to various specific guideline and non-guideline level toxicological assessments that are used by regulatory agencies across the globe. In recent years there has been an increased demand from our stakeholders to formulate more timely approaches and products that can support risk assessment. These approaches include a number of new approach methods including the in vivo genomic dose response studies that can rapidly identify a transcriptomics/molecular point of departure (POD).

Dr. Auerbach noted that NTP first started examining GDR after the Elk River spill. In 2014, a mixture of chemicals including crude 4-methylcyclohexanemethanol (MCHM) and stripped propylene glycol phenyl ether (PPH) was released into Elk River near Charleston, West Virginia. MCHM was able to travel through the water treatment process and into the water supply and many residents showed signs of chemical exposure and toxicity, such as nausea, skin irritation and headaches. In response to the spill, the Center for Disease Control (CDC) developed a drinking water advisory level for the two primary components of the spill, PPH and MCHM.

While there was more data on PPH, the POD for MCHM was based on a median lethal dose (LD50) value by a relatively poorly documented 28-day study. Through its nomination process, NTP was asked by stakeholders to evaluate several hazards and the POD for the drinking water advisory levels issued by CDC, particularly MCHM, and to do the evaluation in less than a year. This presented the opportunity to evaluate the use of in vivo 5-day GDR testing to rapidly characterize a POD and see how it compared to the POD used in the drinking water advisory level.

Dr. Auerbach described that the results from the Elk River GDR studies concluded that the POD used in the drinking water advisory level of MCHM and PPH were adequately protective. However, the group noticed some challenges in reaching these conclusions, specifically the need to standardize how study design, analysis and reporting are conducted. This led to the formulation of the expert review panel on the NTP GDR Analysis Pipeline in 2017. The expert panel included several experts in toxicology, molecular toxicology, and dose-response modeling. Following the expert panel review, the panel formulated the recommendations into an NTP research report that outlined the NTP peer-reviewed approach to GDR modeling.

Dr. Auerbach then discussed the software approaches that enable GDR testing. Around the time of the Elk River spill, the group had worked on updating a software package known as BMDExpress, which models dose-response data from genomic studies via the use of parametric models contained in the EPA BMDS. In the software, the BMD is based on the genes are groups into annotated gene sets such as Gene Ontology Biological Processes, from which a gene set level potency is determined.

Dr. Auerbach outlined the NTP approach to GDR studies. Dr. Auerbach described the approach as an experimental design that is congruent with a dose-response modeling-based approach to data analysis. First, the study is designed to use a wide range of doses of a select target organ in a test species. Following tissue collection and RNA isolation and sequencing, data quality control (QC) and normalization is performed and the data are then analyzed in a three step process using

BMDExpress. First, the measured features (transcripts) are subject to an overall signal filter followed by prefilter, which is a statistical and effect size filter that removes features that do not respond to treatment. Second, the remaining features are fit to multiple models, and the model with best fit to the data is selected. The best-fit model is then be used to determine potency, in our case BMD, for each feature. In the final step, the features and their associated BMD values are further filtered for adequacy of fit, mapped to genes and passed in pre-defined gene sets. Once the gene sets are populated with dose-responsive genes, criteria are used to determine if those gene sets are considered active in response to treatment. Then, an overall potency for each of those gene sets is determined using an agglomerative metric, specifically a median BMD value that is representative of all the gene level potency values for the dose-responsive genes in each gene set. Finally, transcriptomic POD is declared to be the active gene set with the lowest median BMD.

Dr. Auerbach then described in more detail the NTP approach for each step in the GDR guidelines starting with study design. Dr. Auerbach emphasized there were several differences in how GDR guideline studies differ from traditional guideline studies. In traditional guideline studies, there is greater biological replication with use of only a few dose levels. This design is specifically intended for pair-wise analysis of the data, not for fitting multi-parameter, dose-response models. In the NTP GDR study design, there would be more dose levels with fewer biological replicates, which will allow for better coverage of the numerous dose-response relationships in each study, more confident fits of the data, and greater certainty in the BMD estimates for the features. Notably, there has been empirical validation of this design as it relates BMD modelling of data from both guideline studies and, more recently, for transcriptomic studies.

Dr. Auerbach outlined the design parameters to consider for in vivo studies. Dr. Auerbach advised using male and/or female Sprague Dawley rats since they are utilized for modelling toxicological assessment of human environmental exposure. The group argued the option to use one sex or the other (or both). Any given test article could plausibly be determined *a priori* to be more potent in one sex or the other, but often this is unknown, in which case use of both sexes would be justified. Dr. Auerbach recommended the use of a 5-day exposure with 5 repeated doses because other groups have indicated that transcriptomic POD values from 5-day studies were largely concordant with both sub-chronic and chronic POD values. For organ selection, Dr. Auerbach proposed using liver and other expert/algorithmically selected organs or tissues. The liver was chosen because it is the most common target organ in toxicological assessments and findings are often the basis of POD values particularly with non-drug chemicals that do not have more selective pharmacological effects. Finally for dose selection, the group proposed the use of a 5-day maximum tolerance dose, which equates to the dose that produces a decrement in body weight (BW) gain of 20% over a 5-day period without other clear morbidity or mortality. This ensures a clear response at the highest dose level and the identification of responsive features and improved model fitting.

Dr. Auerbach then discussed the NTP approach to the filtering unresponsive genes in GDR studies. This pre-filter step determines if there is any plausible dose-related signal in the data using a one-way ANOVA with a Benjamini-Hochberg correction at $p<0.05$. If one feature passes, then it is deemed that there is signal in the data. This step is not intended to identify

probes or genes for modeling. The use of a trend test is more appropriate to identify dose-responsive features.

Dr. Auerbach described that after pre-filtering, the group recommends finding the dose-responsive features. To do this, a combination of a trend test statistical filter with a nominal p-value and an effect size filter (i.e., fold change) are used to minimize false discovery rates at the gene set level and maximize true discovery and repeatability of findings. Dr. Auerbach noted that optimal prefilter parameters may differ with technology, sample source, and study design, and that it is essential to perform empirical characterization to identify optimal study parameters.

Dr. Auerbach outlined the next step in the NTP approach, which is fitting features to dose-response curves and determination of feature potency. The NTP utilizes eight parametric, continuous models for curve fitting consistent with EPA guidance. The BMR is then selected based on the 10% tail distribution and the best model is selected based on the lowest AIC value. Once the best model is selected, BMD, BMD_L (lower bound) and BMD_U (upper bound) are determined to provide an overall range of uncertainty for each BMD value.

Dr. Auerbach described the subsequent step in the NTP approach, which is estimating gene set level potencies. This process involves filtering the fitted features for adequacy of fit, transforming the fitted features into genes and parsing them into pre-defined genes sets. For a feature to be considered as part of a gene set, its best model must have convergent BMD, BMD_L and BMD_U values, not map to more than one gene, not have a BMD greater than the highest dose, have a global goodness of fit p-value greater than 0.1, and have BMD_U/BMD_L less than 40.

Once the features are filtered and collapsed into genes, they are parsed into gene sets specifically Gene Ontology (GO) Biological Processes. The NTP utilizes GO Biological Processes because it is the largest curated gene sets, offering the most comprehensive coverage of biological space of all the annotated gene sets and their organization by biological process. For a gene set to be active it must have at least three genes and be at least 5% populated. Active gene set parameters may be platform specific. The gene set potency is then determined to be the median BMD value.

Dr. Auerbach outlined the final step in the process as the selection of a genomic/transcriptomic POD and the biological interpretation. The POD is identified as the active gene set with the lowest median BMD, which generally agrees with apical points of departure. However, until a formalized method for biological interpretation is available, the group recommends refraining from providing potentially misleading biological/toxicological interpretation based on the point of departure gene set.

Dr. Auerbach summarized the presentation on the NTP approach to GDR studies. He outlined the NTP expert panel peer review of its approach to GDR studies, modelling and interpretation; reviewed the approach to designing GDR studies and selecting point of departure; emphasized the need to avoid biological interpretation until there is more formalized process in place; and stated that substantial improvements will be incorporated in future iterations of the analysis pipeline.

Dr. Gao asked how the goodness-of-fit value was chosen. Dr. Auerbach clarified that it was based on the guidance of the EPA. He stated, however, that one of the biggest challenges to the data is using one endpoint to derive the POD and, depending on how you look at the data, some sensitive endpoints may be missed. Dr. Edwards asked what criteria were set before the null set

experiments and if there have been reassessments of the data after these experiments. Dr. Auerbach stated that the optimization is re-analyzed with every new dataset that comes in, so the group is constantly evaluating the parameters. Dr. Edwards asked if the group saw any gene sets that appeared consistently across multiple exposures and if a weighted approach would be applied to those sets. Dr. Auerbach stated that they had not formally run that analysis, but they would not be weighted so that there was an unbiased characterization.

Discussion of Panel Roles and Responsibilities

Katherine von Stackelberg, Co-chair, ETAP Panel

Dr. von Stackelberg stated that the group would break into four working groups. Each group was assigned to discuss the first charge question during their day 1 discussion and then their assigned charge question during the day 2 discussions.

Dose-Response Methods and Parameter Refinement

Logan Everett, Bioinformatics Scientist, CCTE

Dr. Logan Everett provided an overview of the datasets used to develop the scientific support analysis. There were two datasets from NTP with three main analyses. The first analysis involved taking the 5-day in vivo transcriptomic dose response data for 14 chemicals with chronic rodent bioassays and using dose concordance of transcriptional and apical response. The second analysis involved combining vehicle controls from both studies to determine the family-wise error rate (FWER). The third analysis involved a 5-day in vivo transcriptomic dose-response data for three chemicals with three inter-study replicates to determine inter-study reproducibility. All studies were performed on the same platform (TempO-seq rat S1500+) and general design (5-day repeat dose) as the proposed ETAP method, processed using the established EPA TempO-seq pipeline (Harrill et al., 2021) and the same outlier removal process in the standard method document.

Dr. Everett outlined how the targeted RNA-seq assay (TempO-seq) worked. TempO-seq utilized targeted probes hybridized to expressed transcripts. It is able to capture gene expression at a lower cost than RNA-seq or microarrays. The S1500+ probe set was designed to maximize biological coverage with ~2,700 genes. This technology has been applied for high-throughput in vitro screening and other research within ORD.

Dr. Everett provided an overview of the datasets used in the analysis. The first dataset was a 5-day repeat dose exposure in rats following recommendations from NTP Research Report 5 with 14 chemicals with chronic apical BMD established from 2-year study. There were more than eight dose groups per chemical with matched vehicle controls and four replicates per group. The transcriptome was profiled from the liver and kidney in each animal. The second dataset was a 5-day transcriptomic study replicated for three of the 14 chemicals with two additional replicate studies per chemical. All replicate studies performed with the same doses, in the same contract laboratory.

