January 13, 2014

Dr. Kenneth Olden
Director, National Center for Environmental Assessment
Office of Research and Development
USEPA Headquarters, Ariel Rios Building
1200 Pennsylvania Avenue, N. W.
Mail Code: 8601P
Washington, DC 20460

Re: Docket EPA-HQ-ORD-2013-0680 "Next Generation Risk Assessment: Incorporation of Recent Advances in Molecular, Computational, and Systems Biology (September, 2013)"

Dear Dr. Olden,

The American Chemistry Council's (ACC) Regulatory and Technical Affairs Department and Center for Advancing Risk Assessment Science and Policy (ARASP) appreciate the opportunity to submit our joint comments on EPA's external draft document entitled "Next Generation Risk Assessment: Incorporation of Recent Advances in Molecular, Computational, and Systems Biology (September, 2013) (Draft Report).

We support continued investment of research resources to improve development and application of the new technologies to enable faster, less expensive and less laboratory animal-intensive hazard, exposure and risk estimations, and to more accurately determine the probability of adverse health outcomes at environmentally relevant exposure levels. While the Draft NexGen Report is a start, it is important for EPA to make clear how the Tox21 and related approaches discussed in the Draft Report will be integrated throughout the Agency's programs. Our attached comments address in detail a number of issues raised in the Draft Report. Further, to assure that these 21st century technologies and approaches are scientifically robust, and optimized as fit for purpose for regulatory use in EPA's programs, we also make a number of recommendations for Agency actions.

Please do not hesitate to contact me if you have any question on ACC's comments or recommendations. I can be reached by e-mail at Rick_Becker@americanchemistry.com or by phone at 202-249-7000.

Sincerely

Richard. A. Becker, Ph.D., DABT
Senior Toxicologist

Attachment: January 14, 2014 Joint Comments of the American Chemistry Council and Center For Advancing Risk Assessment Science And Policy on EPA's September 2013 External Review Draft Report "Next Generation Risk Assessment: Incorporation Of Recent Advances In Molecular, Computational, and Systems Biology"

JOINT COMMENTS OF

THE AMERICAN CHEMISTRY COUNCIL (ACC)

and

ACC's CENTER FOR ADVANCING RISK ASSESSMENT SCIENCE AND POLICY (ARASP)

on

EPA's September 2013 External Review Draft Report

"Next Generation Risk Assessment:

Incorporation of Recent Advances in Molecular, Computational, and Systems Biology"

January 13, 2014
American Chemistry Council
700 2nd Street, NE
Washington DC 2002

Table of Contents

EXECUTIVE SUMMARY

The American Chemistry Council (ACC) and the Center for Advancing Risk Assessment Science and Policy (ARASP) commend the Environmental Protection Agency (EPA) for its initiatives to reflect recent advances in scientific knowledge, tools, and technologies in chemical risk assessment which could more efficiently and accurately characterize potential hazards and risks. EPA's publication, Next Generation Risk Assessment: Incorporation of Recent Advances in Molecular, Computational, and Systems Biology (September, 2013) (Draft NexGen Report), embraces a conceptual framework for tiered, risk-based testing and evaluation strategies, which we support. Many of the approaches presented in the Draft NexGen Report are still aspirational with regards to hazard and risk prediction. At this early stage of their development, some approaches are more accurately labeled as exploratory research projects, and therefore have limited utility in the near term. Furthermore, it is not clear how the EPA's Office of Research and Development NexGen effort has been, or will be, coordinated with the Tox21 activities of the National Center for Computational Toxicology and with the testing and assessment activities of the Office of Chemical Safety and Pollution Prevention (OCSPP). While the Draft NexGen Report is a start, it is important for EPA to make clear how the Tox21 and related approaches discussed in the Draft NexGen Report will be integrated throughout the Agency's programs.

We support continued investment of research resources to improve development and application of the new technologies to enable faster, less expensive and less laboratory animal-intensive hazard, exposure and risk estimations, and to more accurately determine the probability of adverse health outcomes at environmentally relevant exposure levels. Our comments address in detail a number of issues raised in the Draft Report. Further, to assure that these 21st century technologies and approaches are scientifically robust, and optimized as fit for purpose for regulatory use in EPA's programs, including the Office of Chemical Safety and Pollution Prevention (OCSPP), we make a number of recommendations for Agency action as follows:

- Revise the benzene, ozone and benzo(a)pyrene prototypes, taking into consideration public comments, and then submit the revised drafts for independent peer review before using these methods in EPA's Integrated Risk Information System (IRIS) or other EPA programs.

- Build from the NexGen work and the OECD Adverse Outcome Pathway (AOP) activities to develop draft guidance which addresses the development, evaluation and use of AOPs for defined purposes such as prioritization, formation of categories for read across, integrated testing, and screening level hazard/risk assessment.

- Develop an integrated plan, led by the OCSPP, in coordination with National Center for Computational Toxicology, which:

    - » Groups the technologies and approaches into three categories: 1) ready now, 2) undergoing scientific evaluation for relevance, reliability and fitness for purpose, and 3) under development. Such a categorization will provide a clearer picture of the state of the science, the current confidence in use of data from these methods/approaches for specific purposes, and a path forward to improve scientific confidence in the methods.

» Defines the additional datasets and analyses needed to expand the tiered, high throughput screening (HTS)/transcriptomics margin of exposure method to achieve regulatory acceptance for the full range of chemical domains needed by EPA.

» Focuses additional effort on incorporating exposure into the next generation of risk assessment approaches, by building from the consumer model in the ECETOC Targeted Risk Assessment exposure tool and the tiered exposure assessment approaches developed by the ILSI-HESI RISK 21 project

» Develops the framework the Agency will use to establish and document the scientific confidence that is needed for these methods and prediction models to be used for regulatory purposes.

» Employs a more open stakeholder engagement process by maximizing open meetings, broadening consultations and collaborations and conducting peer review in accordance with Agency procedures for influential risk assessment guidance and policies.