The first analysis performed on the two studies was dose concordance, looking at both cancer and non-cancer endpoints. The two studies were evaluated separately and computed average $\log_{10}$(BMD). Concordance was then evaluated by 48 parameter combinations. The objective was to minimize the overall differences between transcriptomic and chronic apical BMDs.

The analysis workflow for the first analysis was conducted in four main steps. The first step was to evaluate the dataset for adequate signal. The second step involved pre-modeling filtering for dose-responsive probes. The third step involved dose-response modeling of individual probes using eight different parametric models and determining the best fitting model based on lowest AIC. Finally, the last step involved the summarization of BMD_L for the known gene sets. All GO Biological Process gene sets were used. The probes with sufficient fit to the known GO Biological Process were used. The Gene Set BMD_L was determined as the median of all valid gene-level BMD_L within a set and the overall BMD_L was the minimum gene set BMD_L.

Once the transcriptional BMD_L was acquired, concordance between transcriptional and apical BMD was evaluated. For each chemical, the transcriptional BMD_L was the minimum gene set BMD from either tissue (liver or kidney) and the corresponding BMD_L. Concordance was evaluated in terms of the RMSD between the two endpoints. The RMSD examined all 14 chemicals and averaged the difference between apical and transcriptional BMDs. Additionally, the Pearson correlations of transcriptional versus chronic apical $\log_{10}BMD_L$s were also evaluated.

In the concordance analysis, 48 different combinations of analysis parameters were evaluated, focusing on those most likely to be dependent on platform and study design. These included pre-modeling probe filtering such as the values for Williams Trend test (p-value ≤ 0.05 or 0.1) or the minimum fold change (≥ 1.5 or 2), dose-response modeling such as determination of BMR and maximum uncertainty ($BMD/BMDL \leq 20$ or $BMD_U/BMD_L \leq 40$), and how the gene set summarization was performed such as the minimum genes per set (3 or 5) and the minimum percentage of gene set coverage (0%, 3%, or 5%).

Utilizing 13 of the 48 different combinations that ensured sufficient sensitivity, the group produced transcriptomic BMD values for all 14 chemicals. Additionally, RMSD (0.567 to 0.958 log10 mg/kg-d) and Pearson correlations (0.804 to 0.917) were computed for all 13 combinations of BMDExpress parameters. Parameter combinations with high correlation coefficients and low RMSD were ranked as highly concordant. Examining the top five parameter combinations, the group noticed that pre-filtering for probes with maximum fold change > 1.5, utilizing a maximum uncertainty in best-fit model: $BMD/BMD_L < 20$, and having a valid gene set BMD minimum of 3, consistently minimize RMSD. The parameters for William's Trend test p-value cutoff ($p < 0.05$ or < 0.1) and the minimum percentage coverage for valid gene sets (0, 3, or 5%) were more variable and had a lesser impact on RMSD.

Dr. Everett also pointed out that even though BMD_L was used as the reference value, the parameter was refined based on BMD because it is a better measure of central tendency for estimating concordance. The use of BMD_L as the reference value is to be more protective given the uncertainty in the BMD estimate.

In the second analysis, Dr. Everett's group combined the vehicle controls from both studies and determined the FWER. The FWER is the chance that at least one test passes for a dataset. In other words, FWER is the chance of erroneously selecting a TRV when there is no apical effect. To accomplish this, the group randomly sampled to create a "sham" dose-response series and examined if the dataset would pass pre-filtering criteria or generate a valid gene set BMD passing all filters.

To generate the "sham" dose-response, samples were used from 14 distinct studies for a total of 53 per tissue after sample-level QC. The group randomly sampled 36 vehicle control replicates from the same tissue. For each dose-response series, there was a vehicle control group and eight dose groups with four replicates per group. The dose values for each of the 14 chemicals were the same as the first analysis. The group then generated 1000 "sham" dose-response series with at least one probe passing ANOVA at 5% FDR filter were run through the complete workflow to determine which percentage of sham series produced at least one valid gene set BMD_L. Overall, using the top five parameter combinations, the FWER < 1%, suggesting validation that the workflow with these parameters is protective.

The third analysis focused on inter-study reproducibility. To accomplish this, the replicates of the transcriptomic studies within three of these studies completed several years apart within same contract labs were utilized. The overall BMD_L was computed for each replicate study based on the most sensitive gene set in either tissue. The group then evaluated the standard deviation (SD) for BMD and BDM_L based on all the unique pairs of replicate studies for the same chemical. Using the top five parameter combinations from the first analysis, the SD for BMD and BDM_L was negligible between each of the top five combinations.

Dr. Everett summarized that the group combined transcriptomic data for 14 chemicals from multiple studies and performed three analyses to refine and validate their workflow. In the first analysis, they demonstrated concordance between chronic apical BMDs from 2-year studies and refined BMDExpress parameters to minimize RMSD. In the second analysis, the group evaluated FWER using "sham" series and demonstrated a FWER < 1% using all workflow filters. In the final analysis, the group evaluated inter-study reproducibility using replicate transcriptomic studies for three chemicals and demonstrated minimal differences in standard deviations between BMD and BMD_L values.

Dr. Post asked which apical endpoints were included in the analysis and whether they were consistent across studies. Dr. Everett stated that he believed they were consistent. Dr. Post further asked if these endpoints were clinical chemistry or organ weights. Dr. Devito stated that the endpoints were all NTP guideline and chronic endpoints, so clinical chemistry and organ weight do not drive them. Dr. Post asked if the endpoints were histopathological endpoints. Dr. Devito confirmed that they were. Dr. Wehmas stated that this dataset was extracted from multiple studies so a few were chronic assays where cancer endpoints were extracted and some were short-term studies. Dr. Rowlands suggested that this discussion continue in the later portions of the agenda.

Dr. Edwards asked why the FWER was divided by rRMS. Dr. Everett specified that the difference in the RMSD range is different so there is a larger spread of values, and that the concordance to apical BMD is the behind this motivation. Dr. Wright further elaborated that applying a criteria minimum may change the number of gene sets and have a slightly higher rate but tends to be more protective.

Dr. Xiong asked about the inter-study reproducibility and how it was calculated. Dr. Everett stated that each *xy* pair is the same for each chemical and that it is an overall estimate of SD, not a chemical-specific calculation.

Concordance Analysis with Inter-study Variability

Kelsey Vitense, Research Scientist, CCTE

Dr. Kelsey Vitense emphasized that part of the assessment of ETAP has involved evaluating concordance between BMD values from short-term transcriptomic studies and apical BMD values from chronic bioassays. However, inter-study variation exists for both the transcriptomic and apical BMD values and can impact their concordance. Estimating and considering inter-study variability is important for interpreting concordance metrics and our confidence in application of ETAP. Dr. Vitense's group estimated the lower bound of expected Mean Squared Difference (MSD) given inter-study variances for comparison between transcriptomic and apical BMDs with the concordance MSD of the top ETAP model.

Using a plot of apical BMDs against transcriptomic BMDs, Dr. Vitense demonstrated inter-study variability in both axes would cause an observed pair of apical and transcriptomic BMDs to not fall exactly on the identity line, even in cases where BMD matches between the two. This can affect metrics, like MSD, which are used to evaluate concordance.

Dr. Vitense described how the lower bound of MSD was calculated using the sum of apical and transcriptomic variances and summarized notations for the transcriptomic BMD (X_c) and apical BMD (Y_c) for chemical c. Dr. Vitense also outlined that the group assumed that apical BMDs (Y_c) and transcriptomic BMD (X_c) were random variables with means $\mu_X(c)$ and $\mu_Y(c)$ dependent on study design and chemical and constant variances σ_X^2 and σ_Y^2. They also used a random variable Z_c which was the difference between transcriptomic BMD (X_c) and apical BMD (Y_c) for chemical c. The group also assumed that the mean BMD across chemicals μ_Z would be constant. The variance for a chemical $Var(Z_c)$ was equal to the sum of the variances σ_X^2 and σ_Y^2 and the two variances were assumed to be independent. Dr. Vitense then derived the equation used to calculate the expected MSD statistic as the sum of the variance of a chemical $Var(Z_c)$ and the mean squared μ_Z^2 and that the lower bound MSD is the sum of the apical and transcriptomic variances.

The MSD is expected to be approximately equal to the sum of the inter-study variances when apical and transcriptomic BMDs are the same on average across chemicals.

Dr. Vitense then estimated the transcriptomic BMD variance, σ_X^2, using inter-study replicates from three chemicals (bromodichloroacetic acid, perfluorooctanoic acid, furan), three replicates per chemical and each replicated performed in a 5-day transcriptomic study with same doses, in the same contract laboratory over several years. The group estimated the transcriptomic BMD variance, σ_X^2, using inter-study replicates from three chemicals using an estimator that was one-half the mean squared difference between transcriptomic BMD values for unique pairs of replicates for each chemical. The group computed transcriptomic BMD variance estimates across all dose-response modeling parameter combinations considered and used the minimum and maximum of transcriptomic BMD variance estimates to provide a range.

Dr. Vitense estimated the apical BMD variance using mean squared error (MSE) from a multiple regression model, which estimates inter-study lowest effect level/lowest observed adverse effect level (LEL/LOAEL) variance after accounting for study descriptors and utilized the minimum and maximum chronic apical LOAEL variances to estimate the minimum and maximum for variance estimates of apical BMDs.

Using the minimum and maximum variances estimates for transcriptomic BMD (0.015, 0.352) and apical BMD (0.252, 0.265), Dr. Vitense calculated the expected lower bound MSD (0.267, 0.617). The lower bound provides an estimate of the expected MSD if the apical and transcriptomic BMDs are the equal on average, but inter-study variation exists for both BMDs.

The MSD of the top combination of transcriptomic model parameters (0.321) was computed using mean BMD values for chemicals with replicates. However, using mean BMD values for only some chemicals violates the assumption of equal variance across chemicals used to derive the lower bound of expected MSD. In order to provide a fair comparison with the lower bound estimate, the MSD of the top model was computed using all combinations of single replicates per chemical, with the following MSD minimum and maximum (0.285, 0.386).

Dr. Vitense summarized that the group computed a minimum and maximum MSD using single chemical replicates that was between 0.285 and 0.386, which falls in range of the lower bound estimates for expected MSD (0.267, 0.617). Dr. Vitense concluded that the error associated with the concordance between the transcriptomic BMD values and apical BMD values was approximately equivalent to the combined inter-study variability associated with the 5-day transcriptomic study and the two-year rodent bioassay. Additionally, she concluded that the transcriptomic and apical BMD values are highly concordant in the context of inter-study variation in BMDs.

Dr. Wright asked if there was bias in the calculation of the bounds MSD and how this calculation that summarized transcriptomic inter-study variability. Dr. Vitense stated that there are two ways to think about transcriptomic MSD: in this instance the group treated them as variables however, another way is to approach them as a prediction of the apical BMD. Dr. Rowlands suggested that this question be written in. Dr. Post asked how many variables were used in the apical BMD calculation. Dr. Vitense stated that the data will be in supplemental slides.

Summary

Alison Harrill, Associate Director, CCTE

Dr. Harrill provided a summary of studies supporting the development of transcriptomic BMD. She stated that regulatory agencies face challenges evaluating human health risk of thousands of chemicals and mixtures, and that technologies like transcriptomics have the potential to fill data gaps in toxicity testing and human health assessment.

Dr. Harrill reviewed the EPA guidance documents and case studies, and discussed the technical frameworks developed at the EPA. The EPA guidance, white papers and reports have recognized the potential for transcriptomics and identified potential barriers to implementation including technical framework for genomic data analysis, criteria for data submission and presentation, and consistency in methods for analyzing data. These have been addressed by MAQC/SEQC studies on reproducibility, software development, the NTP consensus report on transcriptomic dose response modeling, and the OECD reporting template.

Dr. Harrill discussed the literature review that was conducted comparing concordance between transcriptomic PODs from short-term in vivo rodent studies with apical PODs from traditional in vivo studies. The review suggested that the error associated with the concordance between the

transcriptomic BMD values versus apical BMD values is approximately equivalent to the inter-study variability in the repeated-dose toxicity study itself.

Dr. Harrill also discussed the refinement of the study design and data analysis process to derive PODs based on the recommendations in the peer-reviewed NTP Approach to GDR Modeling report. The transcriptomic dose response modeling follows a stepwise process that utilizes BMD modeling approaches that are commonly employed in chemical risk assessment. A transcriptomic dose-response modeling process was identified, refined, evaluated, and contextualized with BMD variance in apical endpoints.

Dr. Harrill noted that the overall conclusions from the literature survey, evaluation of the transcriptomic dose-response analysis methods, and the statistical comparison of the concordance with inter-study variances support the use of transcriptomic PODs from 5-day, repeated dose in vivo rodent studies in quantitative human health assessments. The barriers that have limited application of transcriptomics have been mostly addressed and the methods have undergone extensive peer-review in individual publications and NTP's report. ORD is proposing to apply these methods in a standardized human health assessment framework to address the substantial data gaps that exist among chemicals that lack traditional toxicity testing data.

Dr. Harrill stated that the Day 2 agenda will cover the ETAP process and discuss a PFAS substance that was taken through the workflow under review.

Public Comment Period

Tom Tracy, DFO, ORD

Bridget Rodgers (PETA) stated her belief that the next steps in the review process are essential, and that all stakeholders should be involved. She stated that in light of TSCA requirements to consider alternatives to vertebrate animal testing whenever possible and given recent advancements in in vitro and in silico transcriptomics techniques, EPA should consider how it can incorporate these techniques in the near-term. She added that to gain confidence in these techniques, EPA should conduct in vitro transcriptomics studies using the same chemicals reviewed in the ETAP documents, including the 140 chemicals mentioned in the transcriptomics literature review and the 14 chemicals from the concordance evaluation. This would then allow these in vitro tests to be the first tier before considering any in vivo testing. The use of high-throughput in vitro and in silico transcriptomics provides the added advantage of most efficiently generating data that can inform environmental and human health protections. Lastly, she stated that EPA should clearly state in its guidance that the intention is not to follow ETAP with a 2-year chronic toxicity test and the information obtained from ETAP could be used to set protections and negate any need to consider the in vivo 2-year chronic toxicity test.

Dr. Clewell agreed that transcriptomics should be incorporated into in vitro testing but stated that there is a data gap between in vitro and in vivo endpoints. Dr. Teeguarden asked if there were any formal suggestions to close these data gaps.

Day 1 Wrap Up

Annette Guiseppi-Elie, ORD

Dr. Annette Guiseppi-Elie thanked the ETAP panelists and the presenters. She emphasized three points to conclude the day's discussion:

- ETAP does not replace other EPA products but is a new product that fills the gap for data poor chemicals.
- The scientific underpinnings of ETAP are sound, and there is a substantial body of research demonstrating that transcriptomic endpoints are concordant with apical endpoints.
- These new assessment products allow EPA to do a broad characterization of a chemical in a shorter period of time. She stated there will be a case study demonstrating this on Day 2.

Facilitated Panel Q&A

Craig Rowlands & Katherine von Stackelberg, Co-chairs, ETAP Panel

Dr. Rowlands thanked Dr. Everett for his presentation on dose-response methods and model refinement and asked about the workflow involved in the selection of the five parameter combinations that provided the highest concordance. Dr. Everett stated that the group focused on parameter choices and as more chemicals are examined, optimization of these parameters is essential. If there is incorporation of a different platform, different parameters may be important. Dr. Thomas stated that the development of these parameters, platform, and study design is based on the NTP report recommendations and refined for ETAP purposes.

Dr. Chris Vulpe asked how a transcriptomic POD can be derived without apical endpoints for a chemical. Dr. Thomas stated that there needs to be a transcriptomic timepoint. Dr. Edwards followed up asking how the chosen parameters were decided, if there was a consensus, and how the final EPA workflow compares to the NTP workflow. Dr. Thomas stated that the group followed the NTP workflow and the parameters decided were based on the NTP report. Dr. Everett stated that the purpose of the parameter testing was to determine if there is a signal in the dataset. At the probe level, both groups are using trend test; the only difference is the p-value cutoffs.

Dr. Fry asked about the document on gene set aggregation, specifically how it drives subsequent testing. She stated that the document suggests that there is a link between biological processes and gene identity. Dr. Everett responded that they are grouping the genes by their biological process, and that BMD modeling is done at the gene level and then group the genes by biological process.

Dr. Vulpe asked if there was an application of a pre-filter. Dr. Everett stated the first test on the probe is an ANOVA with a 5% FDR, a correction for multiple testing. This is followed by a trend test, which is used for BMD modeling and dose-responsiveness and is less stringent than the first test.

Dr. Vulpe then asked how the decision to set a threshold for GO was determined and pointed out that GO is based on limited understanding of biology and may be applied loosely. Dr. Thomas stated that one of the reasons the gene set approach was used was that NTP report recommendations were a good foundation to start on. Dr. Teeguarden asked if this is why the group was confident in their concordance. Dr. Thomas stated that they would not have been able to leverage it without confirmation from the literature, as Dr. Wehmas presented.

Dr. Wright stated that is seems that the transcriptomic response seems optimal and wondered how the chemicals were chosen. Additionally, he asked that, since apical data are needed, how can this move forward in a way that keeps it "evergreen". Dr. Devito stated that the group was interested in some chemicals that are persistent and some that are receptor mediated. Some of the chemicals had no effect in the 2-year bioassay but did change transcriptomics. The group also attempted to look at chemicals that did not produce liver effects, but it was difficult, and they are still missing developmental toxicological endpoints. Dr. Thomas and Dr. Gwinn agreed that progress needs to be made in terms of methodological refinement.

Dr. Vulpe pointed out that there is a potential that transcriptional changes may be indicative of beneficial, not adverse, effects. Dr. Thomas answered that many of the chemicals examined were known to be toxic and that their group is not examining efficacy. Dr. Thomas stated that the effects are generally assumed to be negative and the testing is assessing a bioactivity threshold.

Dr. Devito stated that milk thistle and ginseng were included because they are environmental chemicals but have very large PODs. Dr. Wright stated that it would be interesting to examine the impact of drugs in these assays. Dr. Thomas stated that many of the chemicals were drugs.

Dr. Rowlands asked how a suitable analogue could be found. Dr. Thomas answered that candidates for ETAP are taken based on the absence of a suitable analogue and, if there is a suitable analogue, read across is conducted.

Dr. Post expressed that across the literature search the apical endpoints and durations of exposure vary widely. She emphasized that this information is important so that similar approaches can be applied across different exposures. Dr. Thomas stated that many of the apical endpoints are in line with traditional NTP studies. Dr. Post stated that there needed to be more transparency about the apical endpoints. Dr. Vulpe stated that he would be interested in examining endpoint-specific differences, but there are not enough studies to do so.

Dr. Post asked about how half-life could impact transcriptomic results. Dr. Devito stated that this question was the reason they chose to study PFOA. He also stated that transcriptionally, receptor-mediated mechanisms start out with rapid changes but over time the signal gets dampened and later in time is when the apical endpoint is assessed. He stated that with variable half-lives, the predicted values were still similar to the apical endpoints.

Dr. Teeguarden stated that the panel needed to provide a judgement on the state of science. He stated that while the concordance database is not large, they are at a point where suggestions for the long-term can be provided. Dr. Liehr asked about study design and if the number of animals (3 per dose) is sufficient to detect outliers. Dr. Thomas stated that these numbers are often sufficient, and it is better to have more dose groups with fewer animals. Dr. Devito stated that since the analysis is non-linear, a power analysis cannot be performed. Dr. Auerbach and Dr. Harrill stated that there are often practical limitations to larger sample sizes, including attrition.

Dr. Teeguarden asked if EPA is going to build a case for ETAP. Dr. Thomas stated that HERA will improve the research methods. He also stated that when he presented this to other stakeholders, there was an interest in helping improve it and willingness to provide resources to do so. Dr. Guiseppi-Elie stated that there are other groups trying to become involved. Dr. Teeguarden affirmed that the more the scientific community invests, the more beneficial the product will be.

Dr. Gao asked if functional classifications other than biological process are going to be used. Dr. Everett stated that the group tried to use full gene ontology (cellular component or molecular function), but they did not observe any change in concordance, providing no additional sensitivity. Dr. Thomas reiterated that the gene sets used were building off NTP recommendations.

Dr. Edwards asked how ETAP would be incorporated into human health assessments. Dr. Thomas stated that ETAP will provide a POD, but no interpretation. He stated that human health assessment is inherently different. Dr. Jones stated that while there are differences, there are many core commonalities, and EPA wants to have a transparent, standardized way of assessing human health risk considerations across different products. Dr. Lambert followed by stating that with all EPA products there is a difference in how the POD is generated, but every other step is same.