I. INTRODUCTION

The American Chemistry Council's (ACC)[1] Regulatory and Technical Affairs Department and Center for Advancing Risk Assessment Science and Policy (ARASP)[2] appreciate the opportunity to submit our joint comments on EPA's draft document entitled "Next Generation Risk Assessment: Incorporation of Recent Advances in Molecular, Computational, and Systems Biology (September, 2013) (Draft NexGen Report). The stated goal of EPA's draft report was "to create a faster and more cost effective system for chemical risk assessment by incorporating new chemical testing data and advances in molecular and systems biology technologies." The Federal Register Notice announcing the release of this document (FR, 78: 59927- 59929) listed the following specific aims for the Next Generation (NexGen) effort: 1) to demonstrate proof of concept that the data and methods from recent advances in biology can inform risk assessment; 2) to identify which of the information resources and practices are most useful for particular purposes (value of information); 3) to develop decision considerations for use of different types of NexGen data and methods to inform different types of assessments; and 4) to identify priority research needs.

The Agency should be commended for its initiatives to improve approaches for chemical risk assessment to reflect recent advances in scientific knowledge, tools, and technologies which could more efficiently and accurately characterize potential hazards and risks posed by chemicals. There has long been recognition that the conventional toxicity testing methods, which are resource intensive in terms of animal usage, financial commitments, and laboratory facilities, have limitations to efficiently address toxicity testing demands. The chemical industry views integrated testing and assessment approaches, which consider all relevant and reliable information, including exposure, as integral to improving both the efficiency and quality of human health and ecological safety evaluations of chemical products.[3,4]

[1] ACC represents the leading companies engaged in the business of chemistry. ACC members apply the science of chemistry to make innovative products and services that make people's lives better, healthier and safer. ACC is committed to improved environmental, health and safety performance through Responsible Care®, common sense advocacy designed to address major public policy issues, and health and environmental research and product testing. The business of chemistry is a $770 billion enterprise and a key element of the nation's economy. It is one of the nation's largest exporters, accounting for 12 percent of all U.S. exports. Chemistry companies are among the largest investors in research and development. Safety and security have always been primary concerns of ACC members, and they have intensified their efforts, working closely with government agencies to improve security and to defend against any threat to the nation's critical infrastructure.

[2] ARASP is a coalition of 19 organizations focused on promoting the development and application of up-to-date, scientifically sound methods for conducting chemical assessments. ARASP members include: Acrylonitrile Group, ACC's Chlorine Chemistry Division, Ethylene Oxide Panel, Formaldehyde Panel, Hexavalent Chromium Panel, High Phthalates Panel, Hydrocarbon Solvents Panel, Olefins Panel, Oxo Process Panel, Propylene Oxide/Propylene Glycol Panel, Public Health and Science Policy Team, Silicones Environmental, Health and Safety Center of North America and Vinyl Chloride Health Committee, American Cleaning Institute, American Petroleum Institute, CropLife America, Halogenated Solvents Industry Alliance, Nickel Producers Environmental Research Association and Styrene Information and Research Center.

[3] ICCA (2011). Global Product Strategy: Guidance on Chemical Risk Assessment http://www.icca-chem.org/ICCADocs/ICCA_GPS%20July2011_LowResWEB.pdf.

[4] Plunkett et al. (2010). An enhanced tiered toxicity testing framework with triggers for assessing hazards and risks of commodity chemicals. Regul Toxicol Pharmacol. 58(3):382-94.

ACC[5] and our member companies have been examining opportunities to improve on the ability to more efficiently identify potential adverse health outcomes associated with chemical exposure in a risk-based context. We strongly support "fit for purpose" chemical assessment, and we has played a constructive role in the research and development of advanced approaches to chemical profiling for use in hazard and risk assessments through ACC's Long Range Research Initiative.[6] In addition, we have fostered approaches for developing the scientific confidence in such tools, which is requisite for use in product stewardship and regulatory programs.[7] In a continuation of these constructive actions, we offer the following comments on EPA's Draft NexGen Report. We also provide a series of recommendations for improving the development, evaluation and application of the Tox21 methods described in the Draft NexGen Report.

II. GENERAL COMMENTS

The NexGen effort launched by the EPA proposes a conceptual framework that aligns with tiered, risk-based testing and evaluation strategies which ACC supports. The approaches discussed in the Draft NexGen Report have generated considerable optimism for the potential to significantly advance the risk assessment process, both in the near term and in the long term. However, as detailed in our attached comments, many of the approaches presented are still aspirational with regards to hazard and risk prediction, and are at such an early stage of development. They are more accurately labeled as exploratory research projects, and therefore have limited utility in the near term. Furthermore, it's not clear how the EPA's Office of Research and Development NexGen effort is integrated with Tox21 activities of the National Center for Computational Toxicology and with the testing and assessment activities of the Office of Chemical Safety and Pollution Prevention (OCSPP).

It is a shared aspiration of EPA, industry and other stakeholders to rely upon faster and less expensive risk assessment methods that use fewer animals to assess the potential hazards and risks of chemical exposures. These methods, nevertheless, have to be accurate in terms of biological relevance and their predictive value (i.e., relevance, reliability, sensitivity and specificity). There should be recognition that changes to the risk assessment paradigm to meet the needs of the 21st century is an evolutionary process which requires development of scientific confidence in these new risk assessment methods.[8] The phrase "more robust" used at various places in the Draft NexGen Report to describe the new risk assessment methods does not necessarily capture this important point. EPA should use these new risk assessment methods for hazard identification and risk assessment only after they have been adequately validated, qualified and determined "fit for purpose", concepts duly recognized by EPA (page 93 of the Draft NexGen Report). Additionally, adoption of new/novel assays and methods needs to be done transparently through multi-stakeholder engagement.

---

[5] Throughout these comments, when ACC is used, this reflects the joint perspectives and recommendations of ACC and ARASP.

[6] ACC (2012). LRI 2012 Research Portfolio. http://lri.americanchemistry.com/LRI-Research-Program/2012-LRI-Research-Portfolio.pdf.

[7] For example, Patlewicz et al (2013). Use and validation of HT/HC assays to support 21st century toxicity evaluations. Regul Toxicol Pharmacol. 65(2):259-68.

[8] Rowlands et al.(2013). FutureTox: Building the Road for 21st Century Toxicology and Risk Assessment Practices. Toxicol Sci. 2013 Dec 18. [Epub ahead of print].