Dr. Vulpe asked for an elaboration on the choice of platforms (TempO-seq) and how concordance could be applied across platforms. Dr. Thomas stated that cost was the major limitation and why TempO-seq was chosen. Before expanding to other platforms, there will need to be a similar optimization process.

Dr. Teeguarden asked if the 2,700 genes covered in the assay are enough to assess the breath of biological pathways. Dr. Auerbach stated that gene expression is coherent if statistical inference is applied. Dr. Wright asked if, in the literature search, they observed differences in terms of gene selection. Dr. Thomas stated that the gene selection bias did not seem to make a difference.

Wednesday, July 12, 2023

Day 2 Welcome

Chris Frey, Assistant Administrator for Research and Development, ORD

Dr. Chris Frey welcomed the meeting participants back for the second day of the ETAP meeting and thanked the panel for their review. He stated his belief in EPA's responsibility to advance equity and environmental justice. He noted that communities that are disproportionately impacted are also disproportionately impacted by chemical exposure, and that the traditional one-chemical-at-a-time approach to toxicity testing will not keep pace. There is a need to protect the most vulnerable from concern and speed up the pace of TRVs for chemicals of immediate and emergent concern. He thanked the panelists for their contributions and emphasized that ETAP has the potential to be of great value to Agency stakeholders.

Day 2 Agenda and Charge Questions

Rusty Thomas, Director, CCTE

Dr. Thomas thanked Dr. Frey and reviewed the agenda for Day 2 as well as the panel's charge questions. He divided the charge question responses into three categories:

- Tier 1: Recommendations – responses necessary to adequately support scientific basis of the ETAP and implementation as a new ORD assessment product or to improve clarity of the presentation.

- Tier 2: Suggestions – responses for EPA to consider to strengthen the scientific basis of the ETAP and implementation as a new ORD assessment product or to improve clarity of the presentation.
- Tier 3: Future Considerations – Advice you may have for scientific exploration or research to inform future work.

Dr. Thomas then introduced the charge questions. Please refer to Appendix C: Charge Questions for the list of charge questions.

ETAP Overview – Introduce MOPA Example

Alison Harrill, Associate Director, CCTE

Dr. Harrill presented the standard methods document for ETAP. Dr. Harrill outlined the objectives of the presentation:

- To discuss ETAP purpose and applications,
- To introduce the ETAP process, and
- To introduce an example ETAP to illustrate the workflow.

Dr. Harrill stated that the current estimates of the size of worldwide and domestic chemical inventories are substantial, with increasing trends in production and release. However, relatively few chemicals in commerce, the environment, or the human body have traditional toxicity data or human health assessments. The purpose of ETAP is to evaluate potential human health risks when limited or no chemical data are available.

The "Standard Methods for Development of EPA Transcriptomic Assessment Products (ETAPs)" report details the methods used to derive the TRV for use in the ETAP. The TRV is defined as an estimate of a daily oral dose that is likely to be without appreciable risk of adverse effects following chronic exposure. The TRV is intended to protect both the individual and population from adverse effects other than cancer or related to cancer if a precursor event does not occur below a specific exposure level. While a TRV is expressly defined as a chronic value in an ETAP, it may also be applicable across other exposure durations of interest, including short-term and sub-chronic.

ETAP is only intended to be applied to data-poor substances, and ETAP TRVs can be updated to incorporate new data or be retired if traditional toxicity studies and human health assessments are published.

Dr. Harrill then introduced the ETAP process. She stated that the combination of standardized methods is intended to facilitate expedient development and release of the assessment. The ETAP consists of three primary components:

1. Initial database searches and systematic evidence map development.
2. Short-term in vivo transcriptomic study for POD identification.
3. Assessment development and reporting.

In the first component, a database search is conducted to assess if the chemical is data poor. If it is data poor, the assessment proceeds to the second component; otherwise, an ETAP is not

performed. In the second component, a 5-day in vivo rat study is conducted in males and females with target RNA-seq in a variety of tissues and identifies a transcriptomic POD, which is defined as the experimentally determined dose at which there were no coordinated transcriptional changes that would indicate a potential toxicity of concern. In the third component, the transcriptomic POD is used to derive a TRV through the applications of uncertainty factors (Ufs). The results from the systematic evidence mapping, 5-day transcriptomic study, and TRV derivation are compiled and reported in a standardized ETAP reporting template.

Dr. Harrill provided an example ETAP for methoxypropanoic acid (MOPA) and applied the ETAP workflow to demonstrate how this process can be applied to data poor chemicals.

Dr. Clewell asked for clarification on what the cancer and non-cancer endpoints were. Dr. Harrill stated that the group limited the analysis to non-cancer endpoints. Dr. Thomas stated that this was updated from Center for Public Health and Environmental Assessment (CPHEA) guidelines. Dr. Burke stated he had difficulty arriving at a RfD, and that with the TRV there seems to be consistency.

Database Search and Systematic Evidence Map

Avanti Shirke, Center for Public Health and Environmental Assessment (CPHEA)

Dr. Avanti Shirke provided an outline for the presentation: she would first provide an overview of systematic evidence maps (SEMs), then she would outline the utilization of SEM in the context of ETAP, and finally she would outline methods that could be used for the generation of SEMs.

She stated that a SEM is a pre-decisional analysis that uses systematic review methods to compile and summarize evidence but does not reach assessment hazard or toxicity value conclusions. She stated that it is a front-end compilation of evidence used for prioritization, problem formulation and scoping, identifying data gaps, and determining the need for assessment updates. SEMs follow standard methods and practices in CPHEA and ORD.

She then outlined how SEMs were utilized for ETAP. The goal of the SEMs was to assess availability of repeated-dose animal toxicity data if no suitable studies are identified in the ToxVal database (ToxValDB). This search is targeted toward POD and may miss recent studies, assessments and pertinent reviews in the open literature. The next step is to include a broader literature search and read-across analogues. ETAP considered whether repeated dose toxicity studies are available from ToxValDB or the SEM.

She outlined the flowchart used for consideration of ETAP evaluation. Once a ToxValDB search has identified a chemical as data poor, a literature search is conducted without data or language restrictions. Preferred chemical name, CASRN, DTXSID, and synonyms are used as the foundation of the search. Synonyms identified from CompTox Chemicals Dashboard indicated as “valid” or “good” are used. If the number of records retrieved are few (i.e., <200), no further filtering is undertaken. Otherwise, pre-set literature search strategies (“filters”) in SWIFT Review software are used to identify human health content (i.e., human, animal models for human health, and in vitro studies).

The information sources utilized for the literature search are three peer-reviewed databases: PubMed, Web of Science, and ProQuest. Other sources include the “grey literature,” which

includes manual review of reference lists in publicly available draft assessments and studies meeting the inclusion criteria, other databases (e.g., ECHA registration dossiers, EPA ChemView, NTP, OECD Chemicals Database and eChemPortal, EPA ECOTOX database) and searches from confidential business information.

She then outlined the PECO inclusion guidelines for the literature review. The population included both human and non-human mammalian species. She outlined the relevant exposures, which included the chemical and relevant metabolites in oral and inhalation studies. The comparators for human populations were reference populations, case control studies, repeated measures and considerations of other study designs. For animal studies, the comparators were vehicle or untreated control animals. The search included all health outcomes.

The screening phase involved each record being reviewed independently by two screeners at the title and abstract and full-text levels to evaluate "inclusion" or "exclusion" or "unclear" based on PECO guidelines. Studies evaluated as "inclusion" or "unclear" were advanced to full-text review. A specialized systematic review software was used to save time and keep track of screening decisions.

She outlined an example literature review using perfluoro-3-methoxypropanoic acid (MPOA). The search identified five unique studies. However, only one study was human-health related. The information was disseminated using the Health and Environmental Research Online (HERO) database. This resource documents all the references that are relevant to MPOA. Screening decisions can also be tracked using an online content management system for health assessments. This resource keeps track of inclusion/exclusion decisions and allows for display of flow diagrams, in which each node is clickable and available to public for transparency.

She summarized that SEMs are a comprehensive approach to identifying data using systematic review methods. Using the SEM approach ensures a high degree of confidence that no data exist, and ETAP is an appropriate next step.

Dr. Clewell wondered how is adequate data determined, since expert judgment is needed, and whether other agencies will be involved. Dr. Shirke stated that decisions regarding whether a paper meets PECO criteria occurs within the assessment team. Dr. Liehr asked about the response time for the literature review. Dr. Shirke stated the duration depends on the size of evidence base, but with data-poor chemicals the time frame is very short. Dr. Thomas stated that this chemical took about a week and that the whole process is 9 months from chemical procurement. He clarified that a chemical does not enter the SEM workflow until it is already determined to be data poor.

Dr. Edwards asked how data quality was evaluated. Dr. Shirke stated that after a study is determined to be PECO relevant, data quality is evaluated by a team of two or more individuals. Dr. Edwards asked if expert judgment is needed for data quality evaluation. Avanti confirmed that expert judgment is needed and that there are predefined questions that are examined.

Dr. Rowlands asked if the group looks at TSCA submissions. Dr. Shirke stated that these are tracked in EPA Chemview. Dr. Post pointed out that this example was clear case where there were no good-quality, relevant studies and asked how to evaluate situations where there is one good-quality study. Dr. Thomas stated that it is evaluated based on whether there is adequate information to support another type of human health assessment. Dr. Teeguarden stated that

"data poor" is not always well defined and more effort needs to be made to define it at each stage in the process. Dr. Post agreed and stated that it would be good to provide an example where a decision is made at each step. Dr. Thomas stated that they plan to do that but are waiting for this review from the Board of Scientific Counselors.

Dr. Clewell asked how to evaluate epidemiological studies for PECO when exposure is determined in the blood along with other chemicals. Dr. George Woodwall stated that study quality is evaluated after the initial screening. Dr. Woodall brought up concerns related to the presence of low-quality studies. Dr. Jones agreed that it is necessary that the assessment tools are transparent and objective. Dr. Woodall mentioned that there could be a scenario where data are collected to derive a PPRTV, but the data are insufficient. In this scenario, the chemical should be considered for ETAP. Dr. Edwards and Dr. Clewell stated that these methods only decide if ETAP is completed, not whether it is necessary, and warned it could end in a scenario in which an ETAP is created when it is not necessary.

Dr. Post addressed that studies may not be appropriate for dose response. Dr. Wright stated that there also needs to be a flowchart to determine study quality or clarification around what determinations result in a chemical being identified as data poor. Dr. Michelle Angrish noted that the purpose of SEM is to make evaluations of data availability without interpretations. Dr. Teeguarden and Dr. Jones stated that there needs to be more clarity in the documentation on SEM.