II.A. The Need to Develop and Document Scientific Confidence in Advanced Molecular Profiling Approaches

As EPA is well aware, having sufficient scientific confidence in causal linkages and dose-responses between key events at different levels of biological organization along the pathway(s) that lead to adverse health effects is the key to using high throughput (HT) and high content (HC) methods to inform hazard and risk characterization. ACC calls EPA's attention to the development and refinement of a framework for evaluating and documenting scientific confidence in HT/HC methods and their prediction models (Patlewicz et al (2013) and Becker et al. (2014).[9] The agency is encouraged to appropriately evaluate HT/HC methods using this or a similar framework.

The Draft NexGen Report at times the document appropriately acknowledges the limitations of the *in vitro* cell culture models (see page 27). , but the report is inconsistent in this regard. The Draft NexGen Report should acknowledge and address the publication by Thomas et al. (2012)[10] whose extensive and robust analysis of the performance of the 600 ToxCast Phase I screening assays conducted on data rich agro-chemicals, led them to conclude that these assays had very low power for predicting *in vivo* chemical hazards. In light of this analysis, many of the statements relating to the ability of ToxCast assays to inform risk assessment should be viewed with considerable caution. For example, the statement in Table 8 on page 66 of the Draft NexGen Report that "False negatives and positives for ToxCast evaluated assays are low" is at odds with the conclusions drawn by Thomas et al. (2012). Furthermore, in a preliminary analysis of data generated under EPA's Endocrine Disruptor Screening Program (EDSP), CropLife America's Endocrine Policy Forum (CLA/EPF)[11] reported a 44% -56% false positive rate for the ToxCast estrogen-related assays. The Draft NexGen Report should have incorporated the insights from these analyses of the ToxCast assays and appropriately modified recommendations for their use. The following excerpt from page 44 of the Draft NexGen Report highlights this concern: "*In vitro* responses appear to have commonalities with *in vivo* responses but also are affected by a number of variables, such as test system, metabolism, cell type, tissue type, time course of events (ozone data only), individual characteristics (intrinsic and extrinsic), and species. These complexities make the identification of a specific disease hazard from *in vitro* only data difficult." The last point is clearly an understatement. Extreme caution should be exercised when using only in vitro data to drive hazard identification and risk assessment decisions.

The Draft NexGen Report would be greatly improved by providing more specifics and substantiated statements on how the NexGen methods will advance the science of risk assessment and regulatory policy. A number of assertions appear overly optimistic or speculative, for example: "Regulatory toxicology as a whole will move toward increasing reliance on predictive approaches to assessing chemical risk, with a greater emphasis placed on understanding chemical perturbation(s) of conserved biological pathways at key junctures,

[9] Op. cit. footnote 7 and Becker et al.(2014). Applying a Scientific Confidence Framework to a HTS-Derived Prediction Model for Endocrine Endpoints: Lessons Learned from a Case Study (submitted).

[10] Thomas et al. (2012). A comprehensive statistical analysis of predicting in vivo hazard using high-throughput in vitro screening. Toxicol Sci. 128(2):398-417.

[11] CLA/EPF (2013). CropLife America/Endocrine Policy Forum (CLA/EPF) comments submitted with regard to the FIFRA SAP on prioritization (January 28, 2013). http://www.regulations.gov/#!documentDetail;D=EPA-HQ-OPP-2012-0818-0026.

including molecular initiating events (MIEs) (e.g., activation or inactivation of specific receptors, enzymes, or transport proteins)" (page 80 of the Draft NexGen Report). The general experience has been that regulatory agencies can be slow in accepting new methodologies and often first require acceptance by standard setting organizations such as the Organisation for Economic Co-operation and Development (OECD). At the State level, where EPA's decisions are often implemented, there is additional delay in embracing newer technologies and improvements in risk assessment. This substantial hurdle is recognized by the document, which calls for improvement in this process and new "fit-for-purpose" approaches. However, care should be taken that these new technologies are not deployed prematurely to classify hazards and risks without adequate confidence not only in their sensitivity to predict an adverse outcome but also in their specificity to avoid an unacceptable level of false positive responses, lest the public and scientific community lose confidence in these new approaches.

II.B. Mode of Action, Dose Response and Systems Biological Approaches are Key

We support using mode of action (MOA) as an organizing principle. We encourage the work of EPA and others in developing and building scientific confidence in MOAs[12] and in adverse outcome pathways AOPs.[13] ACC has consistently urged EPA to employ MOA analytical constructs, such as the Key Event Dose Response framework[14] or the Hypothesis-Based Weight of Evidence framework[15] in risk assessments to assure that key events, dose-dependent transitions and 21st century knowledge of the biology of molecular, cellular and organ responses form the foundation of human health risk assessments. AOPs are also promising tools that can be used to describe the sequential steps and linkages between initial events, intermediate events, key events, and adverse outcomes, and to document methods to measure key events. In cases where these relationships are well established, the goal is to develop and build scientific confidence in qualitative and quantitative prediction models. The Draft NexGen Report notes that the European Union's "Safety Evaluation Ultimately Replacing Animal Testing," SEURAT-1 (http://www.seurat-1.eu/), has begun to develop a conceptual framework that can be used as a basis to combine information derived from predictive tools to support a safety assessment process. The SEURAT-1 research strategy is to adopt a toxicological mode-of-action approach to describe how a substance might adversely affect human health, and to use this knowledge to develop complementary theoretical, computational, and experimental (HT/HC *in vitro*) models.

The Draft NexGen Report accurately emphasizes the need to characterize systems biology to integrate the observations from the 21st century screening tools. Systems biology is still an emerging science, however, and requires a high level of expertise to integrate information from multiple levels of biological interactions and complexity. As systems biology forms the foundation for many of the approaches envisioned in the document, there is a need to

[12] For example, Meek et al. (201) New developments in the evolution and application of the WHO/IPCS framework on mode of action/species concordance analysis. J. Appl. Toxicol. 2014; 34: 1–18.
[13] OECD (2013). Organisation for Economic Co-operation and Development (OECD). Guidance Document On Developing And Assessing Adverse Outcome Pathways. ENV/JM/MONO(2013)6. http://search.oecd.org/officialdocuments/displaydocumentpdf/?cote=env/jm/mono(2013)6&doclanguage=en
[14] Julien et al. (2009). The Key Events Dose-Response Framework: a cross-disciplinary mode-of-action based approach to examining dose-response and thresholds. Crit Rev Food Sci Nutr. 49(8):682-9.
[15] For example, Rhomberg et al. (2010). Hypothesis-based weight of evidence: a tool for evaluating and communicating uncertainties and inconsistencies in the large body of evidence in proposing a carcinogenic mode of action--naphthalene as an example. Crit Rev Toxicol. 40(8):671-96.

invest significant resources to develop this expertise to serve the needs of toxicology and risk assessment communities. There should also be an acknowledgement that the scientific community is several years away from having the needed resources in this area. EPA recognizes the need to anchor the molecular patterns observed to apical outcomes (page 25). This is an important recognition because molecular changes such as transcriptional signatures are generally seen as biomarkers of exposures rather than effects. Hazard identification using gene expression signatures alone has proven to be an elusive goal. Currently, it is uncertain how transcriptomic signatures would rapidly identify target organs and/or other toxicity responses such as neurobehavioral alterations. However, the technology has great promise in generating hypotheses to probe different plausible mode of actions. Furthermore, for data poor substances, the approach advanced by Thomas et al. (2013a, b) cited in the Draft NexGen Report [16] holds considerable promise. This latter approach does not predict an apical effect based on the affected pathways, but simply capitalizes on the correlation between the lowest transcriptional benchmark dose (BMD) based on a select list of tissues and the BMD identified for cancer and non-cancer endpoints from conventional toxicity studies (acknowledged by the document on page 62). Overall, because the transcriptional BMDs are lower than the BMDs from conventional toxicity studies, the transcriptional BMDs can be used as health-protective indicators of activity, and then evaluated with exposure information to produce a risk-based screening level assessment. ACC notes that the statement on page 59 (lines 20-32 & 33-35 of the Draft NexGen Report),pertaining to dose response relationships based on pathway analysis and uncertainty, contradicts the data-driven conclusions reached by Thomas et al. (2013a,b) as well as what was stated elsewhere in the Draft NexGen Report (on page 62).

The Draft NexGen Report acknowledges the importance of addressing the "thresholds" concept through robust dose-response analysis. As discussed by Rhomberg et al. (2011),[17] thresholds for key events are the rule, not the exception. The new technologies have the potential to generate robust dose-response data for dose-dependent transitions and address the existence of thresholds for a number of toxicity endpoints, including DNA-reactive carcinogens.[18] With respect to dose response and additivity to background and thresholds, Rhomberg et al. (2013)[19] document that linear extrapolation is the exception not the rule. In this regard, the new tools will help in probing the cellular protective mechanisms operational at the low end of the dose-response curve. It is important for EPA to clearly indicate how NexGen methods might be used to test and ultimately modify many important defaults that are currently the hallmark of traditional risk assessment (e.g., existence of thresholds for systemic toxicity and no thresholds for genotoxicity; use of 10X for inter- and intra-species differences; use of 10X for database deficiencies; etc.). Thus, the NexGen report should acknowledge the promise of high throughput methods generating high-content information as a means to overcome many of the technical limitations that have led to the current imbedding of highly conservative default assumptions in

[16] Thomas et al. (2013a). Incorporating new technologies into toxicity testing and risk assessment: Moving from 21st century vision to a data-driven framework.. Toxicol Sci.136(1):4-18
Thomas et al (20113b). Temporal concordance between apical and transcriptional points of departure for chemical risk assessment. Toxicol Sci. 134: 180-194.
[17] Rhomberg et al. (2011). Linear low-dose extrapolation for noncancer health effects is the exception, not the rule. Crit Rev Toxicol. 41(1):1-19.
[18] Gollapudi at al. (2013). Quantitative approaches for assessing dose-response relationships in genetic toxicology studies. Environ Mol Mutagen. 2013 54(1):8-18.
[19] Op. cit, footnote 17.

risk assessment policies and practices.[20]

ACC recommends EPA establish a proof of concept framework that starts with the mode of action and adverse outcome and works backwards by identifying key events and various genomic changes that can be phenotypically linked to these important key events, e.g., phenotypic anchoring of omics responses to histological evidence. The proof of concept should require that this elucidation of the mode-of-action, including genomics, be supported by the important dose-temporality concordance table set forth in the WHO/IPCS Human Relevance Mode-of-Action framework.[21]

II.C. Areas for Improvement: Application Issues

The Draft NexGen Report does not provide guidance on how the new technologies can be applied today or in the near future. The majority of the statements in the Draft NexGen Report on the use of these technologies are aspirational. Several of the technologies are still in their infancy/ research mode or practiced in one or a small number of laboratories. A notable example is the mouse population diversity lines proposed for use in toxicity testing whose utility is still highly speculative. It would be beneficial for EPA to categorize the technologies and approaches into three groups: 1) ready now, 2) undergoing scientific evaluation for relevance and reliability and fitness for purpose, and 3) under development. Such a categorization would give a clearer picture of the state of the science, the confidence in the data from these methods/approaches, and the path forward to improve scientific confidence to achieve the degree necessary for application for specific purposes. Furthermore, the agency should clearly acknowledge that it is premature to use the majority of the new technologies described in the document to inform the risk assessment for adverse outcomes either quantitatively or qualitatively without data to establish causal linkages among different levels of biological organization to show how specific responses and dose levels may lead to adverse outcomes.

In order to gain greater confidence in the ability of NexGen methods to more accurately predict risk, it is questionable whether an epidemiologically-based focus will be fruitful, since the causative agents for many human health outcomes often cannot be truly established with sufficient scientific confidence. For the NexGen methods to be used to predict human health risk (vs. e.g., prioritization) it is important to establish the relevance, reliability, sensitivity and specificity of these methods to predict a full range of indisputable disease outcomes, with associated information related to dose-response relevant those outcomes, in experimental animal test systems.

In a number of instances, the Draft NexGen Report makes fairly broad statements regarding the applicability of a particular methodology or approach based on a single study or data from a single laboratory. Furthermore, the concept of using *in vitro* cytotoxicity data from genotypically diverse cell lines to predict *in vivo* response is still an exploratory research concept and the statement "This approach will provide significant new insights into human variability in response..." is optimistic. Sufficient scientific confidence in such experimental models must be established prior to their use to inform risk assessment decisions.