In vivo Study Design

Leah Wehmas, Genomics Scientist, CCTE

Dr. Wehmas stated that, as a follow-up to SEMs, the next step in the workflow is the in vivo study design stage. The in vivo study design covers analytical chemistry, dose formulation and selection, study design, tissue collection, and RNA isolation and sequencing.

In the first step, the chemical of interest is procured commercially, by synthesis, or through a reliable third party. Ideally, the purity would be greater than 95%. Dr. Teeguarden asked about low purity (20%) chemicals. Dr. Thomas stated that at the moment there is no lower bond. Dr. Teeguarden stated that there should be more criteria to define purity. Dr. Wehmas stated that in the MOPA example, purity was greater than 98%.

After verifying purity, dose formulation and dose selection are the next steps. To pick an appropriate vehicle, solubility and stability are assessed visually or analytically. The highest dose is based on the rat tolerability test or in silico prediction, and there are recommendations for ≥ 5 dose-levels. For the dosing, half-log spacing is utilized for all doses except the lowest positive dose, which is a full log scale lower. The aqueous solutions are also tested for pH. Dr. Gao asked for the definition of an acceptable pH. Dr. Wehmas answered that recommendations were based on veterinarian assistance to determine a tolerable pH. In the instance of MOPA, it was stable in water with less than 10% lost in the dose solution. The highest dose was set to 60 mg/ml with 9 doses selected. The solution was acidic, so adjustments were made to the pH.

In the second step, the group conducted a 5-day repeated dose study. Dr. Wehmas discussed the housing conditions of the animals and study design. Sprague Dawley rats were acclimated and randomized to a treatment group. On days 1-5, BW was recorded daily, the animals were exposed at 5 to 10 ml/kg via oral gavage, morbidity and mortality was recorded twice daily, and

rats that showed overt toxicity were removed. For MOPA, 5 animals per sex/dose-level were assigned to the treatment group, and 8 animals per sex were assigned to the vehicle control group. A few animals died, and two of the highest-dose females had breathing problems. On day 6, BWs were recorded, and the rats were euthanized. Tissues were collected (liver, brain, adrenals, kidneys, heart, thymus, lungs, spleen, thyroid, blood, testes, ovaries, and uterus). For MOPA, there were three aliquots of tissue collected (2 in RNALater and 1 snap frozen), and 6 aliquots of blood were collected. The mean BW change was tolerated for most animals. Although there was a decrease in the highest dose group.

In the third step, RNALater tissue was used for RNA isolation and evaluated for RNA quantity and purity. Currently, no specific recommendations are made for RNA purity or integrity. Targeted RNA-sequencing using BioSpyder TempO-seq rat S1500+ is used with ≥ 1 million mapped reads/sample.

Dr. Wehmas summarized the in vivo study design. The group developed a standardized design for the 5-day *in vivo* studies for the ETAP using MOPA, verified at >98% purity by liquid chromatography with tandem mass spectrometry (LC/MS/MS). MOPA was formulated in deionized water with <10% loss and delivered daily to male and female rats via gavage at nine dose levels for 5 days. Following 5 days of exposure, 12 tissues were removed, RNA isolated, and targeted RNA sequencing performed using the BioSpyder TempO-Seq rat S1500+.

Dr. Clewell asked about the rationale behind the dose spacing. Dr. Wehmas answered that the lowest dose was still high, and using the full log dose allows the group to get a high and low statistical bounds. Dr. Thomas followed up by stating that, for BMD modelling, it is important to anchor the high-dose and low-dose bounds. Having no signal below the lower dose allows for the definition of BMD_L and a robust estimate for POD. Dr. Clewell asked about chemicals where LOD will be an issue and whether blood was examined. Dr. Thomas explained that blood is not one of the 12 required tissues for an ETAP. They collected blood for dose measurements analysis but states that collecting blood for a toxicokinetic study is an option.

Dr. Post followed up asking if the group was examining blood levels of the chemical. Dr. Wehmas stated that blood was collected to analyze dose measurements. Dr. Thomas stated that derivation of the POD was based on the administered dose. Although it is possible to perform toxicokinetics (TK) and derive POD with these levels, POD using administered dose is acceptable. Dr. Post stated that the optional TK studies are not defined. Dr. Thomas answered the group wants to keep it optional since POD is based on the administered dose. Dr. Post stated that there should be clarification on how TK is defined in the report.

Dr. Wright stated that there needs to be clarity on how the minimum number of doses was selected and what "randomization by weight" means. Dr. Wehmas stated that the randomization by weight means animals were randomly assigned by weight and that there was no statistically significant difference by weight. Dr. Thomas stated that the standard in vivo design would arrange a study to define the highest dose; however, these are data poor chemicals. He stated that the dose selection and number of doses is outlined in the methods document but more clarity is needed there.

Dr. Edwards asked about the lowest possible dose describing where nothing is found transcriptionally. Dr. Thomas clarified that range finding is performed to find the highest dose.

With at least five dose levels utilized, then dose levels drop a half-log until the lowest dose, which is a full log below the second-lowest dose.

Dr. Vulpe asked about the choice of tissues and why blood samples were not included. Dr. Wehmas stated that there were no recommendations in the documents. She stated that the group used entire tissue except for liver and lung (left lobe for liver and right lung).

Transcriptomic Dose Response Analysis

Logan Everett, Bioinformatics Scientist, CCTE

Dr. Logan Everett described how the transcriptomic dose-response analysis was performed after the RNA sequencing from the 5-day in vivo study. The analysis involves two main steps: (1) data pre-processing and QC and (2) dose response modeling and summarization.

Dr. Everett discussed the ETAP report on MOPA. He stated that the group tested 9 doses ranging from 0.01 to 300.0 mg/kg-day. A total of 104 animals were profiled (52 males, 52 females). Transcriptomic profiling was performed for both sexes on a total of 12 tissues using TempO-seq S1500+ covering 2,600 genes.

To perform data pre-processing and QC, the raw sequencing reads were aligned to known probes sequences using reproducible pipeline and QC procedures utilized across multiple EPA/CCTE projects.

After alignment, the sample-level quality metrics were computed and quality was assessed. Samples were removed if (1) the total mapped reads were < 10% of target depth, (2) the percentage reads uniquely aligned was < 50%, and (3) the total number of probes with 5+ reads was < 1,200 probes. No signal distribution cutoff was applied in this QC but the minimum number of probes capturing 80% of the total reads (N_{80}) was evaluated and used in outlier review. For MOPA, seven out of 1092 samples (< 1%) failed to yield high quality TempO-seq data.

After performing the sample-level quality metrics, samples that fail initial QC are removed and all proves with a mean count < 5 are removed across same study/tissue. The counts are then normalized to $\log_2$(CPM+1) for BMDExpress input.

Following normalization, principal component analysis (PCA) for each study, tissue and sex to identify outliers. Strong outliers, defined as samples where the distance to all other samples >2x the span of all other samples (PC1 or PC2), were always removed. Moderate outliers were removed if (1) vehicle sample was outliers on both vehicle PCA and dose-response PCA, (2) QC metrics were below matching samples, (3) outliers were in both PC1 and PC2 with relatively large Euclidean distance, or (4) there was a larger distance from corresponding replicates or adjacent dose groups. A total of 10 out of 1085 samples were removed based on PCA outliers.

Dr. Everrett summarized the QC findings for MOPA. A majority of samples were above the target depth (1 million mapped reads). Samples below 10% target depth (100,000 reads) were removed. A majority of samples mapped >80% of the reads, and samples were removed if they mapped <50% of the reads.

Dr. Everett then described the dose response modelling. First, the dataset was evaluated for adequate signal for each tissue and both sexes. All tissues were profiled passed an ANOVA test

for at least one probe. Following this evaluate, a pre-model filter was used for dose-responsive probes using a William's trend test with a p-value < 0.05 and |Fold-Change| > 1.5. After pre-filtering, dose-response modeling was used for individual probes with 8 different parametric models. The best-fit model was selected for each probe based on AIC. Finally, a BMD_L value was generated using Gene Ontology Biologic Process (GO:BP). The Gene Set BMD_L was defined as the median gene-level BMD_L and the overall BMD_L was defined as the Minimum Gene Set BMD_L across all tissues for both sexes.

Dr. Everett outlined several considerations when arriving at a POD. If there are multiple gene sets that have the lowest overall BMD_L, evaluate which gene set has lowest median BMD_L. "No value" ETAP can be declared if no tissue produces a valid GO:BP class or the most sensitive GO:BP class BMD is greater than 3-fold below the lowest positive dose.

Dr. Everett described the POD for the MOPA study. The lowest BMD in males was observed in heart with BMD=2.8 mg/kg-day and BMD_L= 0.893 mg/kg-day. For females, the uterus was the most sensitive tissue with BMD=0.872 mg/kg-day and BMD_L=0.121 mg/kg-day.

Dr. Everett summarized his transcriptomic dose-response analysis. The group presented a standardized transcriptomic dose-response analysis and POD identification. In applying the ETAP process to MOPA, 1,075 of 1,092 samples (98 %) passed all quality control criteria. All tissues passed initial criteria for dose-response modeling. The uterus was the most sensitive tissue (lowest valid gene set BMD) with median BMD=0.872 mg/kg-day and median BMD_L=0.121 mg/kg-day used to compute the human equivalent dose (HED).

Dr. Wright asked if changing the threshold for the number of reads (from 10 million reads to 1 million reads) would affect the BMD_L. Dr. Everett stated that it did not affect the BMD_L. Dr. Wright stated the change should change the variation. Dr. Thomas stated that this is under the assumption that technical variability is greater than animal variability. Dr. Wright stated that it should move the BMD to the left and reduce variation. Dr. Everett stated that this is not necessarily true, since the BMD will move to the left, but the confidence interval will be tighter. He stated that technical variance was not the main driver of effects compared to biological variance.

Reference Value Derivation and Reporting

Jason Lambert, Supervisory Toxicologist, CCTE

Dr. Jason Lambert first outlined the methodological workflow to derive a TRV, which occurs after a POD has been identified from the 5-day study.

Dr. Lambert provided a definition of POD for human health assessment as the dose-response point that marks the beginning of a low-dose extrapolation. This point can be the lower statistical bound on a dose for an estimated incidence or a change in response level from a dose-response model (e.g., BMD), or a LOAEL or no observed effect level (NOAEL) for an observed incidence, or change in level of response.

The transcriptomic POD is defined as the administered dose at which there were no coordinated transcriptional changes that would indicate a toxicity of concern. The transcriptomic POD is not associated with a specific hazard and does not necessarily discriminate between non-cancer or

cancer effects, adverse or adaptive responses, nor is it used to infer mechanism or mode of action.

In the traditional EPA human health risk assessment practice, candidate PODs from experimental animal studies must be converted to a HED. This involves physiologically-based toxicokinetic modelling, the chemical-specific cross-species TK data, and cross-species allometric BW scaling however this process cannot be accomplished with data-poor chemicals.

Dr. Lambert then described the method for cross-species dose conversion for transcriptomic POD. Using a terminal BW (for humans it has been revised up to 80 kg), the BMD_L is calculated using the following equation:

$$BMDL_{HED} = BMDL \times DAF = BMDL \times \frac{BW_{Animal}^{1/4}}{BW_{Human}^{\frac{1}{4}}}$$

where, BMD_L = animal assay-based POD and,

DAF = dosimetric adjustment factor.

For example, the chemical MOPA has the lowest overall BMD_L of 0.121 mg/kg with a terminal weight for rats as 0.277 kg and humans as 80 kg, which derives the $BMD_{L(HED)}$ of 0.0279 mg/kg per day. The $BMD_{L(HED)}$ was identified as the POD for the derivation of a TRV for MOPA.

To derive the TRV, the transcriptomic POD uses a standard set of Ufs which are the 5 areas traditionally used for human health risk assessment. A standardized set of Ufs are proposed due to the carefully prescribed design of the animal studies and data analysis procedures. The Ufs considered in a ETAP are intended to account for:

1. Uncertainty in extrapolating animal data to humans (i.e., interspecies uncertainty) (UF_A).
2. Variation in susceptibility among members of a human population (i.e., inter-individual or intraspecies variability) (UF_H).
3. Uncertainty in extrapolating from a LOAEL rather than from a NOAEL (UF_L).
4. Uncertainty in extrapolating from data obtained in a study with less-than-lifetime exposure (i.e., extrapolating from subchronic to chronic exposure) (UF_S).
5. Uncertainty associated with an incomplete database (UF_D).

Using these Ufs, the TRV is calculated as follows:

$$TRV = \frac{BMDL_{HED}}{UF_A \times UF_H \times UF_L \times UF_S \times UF_D}$$

Dr. Lambert then described each of the Ufs in more detail.

The animal-to-human interspecies uncertainty factor (UF_A) accounts for the extrapolation of laboratory animal data to humans, and it generally is presumed to include both TK and toxicodynamic (TD) considerations. For TK, the cross-species scaling is applied to convert animal POD to POD_{HED} and for TD, sufficient comparative cross-species are not expected to be available for an ETAP. In the derivation of a TRV, a UF_A of 3 is applied in conjunction with calculation of a POD_{HED}.

The intraspecies variability uncertainty factor (UF_H) is used to account for the possibility that the evidence considered is not representative of the dose-response relationship in subgroups of the human population that are most sensitive to the health hazards of the chemical being assessed. In the derivation of a TRV, a UF_H of 10 is applied.

The uncertainty in extrapolating from LOAEL to a NOAEL (UF_L) is traditionally applied when the lowest tested dose in a given study induces a statistically and/or biologically significant change in anatomy/physiology, compared to control(s). BMD modeling is used in the ETAP to identify transcriptomic PODs. In the derivation of a TRV, a UF_L of 1 is applied when a GO:BP BMD_L value is successfully identified for one or more classes using the ETAP method.

The subchronic to chronic uncertainty factor (UF_S) is applied to the derivation of a chronic reference value if the toxicity data are based on a subchronic duration study. Observed differences in 5-day transcriptomic and chronic apical BMDs are largely driven by inter-study variability, rather than systematic differences. In the derivation of a TRV, a UF_S of 1 is applied.

The database uncertainty factor (UF_D) is intended to account for the potential for deriving an under-protective reference value as a result of an incomplete characterization of the chemical's toxicity. A complete database of toxicity information is not expected to be available in a ETAP so an UF_D of 10 is applied. Using the equation for TRV, the TRV for ETAP is calculated as:

$$TRV = \frac{BMDL_{HED}}{Composite\ UF\ (300)}$$

Dr. Lambert then used MOPA as an example for calculating TRV:

$$TRV = \frac{0.0279\ mg/kg\ -day}{300} = 0.00009\ mg/kg - day$$

Dr. Lambert then compared the RfD for PFPrA, PFBS and GenX.

Dr. Lambert described TRV in more detail. A TRV is defined as an estimate of a daily oral dose that is likely to be without appreciable risk of adverse effects following chronic exposure. The TRV is meant to protect both the exposed individual and population from effects other than cancer, or related to cancer if a necessary key precursor event does not occur below a specific exposure level. While a TRV is expressly presented as a chronic value in an ETAP, it may also be applicable across other exposure durations of interest including short-term and subchronic.

Dr. Lambert discussed how ETAP information is reported. The ETAP will be provided in a highly standardized assessment template with a minimal, free-form text with no subjective interpretation. The report will have six sections: (1) Background, (2) Assessment Review, (3) Chemical Identity and Physical Properties, (4) Literature Survey, (5) Animal Study, (6) HED and TRV. Additional detailed results for each study will be provided in the appendices.

Dr. Lambert described the quality assurance (QA) and technical review process for ETAPs. After there is a draft of the ETAP, an audit of data quality will be evaluated for how the study was conducted and, in parallel, the draft will undergo review by at least two scientists at ORD.

For QA, all ETAP activities and testing covered under a standard EPA Category A Quality Assurance Project Plan, and each individual ETAP will undergo independent Audit of Data Quality (ADQ) by the CCTE QA team. In the ETAP ADQ process, the final reports are received

from EPA contractors, and the CCTE QA team evaluates documentation using the ADQ checklists developed for each process, documenting analytical chemistry, the animal study, and the transcriptomics.

Dr. Lambert presented how QA was performed with MOPA. For MOPA, the team found that the review, verification, and validation processes for the analytical chemistry support were methodical, detailed, and thoroughly documented, the chain of custody documentation was complete throughout the MOPA ETAP, and documentation generally flowed well between multiple contractors and EPA staff. Four calculation errors were identified, corrected and noted as points of emphasis for future studies. The processes generating each of the key documents should undergo periodic audits to ensure the accuracy of the key reports.

Dr. Lambert then briefly discussed the technical review process. Dr. Lambert stated that each ETAP will be reviewed by a minimum of two ORD technical experts. Following successful completion of ADQ and technical review, individual ETAPs will be publicly posted to an EPA website.

Dr. Liehr asked if there is ever a point in the QA process where a ETAP study will produce a bad product and the study needs to be redone. Dr. Lambert stated that the QA process is meant as documentation in the report not evaluation of the study.

Dr. Edwards asked how the Ufs were applied at each step and if there was script for them. Dr. Thomas stated that this number is doubled checked by QA but if the approach is operationalized, there could be a script. Dr. Edwards and Dr. Thomas both agreed that if the process were standardized into a pipeline there could be a reduction in the error.

Comparison of Transcriptomics Reference Doses (TRVs) with Reference Doses/Concentrations (RfDs/RfCs) and Summary

Alison Harrill, Associate Director, CCTE

Dr. Harrill stated that formal statistical evaluation of concordance between TRV and RfD has primarily focused on BMD results. However, since the reference value is ultimately used to evaluate chemical risks, comparing traditional RfD and TRV values provides some understanding of the relative level of protection afforded by the ETAP. One finding the group noticed was that critical effect in four of the seven chemicals occurred in species other than the rat, which is the species utilized in ETAP. For six of the seven chemicals, the TRV was lower than the RfD or provisional RfD (p-RfD), with PFOA as the only chemical with slightly higher TRV.

Dr. Harrill also evaluated an additional set of 20 chemicals was identified from the literature review. A subset had multiple time points, tested species, or tissues with reported transcriptomic POD values. The transcriptomic POD values were adjusted to HED using default BWs for species, strain, and sex used in the study. While study designs and BMD analysis were not standardized across literature surveyed, a composite UF of 300 was used to calculate a TRV across all chemicals.

A total of 28 of the 47 (~60%) combinations had TRVs that were more sensitive than the RfD/reference concentration (RfC). The Median Absolute Ratio was 2.3 ± 1.1 (median absolute

deviation, MAD); Median Absolute Ratio (Non-Matched Species) were 3.2 ± 1.3 (MAD); and, Median Absolute Ratio (Matched Species) = 1.5 ± 1.1 (MAD).

Dr. Harrill then discussed comparisons with the PBDEs. The maximum absolute ratio was 59-fold for 2,2',4,4'-tetrabromodiphenyl ether (BDE-47). However, the RfD for BDE-47 used a composite UF of 3,000 to account for database uncertainties. By comparison, the absolute ratio between the TRV and RfD for 2,2',3,3',4,4',5,5',6,6'-decabromodiphenyl ether (BDE-209) was only 1.64-fold and BDE-209 used a UF of 300. In addition to the bromodiphenyl ethers, the TRV value for naphthalene was approximately 19-fold higher based on the mouse lung compared with the RfC.

Dr. Harrill summarized the MAD values. Overall, the results of the analysis suggest that the TRV provides a similar level of protection relative to the traditional RfD, p-RfD, and RfC values.

Dr. Harrill stated that the ETAP provides a standardized process for systematic evidence mapping, in vivo exposure, sample collection, and RNA-seq analysis, transcriptomic dose response modeling, POD identification, and application of standard uncertainty factors to derive a TRV. Comparison of transcriptomic toxicity values with traditional RfDs demonstrated similar levels of protection across a broad range of chemicals and effects.

Dr. Teeguarden asked for clarification on the concordance with RfDs. Dr. Thomas stated it was meant to give regulators more confidence in the comparison to RfDs. But scientifically they wanted to compare to the PODs. Dr. Edwards stated that it was provided in the methods, but not the scientific support document. Dr. Thomas stated that it was because the scientific support document stops at the POD.

Dr. Post stated that the process that used traditional data had similar results. She also stated that there is a PFOA draft document based on human data that is three orders of magnitude lower. The candidate RfD was based on the 2016 assessment, and the new animal reference values that are candidates are comparable. Dr. Thomas stated that was a good reference and that EPA is still developing the report and reference values.

Questions from ETAP Panel on ETAP Method

Craig Rowlands and Katherine von Stackelberg, Co-chairs, ETAP Panel

Dr. Frye stated that since diets and water often have chemical contaminants, she suggested that the water and food should be analyzed, if possible. Dr. Edwards asked about the use of read across in the ETAP process. Dr. Jones stated the definition of data poor usually means that read-across methodology is not able to be applied. However, EPA is considering incorporating read across more broadly.