[20] Bus JS, Becker RA. (2010). Toxicity testing in the 21st century: a view from the chemical industry. 112(2):297-302.
[21] Op. cit. footnote 12

There was also little discussion on how the NexGen technologies and approaches can modernize and potentially reduce the resource intensive conventional toxicology testing. For example, the need for chronic long-term studies could be re-examined using the new approaches and simplified in most instances or completely eliminated in certain cases. The approach described by Thomas et al. (2013a,b; cited in the report) could essentially eliminate the need for carcinogenicity studies in those cases where the exposure potential is extremely low.

II.D. Areas for Improvement: Biomarkers, Exposure and Dosimetry – Placing Activity Data and Similar Hazard-Related Information into a Risk Context

In the Summary section and at various other places in the Draft NexGen Report, molecular signatures are often referred to as "biomarkers of exposure and effect" and it is asserted that such biomarkers could be used to inform quantitative risk assessment. While it is accurate to state that these molecular events or signatures represent biomarkers of exposure, they should not be equated to biomarkers of effect without adequately qualifying them for that purpose using established standards such as those developed by Institute of Medicine (IOM) and the Food and Drug Administration (FDA).[22] An example of this is the identification of 16 significantly altered genes as biomarkers of benzene exposure (page 18 of the Draft NexGen Report). Ideally, biomarkers should undergo thorough scientific evaluation according to the standards established in human medicine.

As the toxicology community embraces more *in vitro* based methods and prediction models, there will be a growing need to model the *in vitro* concentrations to *in vivo* exposures as well as internal systemic doses so that the data from *in vitro* studies can be interpreted in the context of risk assessment. EPA should build from the work of the Hamner Institutes in developing and applying *In Vitro* to *In Vivo* Extrapolation (IVIVE) for dosimetry[23], by expanding IVIVE to sets of substances employed in ToxCast and Tox21.

While the Draft NexGen Report acknowledges the importance of internal doses and actual human exposures in the risk assessment of environmental and industrial chemicals, the document fails to cite the extensive work by private sector, academic, Centers for Disease Control and Prevention (CDC), Health Canada and EPA scientists in developing and applying Biomonitoring Equivalents (BEs)[24] to interpret human biomonitoring data, such as that from NHANES or the Canadian Health Measures Survey in a population risk assessment context. Furthermore, EPA is encouraged to expand the development and application of BEs as part of the Agency's activities to improve risk assessment and accelerate the incorporation of advanced approaches into EPA's hazard and risk assessment activities.

---

[22] Institute of Medicine (IOM), 2010. Evaluation of biomarkers and surrogate endpoints in chronic disease. ISBN: 978-0-309-15129-0.; FDA (2013). Biomarker Qualification Context of Use. http://www.fda.gov/Drugs/DevelopmentApprovalProcess/DrugDevelopmentToolsQualificationProgram/ucm284620.htm.

[23] Wetmore et al. (2013). Relative impact of incorporating pharmacokinetics on predicting in vivo hazard and mode of action from high-throughput in vitro toxicity assays. Toxicol Sci. 132(2):327-46; Yoon et al. (2014). Evaluation of simple in vitro to in vivo extrapolation approaches for environmental compounds. Toxicol In Vitro. 28(2):164-70.

[24] Aylward et al. (2013). Evaluation of biomonitoring data from the CDC National Exposure Report in a risk assessment context: perspectives across chemicals. Environ Health Perspect. 121(3):287-94; BECKER ET AL. (2012). Development of screening tools for the interpretation of chemical biomonitoring data. J Toxicol. 2012;2012:941082. doi: 10.1155/2012/941082.

Exposure should also be considered in applying HT/HC methods for prioritization. Molecular signatures alone should not form the basis for chemical prioritization as suggested at several places in the document. But instead, these signatures should be combined with human exposure estimates to calculate the margins of exposure (MOE) and the MOEs can then be used to prioritize chemicals for further testing (Thomas et al., 2013a,b). In addition, simply showing that a chemical produces a "signature" similar to a disease state, absent dose (or concentration) considerations and exposure potential, is much less likely to result in resources being focused on substances of the most risk to humans or the environment.

An important missed opportunity in the Draft NexGen Report was not specifying a limit to the top dose/concentration tested in the NexGen assays rather than testing exorbitantly high doses/concentrations, concentrations equivalent to the maximum tolerated dose (MTD). It is now widely recognized that the effects observed at the MTD are not relevant in the majority of cases to exposures encountered in the environment. The document proposed testing at a wide range of concentrations, including those relevant to human exposures, but had little discussion about what the highest tested dose should be. Irrelevant effects observed *in vitro* or *in vivo* at MTDs are unlikely to advance the accuracy of risk assessment for environmental exposures. The Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) expert report on *in vitro* endocrine screening assays recommends that the limit concentration should be 1 mM, but notes that the solubility characteristics of each test substance must be taken into consideration.[25]

III. SPECIFIC COMMENTS ON SECTION 3.1.1: BENZENE-INDUCED LEUKEMIA

The benzene prototype brings together selective data and perspectives from traditional measures as well as data from new technologies to evaluate the health risk(s) following exposure to benzene. While this exercise is interesting and could be viewed as a step in the right direction for an eventual practical application, this prototype fails to thoroughly consider and acknowledge a number of factors essential to a scientifically sound and systematic risk assessment, and therefore, should not be used to inform benzene risks. Factors that warrant particular emphasis are discussed in more detail below and are presented within the context of five overarching themes:

A. Clarity needed on EPA's intentions regarding this case study to benzene risk values.
B. Omitted perspectives and references.
C. Lack of acknowledgement and consideration of the uncertainties associated with the use of genomics to inform benzene risk assessment.
D. Lack of acknowledgement and consideration of the uncertainties for predicting hazards for data-poor chemicals.
E. Lack of acknowledgement and consideration of the uncertainties for extrapolating risk to low doses.

[25] ICCVAM. (2002). Expert Panel Evaluation of the Validation Status of In Vitro Test Methods for Detecting Endocrine Disruptors: Estrogen Receptor and Androgen Receptor Binding and Transcriptional Activation Assays http://ntp.niehs.nih.gov/iccvam/docs/endo_docs/expertpanfinalrpt/panelrpt1102.pdf.