Dr. Burke stated that non-targeted analysis is important for identifying environmental contaminants and asked about documentation on analytical detection in non-targeted analysis. Dr. Thomas stated that non-targeted analysis was not included in these reports. Dr. Burke asked if they planned to incorporate non-targeted analysis in the future. Dr. Thomas stated that it would largely depend on demand.

Dr. Burke also asked about mixtures analysis. Dr. Thomas answered that he believed ETAP is amenable to the mixture process. Dr. Guiseppi-Elie stated that there are programs for

contaminants of emerging concern, so there is ongoing work on chemicals beyond PFAS. Dr. Lambert stated that one area of improvement for ETAP and the whole EPA portfolio is rapid assessment for emerging contaminants, whether it be read across or other sources.

Dr. Post stated that for non-targeted analysis some of the chemicals may not be available and may need to be synthesized. She stated that for PFAS, or other chemicals found in an industrial source, it would be nice to have the company provide them for ETAP. Dr. Thomas stated that this would require a legislative change and chemicals cannot always be easily or cheaply synthesized.

Dr. Wright asked about the price comparison between an ETAP study and a 90-day or 2-year chronic assay. Dr. Thomas stated that it would depend on a number of factors, including the cost of the chemical, analytic chemistry, and dose formulation. However, if the cost of the chemical is removed, it is about $150,000 to $200,000 with sequencing costs, though these costs will decrease over time. He stated that a 90-day study will cost around $250,000, while a 2-year study will cost around $1,000,000 to $4,000,000.

Dr. Wright stated that it seems the 90-day study provides more information than ETAP. Dr. Thomas stated that the cost per chemical for a 5-day study is much less than a 90-day or 2-year study. The costs increases considerably when chemicals become harder to acquire. Additionally, he stated that the time to conduct these studies is also an important factor to consider.

Dr. Post stated that a benefit of the 90-day study is the ability to know the cause of the toxicity. Dr. Teeguarden stated that they have demonstrated concordance with 90-day studies. Dr. Post agreed that the PODs are equivalent but the source of toxicity may be different.

Dr. Gao asked, since they did not check RNA sample quality, would the group examine FASTQ sample quality. Dr. Thomas stated that they would check FASTQ quality. Dr. Gao followed up asking why RNA quality would not be evaluated before sequencing, since low quality data are produced when the initial RNA quality is low. Dr. Thomas stated that some samples fail FASTQ quality even if they have a good RNA integrity number (RIN) and that this eliminates two QC checks.

Dr. Wehmas stated that the Agency does evaluate quality but did not provide any criteria for exclusion. Dr. Harrill stated that some tissues have a higher RIN number. Dr. Vulpe asked why the ETAP did not consider quality if a check was conducted. Dr. Auerbach stated that the RNA isolation is a specific time cost. Dr. Everett stated that they still use the data, but do not set strict cutoffs. Dr. Thomas encouraged suggestions and recommendations for criteria to use, but noted that time and other considerations will need to be taken into account.

Dr. Rowlands asked, if the group found all the useful criteria, whether they could expand the analysis to other chemicals and reinforce with more transcriptomic data. Dr. Thomas stated that the limiting factor is the transcriptomic data. Dr. Wehmas stated that the literature search was limited to mouse and rat data. Dr. Teeguarden asked if this meant choosing a small number of specific probes (1 gene) that optimize concordance, noting, however, that this would lead to noisy results. Dr. Clewell clarified that it was not one gene being predicted, but genomic responses (i.e., gene sets). Dr. Fry stated that this was her statement yesterday: that gene sets connect genes to their identity. Dr. Thomas stated that a lot was borrowed from the NTP report.

Dr. Teeguarden asked about genes with conflicting probes. Dr. Everett stated that there are very few conflicting probes assigned to the same gene. Dr. Edwards mentioned that the false discovery rate picked up things that are low by chance, and the gene set increases confidence in the noise when looking at the coherence in the pathway, so that real biology can be detected with higher power than looking at just genes. Dr. Teeguarden asked if the EPA ETAP team could explain that in the supporting documents.

Dr. Wright stated that when examining overall transcriptomic change, there is a chance that there are activated pathways randomly and that statistical research should investigate how random genes may activate gene sets. Dr. Edwards stated that while some GO processes are not specific, GO does allow you to compare how GO sets change at different time points. Dr. Clewell added that pathways are an effective way of grouping and examining gene networks.

Dr. Wright asked about how Ufs were applied and how that would change if they used data from sub-chronic to chronic. Dr. Post stated the UF would be less but that the group is using a more stringent UF. Dr. Clewell added that they are using database Ufs for sub-chronic to chronic. Dr. Post added that TK can also influence UF calculations. Dr. Wright wondered if the same approach to Ufs would be different to apical studies. Dr. Clewell stated that he was glad they considered a UF for lack of database. Dr. Post believed that the main data gaps were lack of RfDs for developmental, reproductive, immune, or neurological effects.

Dr. Vulpe asked what the rationale was for Ufd of 10 versus 3. Dr. Lambert stated that there is a guidance document and that a Ufd of 10 was used since there is no existing data for ETAP. Dr. Vulpe asked if the UF was to account for what was not tested. Dr. Post stated that it is because developmental or immune effects could occur at a lower level. Dr. Jones said that it was a science-based, not policy-based decision, for what Ufs to apply and wanted to hear from any committee members that would like to weigh in on those details.

Dr. Lambert stated the ETAP method document outlining Ufs is highly consistent between both apical and transcriptomic endpoints. He also believed that transcriptomic signatures should relate to a phenotypic outcome. Dr. Vulpe stated that the three gene threshold was based on expert opinion but it may be valuable to do internal validation. Dr. Burkey asked if the team cross checked against Agency guidance that they have met BOSC review requirements. Dr. Thomas stated that he believed they have, and that ETAP was already based on peer-reviewed methods.

Dr. Vulpe asked about small technical details such as how the PCA was performed. Mr. Tracy stated that he would compile these questions and coordinate communications to the appropriate groups. Dr. Clewell asked about why the document mentioned cancer effects. Dr. Thomas responded that Dr. Clewell should expand on that in his response.

Dr. Edwards asked what the criteria would be for the addition of other tissue and what the costs would be for a comparable in vitro study. Dr. Thomas stated that it would depend on the cell types, and whether primary cells or cell lines are used. Dr. Clewell stated that the animals will be a big cost. Dr. Thomas added that in terms of tissue selection it is about the biological processes we want to cover.

Dr. Liehr asked what justifies the 9-month timeline. Dr. Thomas said that this comes post-procurement and it is an estimate based on an earlier study. Contracting can have certain delays or pitfalls to set things back. Dr. Thomas stated that if ETAP becomes a high demand product,

efficiency could be improved. Dr. Devito stated that they have been doing quick tolerability to determine the top dose but the chemistry is the difficult part (solubility, pH, etc.). Dr. Post asked how long it takes to produce a report once animals are dosed. Dr. Devito stated that it takes 2 weeks to dose males and females, 2-3 weeks to plate the tissues and prepare the RNA for sequencing, and about 2 months for sequencing itself.

Charge Question Discussion & Report Out

Craig Rowlands & Katherine von Stackelberg, Co-chairs, ETAP Panel

Charge Question 1

Given the literature review and the data analysis presented in the documents, please comment on whether the approach outlined for transcriptomic benchmark dose analysis and gene set summarization following a 5-day in vivo exposure are clearly described and provide a scientifically supportable estimate of the point-of-departure for chronic toxicity for data poor chemicals. (Topic primarily covered in Day 1)

Dr. von Stackelberg thanked everyone and initiated the discussion of charge question 1. The group suggested changing the wording of the question because it was too broad. The group believed the approach was reasonable and scientifically sound. If it is sufficiently representative of the concordance data, the BMD_L effort is significantly sufficient. There were questions about how statistical significance was established.

The group was looking for stronger justification that ETAP works better than the bioassay. Additionally, there was a call for more precise language. Several examples include:

- clarification around the term "sensitive," specifically when discussing gene set analysis,
- definition of median BMD of gene set, and
- consistency in the use of POD.

The group also stated that there needed to be more transparency in the description of the apical endpoints. There was also a recommendation to provide a graphic to illustrate the rationale of dose spacing. The group also wanted the discussion of TRV to be less cancer centered. It was stated that PFOA needed a footnote, since the basis of the drinking water maximum contaminant load is different than what is shown in the table.

As the use of this database grows, it would be beneficial to do a retrospective analysis. Additionally, this approach could be applied to data-rich chemicals to provide a confirmatory analysis. A future consideration is to bring in epidemiological or human health assessment data, perhaps by using human cells for confirmation. There was an additional suggestion of cutting down the number of tissues based on need.

Dr. Teeguarden wanted to reinforce the fact that most of the recommendations are asking for clarification and reflect that the EPA team has done a great job of building the scientific justification and description.

Charge Question 2

EPA has proposed standard uncertainty factors to account for intraspecies variability (UF_H), interspecies differences (UF_A), database limitations (UF_D), duration (UF_S), and LOAEL-to-NOAEL extrapolation (UF_L) in the standard methods document. Are the uncertainties in the derivation of the reference values clearly described, and are the uncertainty factors scientifically justified? (Topic primarily covered in Day 2)

Dr. Post led the discussion of charge question 2. One recommendation was the revision of the document so that it states that Ufs can be increased or decreased because it might be more appropriate but recognizes that it may need to be peer reviewed. It was discussed that similar chemical structures can be evaluated, such as read across. There were additional questions about Ufs. The group asked if there could be an application of an animal-to-human UF. Additionally, there was a strong recommendation for incorporation of TK in UF derivation.

Dr. Clewell stated that they want to indicate disproportionate human or animal bioaccumulation, hoping that read across could help with that. Dr. Teeguarden stated that they applied read-across methodologies to a small subset, but that the group does not yet recommend it since there is not a standardized method. Dr. von Stackelberg stated that read across could be run and one could immediately recognize two things: a very long half-life and rate of excretion. Dr. Gwinn stated that novel chemistry can also be used, but that is still in an early phase.

Charge Question 3

To facilitate timely development and release of ETAPs, EPA is proposing to have the standard methods document undergo peer-review. Individual ETAP reports based on these peer-reviewed methods would undergo internal technical and quality control review but not need to be individually peer-reviewed externally. Please comment on this proposed approach. (Topic primarily covered in Day 2)

Dr. Liehr summarized the response to charge question 3. ETAP is innovative. The group does not believe there is a need for external review of the ETAP approach, because the approaches are standardized and have already been peer reviewed. However, the group did flag that the UF may be modified, which could warrant external review. Dr. Jones stated that it helps with determining language in order to avoid being restrictive.

Charge Question 4

To facilitate rapid development and review of each ETAP, the results from the systematic evidence mapping, 5-day transcriptomic study, and TRV derivation are compiled and reported in a standardized ETAP reporting template with minimal free-form text. The ETAP template and an example ETAP using empirical data are provided for your review. Please comment on the extent to which the content and format of the reporting template and the example ETAP provide the important quantitative human health assessment information for a data poor chemical, with suggestions for improvement if warranted. (Topic primarily covered in Day 2)

Dr. Edwards discussed charge question 4. The group recommended that standard language be used to discuss the limitations in the background and assessment review. The group also asked that more detail about TempoSeq (what version was used, which genes are included, etc.) is included. These details should go into the template in case any changes are made in the future.

Additionally, there is a need to clarify what a cancer versus non-cancer TRV is. Future considerations could also include whether better gene sets or approaches may come online.

Dr. Liehr stated that she thought ETAP is a great product but that the document could benefit from better wording. In addition, she stated that there needs to be clarification on which specific steps are part of the 9-month timeline.

Wrap-Up and Adjournment

Tom Tracy, DFO, ORD

Rusty Thomas, Director, CCTE

Dr. Thomas thanked the panelists for their work. The meeting was adjourned.

DRAFT

Appendix A: Agenda

United States Environmental Protection Agency
Office of Research and Development (ORD)
EPA Transcriptomic Assessment Products (ETAP) Panel
Meeting Agenda – July 11-12, 2023
Research Triangle Park, North Carolina

Day 1: July 11, 2023

Time	Duration	Topic	Speaker
9:00-9:10 am	10 minutes	Welcome	Maureen Gwinn
9:10-9:20 am	10 minutes	Introduction to the Panel	Tom Tracy
9:20-9:45 am	25 minutes	EPA ORD Portfolio Approach and Where ETAP Fits	Samantha Jones
9:45-10:00 am	15 minutes	Day 1 Agenda, Introduction of ETAP Team, and Charge to the Panel (Review Charge Qs)	Rusty Thomas
10:00-10:30 am	30 minutes	Break	
10:30-11:00 am	30 minutes	Science Support Introduction/Background	Alison Harrill
11:00-11:30 am	30 minutes	Literature Review	Leah Wehmas
11:30-12:00 pm	30 minutes	NTP Genomics Report Overview	Scott Auerbach
12:00- 1:00 pm	60 minutes	Working Lunch 12:00-12:30 pm Break 12:30-1:00 pm Discussion of Panel Roles and Responsibilities	
1:00-1:30 pm	30 minutes	Dose Response Methods and Parameter Refinement	Logan Everett
1:30-2:00 pm	30 minutes	Concordance Analysis with Inter-study Variability	Kelsey Vitense
2:00-2:10 pm	10 minutes	Summary	Alison Harrill
2:10-2:30 pm	20 minutes	Break	
2:30-3:30 pm	60 minutes	Facilitated Panel Q/A	Co-Chairs: Craig Rowlands and Katherine von Stackelberg
3:30– 4:30 pm	60 minutes	Public Comment Period	Facilitator: Tom Tracy
4:30 – 4:45 pm	15 minutes	Wrap Up	Annette Guiseppi-Elie

DRAFT

4:45 – 5:45 pm	60 minutes	Break up into Charge Question groups 1-4 and Initial Discussions (closed session)	Co-Chairs: Craig Rowlands and Katherine von Stackelberg

Day 2: July 12, 2023

Time	Duration	Topic	Speaker
9:00-9:10 am	10 minutes	Welcome Back	Chris Frey
9:10-9:20 am	10 minutes	Day 2 Agenda and Charge Qs	Rusty Thomas
9:20-9:35 am	15 minutes	ETAP Overview- Introduce MOPA as the example	Alison Harrill
9:35-10:00 am	25 minutes	Database Search and Systematic Evidence Map (SEM)	Avanti Shirke
10:00-10:30 am	30 minutes	In vivo Study Design	Leah Wehmas
10:30-11:00 am	30 minutes	Break	
11:00-11:20 am	20 minutes	Transcriptomic Dose Response Analysis	Logan Everett
11:20-11:50 am	30 minutes	Reference Value Derivation and Reporting	Jason Lambert
11:50-12:05 pm	15 minutes	Comparison of Transcriptomics Reference Doses (TRVs) with Reference Doses/Concentrations (RfD/Cs) and Summary	Alison Harrill
12:05-1:05 pm	60 minutes	Working Lunch 12:05-12:35 pm Break 12:45-1:15 Begin questions from Panel on ETAP Method	
1:05-2:00 pm	55 minutes	Continue Questions from Panel on ETAP Method	Co-Chairs: Craig Rowlands and Katherine von Stackelberg
2:00-4:00 pm	120 minutes	Break up into Charge Question Groups 1-4 (closed session)	Co-Chairs: Craig Rowlands and Katherine von Stackelberg
4:00-5:00 pm	60 minutes	Report out and Charge Question Discussions	Co-Chairs: Craig Rowlands and Katherine von Stackelberg

DRAFT

5:00- 5:15 pm	15 minutes	Wrap Up and Close Meeting	Rusty Thomas and Tom Tracy

Appendix B: Participants

ETAP Panel Members:

Katherine von Stackelberg, *Chair*
Justin "Craig" Rowlands, *Co-Chair*
Thomas Burke
Harvey Clewell
Stephen Edwards
Rebecca Fry
Xiugong Gao
Gladys Liehr
Gloria Post
Justin Teeguarden
Chris Vulpe
Fred Wright
Lauren Zeise*

*participated virtually

EPA Designated Federal Official (DFO): Tom Tracy*, Office of Research and Development*

EPA Presenters:

Scott Auerbach, Molecular Toxicologist, Division of Translational Toxicology, National Institute of Environmental Health Sciences (NIEHS)
Logan Everett, Bioinformatics Scientist, Center for Computational Toxicology and Exposure
Chris Frey, Assistant Administrator, Office of Research and Development
Annette Guiseppi-Elie, Associate Director of Science, Center for Computational Toxicology and Exposure
Maureen Gwinn, Principal Deputy Administrator, Office of Research and Development
Alison Harrill, Associate Director of Toxicology, Center for Computational Toxicology and Exposure
Samantha Jones, Associate Director for the Assessment of Science, Center for Public Health and Environmental Assessment
Jason Lambert, Supervisory Toxicologist, Office of Research and Development
Avanti Shirke, Biologist, Office of Research and Development
Rusty Thomas, Director, Center for Computational Toxicology and Exposure
Kelsey Vitense, Research Scientist, Office of Research and Development
Leah Wehmas, Genomics Scientist, Center for Computational Toxicology and Exposure

Other EPA Attendees:

Olivia Anderson	Heidi Bethel	Amanda Brennan
Michelle Angrish	Adam Biales	Joseph Bundy
Meagan Bell	Jackson Bounds	Susan Burden

Jone Corrales
John Cowden
Sarah Davidson-Fritz
Mike DeVito
Kathie Dionisio
Kellie Fay
Robert Flick
Jermaine Ford
Stiven Foster
Derik Haggard
Tracy Hancock
Joshua Harrill
Monique Hazemi
Maria Hegstad
Carolyn Holmes
Weichun Huang
Michael Hughes
Thomas Jackson
Amanda Jurgelewicz
John Kenneke
Matt Klasen
Jason Lambert
Tyler Larson
David Lattier
Candice Lavelle
Monica Linnenbrink
Lucina Lizarraga
Denise MacMillan
Roman Mezencev
Viktor Morozov
Holly Mortensen
Nora Mosch
Esra Mutlu
Beth Owens
Katie Paul Friedman
Robert Payne
Dan Petersen
Allison Phillips
Kathryn Renyer
Jesse Rogers
Mary Ross
Kate Saili
Reeder Sams
Zachary Stanfield
Michael Stewart
Ravi Subramaniam
Jennifer Terlouw
Kris Thayer
Olivia Torano
Elin Ulrich
Scarlett VanDyke
Sara Vliet
Sean Watford
Chelsea Weitekamp
Barbara Wetmore
Paul White
George Woodall
Douglas Young

Other Participants:

Selim Atak
David Balshaw
Tara Barton-Maclaren
Richard Becker
Scott Belcher
John Bucher
Yenwei Chen
Amy Clippinger
Eduardo Costa
Matthew Daidola
Jeremy Erickson
Graham Everett
Melinda Gullett
Will Gwinn
Jui-Hua Hsieh
Christel Johnson
Agnes Karmaus
Joseph Manuppello
Kimberly McAllister
Barry McIntyre
Frannie Nilsen
Rick Paules
Richard Raymond
Leslie Recio
Bridget Rogers
Lillian Rosenberg
Jeffrey Rosenberg
Ivan Rusyn
Dan Selechnik
Daniel Shaughnessy
Andrew Turley
Anna van der Zalm
Nigel Walker
Lindsay Wilson
Randall Yang

Contractor Support (ICF):

Sagi Gillera
Ali Goldstone
Catherine Smith
Sam Snow
Leah West

Appendix C: Charge Questions

1. Given the literature review and the data analysis presented in the documents, please comment on whether the approach outlined for transcriptomic benchmark dose analysis and gene set summarization following a 5-day *in vivo* exposure are clearly described and provide a scientifically supportable estimate of the point-of-departure for chronic toxicity for data poor chemicals. (Topic primarily covered in Day 1)
2. EPA has proposed standard uncertainty factors to account for intraspecies variability (UF_H), interspecies differences (UF_A), database limitations (UF_D), duration (UF_S), and LOAEL-to-NOAEL extrapolation (UF_L) in the standard methods document. Are the uncertainties in the derivation of the reference values clearly described, and are the uncertainty factors scientifically justified? (Topic primarily covered in Day 2)
3. To facilitate timely development and release of ETAPs, EPA is proposing to have the standard methods document undergo peer-review. Individual ETAP reports based on these peer-reviewed methods would undergo internal technical and quality control review but not need to be individually peer-reviewed externally. Please comment on this proposed approach. (Topic primarily covered in Day 2)
4. To facilitate rapid development and review of each ETAP, the results from the systematic evidence mapping, 5-day transcriptomic study, and TRV derivation are compiled and reported in a standardized ETAP reporting template with minimal free-form text. The ETAP template and an example ETAP using empirical data are provided for your review. Please comment on the extent to which the content and format of the reporting template and the example ETAP provide the important quantitative human health assessment information for a data poor chemical, with suggestions for improvement if warranted. (Topic primarily covered in Day 2)