Considering these shortcomings, the benzene prototype falls significantly short of a "proof of concept" model. At best, this prototype is an example of how computational methods and other non-traditional sources of information *might* be used to generate hypotheses, or how non-traditional data *might* one day be integrated to inform regulatory decisions. This is somewhat acknowledged within the document where the conclusions across the Tier 3 prototypes (section 3.1.4) tend to reflect the *potential* utility of these new data types and the need for more research to substantiate their use in risk assessment. However, this perspective is inconsistent with the overarching conclusion stated in the Draft NexGen Report that the prototypes demonstrate a definitive "proof of concept" suitable for informing hazard identification, characterizing exposure-response, and predicting hazard of data poor chemicals. It is recommended the Draft NexGen Report consistently acknowledge these shortcomings, and alter the conclusions to reflect the uncertainty associated with reliance on the genomics data. The report would also be improved by inclusion of a path forward on how to improve scientific confidence to the extent necessary for reliable application of these data in benzene risk assessment.

III.A. Clarity needed on EPA's Intentions regarding this Case Study to Benzene Risk Values

The benzene prototype assessment does not sufficiently explain how close (or far away) a genomics-based benzene risk assessment is at this time. It seems that at this point, health risks of benzene would likely continue to be based on current empirical and weight of evidence observations in human populations.

Further, the assessment, while attempting to highlight what is possible regarding advancing benzene risk estimation for the endpoint of leukemia (namely acute myeloid leukemia (AML), adds references and opinion regarding other effects of benzene, such as lymphoma and effects on blood cell counts. The discussion of benzene effects on blood counts is not relevant to the title of this assessment "benzene-induced leukemia". The intentional inclusion of other blood effect observations does not add clarity to what is possible regarding a pathway to understanding key events for AML.

III.B. Omitted Perspectives and References

The ability to better assess the scientific validity of the conclusions about benzene and the recommendations provided in the Draft NexGen Report would be enhanced by better documentation of the overall process for study selection. Although the draft report states that it is "not intended to be a comprehensive review of all available data that might be used in risk assessment", the literature selection process should be systematic, and transparently documented. Additionally, references given for several issues are selective and therefore an incomplete story is presented. Specifically regarding blood cell count effects of benzene, the Draft NexGen Report references Lan et al. (Science 306:1774-6, 2004) as a key study but fails to present a complete line of evidence. Several other studies have not confirmed the Lan et al. (2004) observations regarding effects of benzene exposure on blood cell count. Omitted references include:

- Tsai et al., 1983 and 2004 (J. Occup. Med.25(9):67-73, 1983; Regulatory Toxicology and Pharmacology 40(1):67-73, 2004) ;

- Schnatter et al., 2010 (Chem. Biol. Interactions 184(1-2): 174-81, 2010):
- Swaen, et al., (2010) Chem. Biol. Interactions 184(1-2): 94-100, 2010; and
- an example of a recent relatively major study regarding non Hodgkins lymphoid neoplasms (Wong, et al., 2010, Chem. Biol. Interactions 184(1-2): 129-46, (2010)).

The continuing uncertainty surrounding the mechanism of benzene-induced AML has been amplified recently with a publication which questions the current paradigm of the similarities between the origins of benzene and therapy related AML.[26]

III.C. Lack of Acknowledgement and Consideration of the Uncertainties Associated with the Use of Genomics to Inform Benzene Risk

The state of the science for benzene-mediated hematotoxicity does not currently support the use of genomics data to inform the dose response in risk assessment or to predict hazards for data-poor chemicals. As described in more detail below, the primary reasons for this are: lack of consensus on how benzene mediates toxicity; insufficient support for a causal linkage between the genomic events and the apical outcome; and inadequate understanding of the relevance of the current test system to inform benzene risk.

Considering these shortcomings, the benzene prototype falls significantly short of a "proof of concept" model. At best, this prototype is an example of how computational methods and other non-traditional sources of information *might* be used to generate hypotheses, or how non-traditional data *might* one day be integrated to inform regulatory decisions. The agency should clearly acknowledge that it is premature to use the genomics data described in the document to inform the risk assessment either quantitatively or qualitatively without chemical specific data to establish causal linkages to specific responses and the adverse outcome. It is recommended the Draft NexGen Report consistently acknowledge these shortcomings, and alter the conclusions to reflect the uncertainty associated with reliance on the proposed genomics data.

III.C.1. Lack of Clarity Regarding Mode of Action of Benzene Mediated Disease

The responsible incorporation of molecular epidemiologic and molecular clinical data sets into risk assessment requires a well-established and consensus driven knowledge base to anchor scientific observations. The knowledge base should include, at a minimum, an understanding of the disease process and how the mechanistic observations correlate to the disease stages observed in an individual. Although not emphasized in the Draft NexGen Report, there is a lack of consensus within the scientific community on the mode of action of benzene meditated hematopoietic disease.[27] Equally important, the understanding of the biological processes responsible for malignant transformation and disease progression in the bone marrow continues to evolve within the hematology community. The uncertain etiology of AML in both *de novo* and chemically induced disease states, profoundly impacts the usefulness of the genomics data to benzene risk assessment.

[26] Irons et al. (2013). Acute myeloid leukemia following exposure to benzene more closely resembles de novo than therapy related-disease. Genes Chrom Cancer. 52:887-94.
[27] Meek et al. (2010). Proposed mode of action of benzene-induced leukemia: Interpreting available data and identifying critical data gaps for risk assessment. logical Interactions, Chem-Biol Interact. 2010. 184:279-285.

- **Bone Marrow Niche-Mediated On Cogenesis Has Not Been Considered**
  The "probable mechanism" of benzene-mediated AML described in the Draft NexGen Report (Box 5) involves a number of generic cellular events with the hematopoietic stem cell serving as the primary target for cancer initiation, promotion, and progression. While this hypothesized mechanism captures many of the common biological events shared across the multitude of proposed hypotheses for benzene mediated disease, this reductionist hematopoietic stem cell-centered approach to understanding hematopoietic deregulation has limitations and important implications to the utility of the proposed systems biology approach to benzene risk assessment. Although the scientific community may be in agreement that benzene must be metabolized to initiate leukemogenesis, there is little agreement regarding what cellular target the reactive metabolite interacts with or even what cell type the metabolite targets to initiate leukemia.

  In the hematology community, it is increasingly recognized that the tumor microenvironment "niche" plays a pivotal role in cancer initiation and progression.[28] This niche-mediated model of oncogenesis challenges the existing premise that the initiating events in cancer are completely or predominantly cancer cell autonomous. For decades, benzene has been widely accepted as the prototype chemical leukemogen for which disease pathogenesis is assumed to involve an obligatory role for clonal cytogenetic injury occurring in the leukemia initiating cell (i.e., hematopoietic stem cell). Remarkably, these assumptions are based entirely on non-quantitative, somewhat subjective human studies, clinical observations, indirect experiments employing surrogate cells that are not stem cells (e.g., peripheral lymphocytes), and *in vitro* studies that do not adequately reproduce the bone marrow microenvironment *in vivo*. However, data is accumulating to support a niche-mediated mechanism for benzene-mediated leukemogenesis.

  The niche-mediated model of oncogenesis is a multistep process; the initiating step occurs in the heterologous cells comprising the bone marrow stroma and leads to secondary genetic changes in other cells. Evidence supporting a niche-mediated model for benzene-mediated leukemogenesis comes from studies of myelodysplastic syndrome (MDS) and AML in exposed populations in Shanghai.[29] These molecular pathology and cytogenetic studies revealed immune-mediated inflammatory changes as well as a surprising lack of clonal cytogenetic abnormalities in the diseased cells. These findings highlight a critical role of the supportive bone marrow microenvironment in initiating benzene–mediated neoplasia. These suggest that immune-mediated targeting of bystander progenitor cells, not direct cytogenetic injury in hematopoietic stem cells, may be sufficient to result in the evolution of a neoplasm-initiating cell phenotype. This hypothesis is consistent with the evolving paradigm within the hematology community where a combination of intrinsic and extrinsic influences in the microenvironment plays an integral role in maintaining stem cell quiescence or driving proliferation and differentiation.

[28] Raaijmakers et al. (2011). Niche contributions to oncogenesis: emerging concepts and implications for the hematopoietic system. Haematolgica 96:1041-1048.
[29] Op. cit. footnote 26 and Irons et al. (2005). Chronic exposure to benzene results in a unique form of dysplasia. Leukemia Research, 29(12): 1371-1380

The Draft NexGen Report concludes (section 3.1.4) there is a strong association between the human molecular genomic signature and specific human disorders and diseases (presumably AML). However, more work needs to be conducted to support the associations proposed in the Draft NexGen Report. Considering the biological events mediating benzene-induced leukemia are unclear, the assumptions made in the benzene prototype call into question the relevance of the proposed systems biology approach. That is, the connections between the observed genetic perturbations and leukemia are unknown. Therefore it is unclear if the observed genetic changes are biomarkers of effect and precursors to disease onset, biomarkers of exposure, or irrelevant to disease progression. Without an understanding of the sequential causal events (i.e., mode of action) leading to leukemia and the linkages between the observed genetic changes and these sequential molecular events, it is difficult to distinguish true effects and adaptive responses from unrelated background signals. These compounding factors, in the context of evaluating risk, make the significance and utility of the NexGen benzene prototype unclear.

- **The Disease Progression Phase that the Gene Signature is Meant to Reflect is Unclear**

  It is unclear if the Draft NexGen Report is distinguishing between gene changes that precede the disease or gene changes that are a result of AML. Of course, recent paradigms use specific genetic anomalies as criteria for diagnosing specific leukemias, including AML. However, there is a strong possibility these genetic anomalies are the result of, rather than the cause of, specific subtypes. The Draft NexGen Report should clearly indicate, and provide the scientific support for, the stage of disease progression these genes are meant to represent. If the authors are indicating the genomic signature is reflective of gene changes made as a result of the disease rather than the cause of the disease, it is hard to comprehend how this could be useful for informing risk characterization. If however, the Draft NexGen Report is suggesting that these gene changes are indicative of disease initiation, the scientific support for this should be discussed. This is a critical distinction, particularly because the NexGen Draft Report case study employs a genomic signature from peripheral blood of exposed *individuals who do not have leukemia*.

III.C.2. Insufficient Support for a Causal Linkage between the Genomic Events and Apical Outcome

The degree to which the proposed genomic signature will be useful for informing benzene risk assessment depends upon the scientific data supporting a causal linkage between the genomic changes and leukemogenesis. The Draft NexGen Report cites the following observed similarities in pathway disruptions as additional evidence for a causal relationship between alterations in the specific gene signature and increased leukemia risk*:* "(1) caused by other chemical leukemogens, (2) observed in leukemia of unknown origins, and (3) reversed by certain leukemia chemotherapeutic agents." While qualitatively, these observations support an association of this gene signature with AML and possibly identify genes important for disease maintenance, such observations do not establish a causal relationship between specific gene/pathway alterations and benzene-mediated leukemia. To conclude causality, it is important

to understand the role of these gene pathways at various stages of disease progression and to be able to discern those that are necessary from those that are sufficient, those that initiate the disease from those that are responsible for maintaining it, and those that simply become fixed in the cell population with potentially little consequence. Association does not prove causation.

Connections between pathway perturbations and apical events can be viewed as a cascade of causative processes, with the outputs of the earlier processes constituting the causes of the later ones. There are a number of considerations that need to be addressed when developing a data basis for causality, one being dose-responsiveness.[30] The Exposure-Dose-Response Assessment section of the Draft NexGen Report concludes that the genomic signature varies in a dose-dependent manner. However it is unclear if this assertion has actually been scientifically demonstrated. The referenced McHale et al. 2011 report concludes that the expression of each of the 16 signature genes across the five exposure categories shows a pattern of dose-responsiveness, noting, however, the higher expression is actually found in the low exposure group. Therefore the dose-responsiveness of this gene signature is not entirely clear. More information should be provided to clarify the actual basis for the dose-responsive conclusion. Assuming the genomic response is dose-responsive, this alone is not enough to support causality. There are a number of additional considerations that would need to be systematically demonstrated to support causality. Therefore, the current state of the science is inadequate to support using such a signature to extend the exposure range of traditional epidemiology studies to lower exposures. Similarly, the state of the science is inadequate to support the authors' conclusions that the proposed genomic signature is both a biomarker of exposure and effect. The Draft NexGen Report must acknowledge that it remains to be determined if the altered gene pathways identified in this report are responsible for disease initiation, progression, maintenance, or are irrelevant. Without this understanding it is premature to conclude this genomic signature is a biomarker of effect and therefore appropriate for low dose risk extrapolation.

At this time it would be difficult to make a testable case for any particular direct pathway from primary genomic event(s) to leukemogenic (or other blood related) process(es). Current observations of genomic responses could or could not be related to leukemogenic responses or may be secondary to initial events and therefore more related to pathogenesis than primary effects (if any). It is therefore troubling that the report expresses confidence that 16 signature genes would form a "biomarker of exposure (and associated leukemia)" for future work (see page 18 of the Draft NexGen Report). While such exploratory work might be useful, at present it should continue to be viewed as hypothesis generating rather than hypothesis testing.

III.C.3. Relevance of the Test System Is Not Established

In order to use a systems biology approach for hazard identification and risk characterization, it is important the test systems and observations are appropriately anchored to the adverse outcome. As discussed above, it is important there is a scientific basis for linking the genetic changes to stages of disease initiation and/or progression to appropriately inform risk. In the approach described in the Draft NexGen Report, there are a number of limitations with the proposed test system that question its usefulness to risk assessment.

[30] Boobis et al. 2006. Crit Rev Toxicol. 36:781; Boobis et al. 2008. Crit Rev Toxicol. 38:87; Sonich-Mullin et al. 2001. Regul Toxicol Pharmacol. 34:146; USEPA, 2005. Guidelines for Carcinogen Risk Assessment, EPA/630/P-03/001F.

- **Limitations of the Cell Types Being Analyzed for the Genomic Changes Should Be Considered.**
  One limitation to the proposed genomics approach described in the Draft NexGen Report is the cell types being used in the genomics analysis. As noted in the Draft NexGen Report, exposure-dependent alterations in genes and pathways are often monitored in peripheral blood mononuclear cells (primarily lymphocytes). This is an important consideration as 1) the target cells are unknown and may be one of many comprising the bone marrow niche, and 2) the diseased cells are of the myeloid lineage residing in the bone marrow which is both a different biological compartment and cell type than what is being analyzed in these studies.

To conclude that the gene changes observed in these "surrogate" circulating peripheral blood cells are predictive of the cancer initiation events or representative of disease manifestation requires further justification. Without a clear connection between the cell types displaying the genomic signature (i.e., peripheral blood cells), the target cells and/or the diseased cells (e.g., hematopoietic stem cells), it is possible the genomic events reported are merely general measures of alteration of genetic material without known associations to toxicity (i.e. biomarkers of exposure).

- **Lack of Disease Progression in the Sample Population Exhibiting Genomic Changes Should Be Considered**
  Although the genomic signatures are appropriately observed in cells obtained from individuals who have been exposed to benzene, there is no indication if these exposed individuals with these gene changes develop leukemia. The Draft NexGen Report cites observed similarities in pathway disruptions in leukemia of unknown origins as well as by other chemical leukemogens. It is unclear, however, if these genomic changes are actually necessary for disease initiation. It is unknown if these genetic changes are early markers for the development of leukemia or are markers of exposure.

- **Dose Response of the Proposed Genomic Changes Is Poorly Characterized**
  Related to the point above and discussed in more detail earlier in this document, the dose response relationship for these gene changes is poorly characterized.
  Dose-responsiveness is one of the critical criteria for supporting a causal linkage between the genomic observations and the apical event. Characterizing the dose response is also critical to appropriately inform extending the exposure range of traditional epidemiology studies to lower exposures.

- **Appropriateness of the Disease Outcome Should Be Considered**
  The Draft NexGen Report states benzene causes MDS. However, benzene mediated MDS is not given appropriate weight in the report. One of the leading hypotheses in AML disease progression begins with benzene inducing myelodysplasia, which may take the form of MDS, followed by the subsequent evolution of either state to AML. Not including the MDS disease progression pathway in the benzene prototype is a major oversight. If MDS is the more appropriate risk measure, a systems biology approach focused on AML is problematic. Phenotypic anchoring to a proven histopathological precursor lesion, such as MDS, is an important proof of concept requirement that extends to any number of cancers that normally following a chronic inflammatory course in a specific tissue or organ.

III.D. Lack of Acknowledgement and Consideration of the Uncertainties for Predicting Hazards for Data-Poor Chemicals

The Draft NexGen Report discusses the potential utility of the genomic signature to inform hazard identification of potential leukemogens with limited data sets. However, it falls short of the report's articulated goal of describing a systems biology model suitable for informing hazard identification. Without an understanding of the linkage between the gene signature and disease progression, it is possible the genomic signature represents genes responsible for disease maintenance and not disease initiation. If this is the case, and these genes are relied upon for disease prediction (as proposed on p.25), this could potentially fail as a hazard identification indicator as the activation of these genes would not occur upon simple exposure, but would actually require disease manifestation. Furthermore, it is unclear in the Draft NexGen Report if the 16 gene signature is specific enough to AML to adequately predict disease risk for data-poor chemicals. For instance, if a majority of these genes are also activated in a general inflammatory response, they will not specific be enough for informing the leukemia risk for a data poor chemical.

Interestingly, the Draft NexGen Report acknowledges the inadequacy of the proposed systems biology approach to hazard identification stating, "Anchoring of the molecular patterns to apical outcomes, considerable systems biology knowledge, and high-quality data; however, appear necessary to define the disease signature against which data-limited chemicals could be compared. (sic)" However, this acknowledgement seems to have been ignored when generalizing the accomplishments of the tier 3 prototypes in the report. It is recommended the NexGen report more appropriately reflect these shortcomings in the overall conclusions and accomplishments of the tier 3 prototypes.

III.E. Lack of Acknowledgement and Consideration of the Uncertainties for Informing Low-Dose Extrapolation

As discussed in detail above, there are a number of limitations which make it premature to use the genomics data to inform risk at lower exposure levels. These deficiencies are 1) insufficient support for a causal linkage between the genomic events and the apical outcome; 2) an inadequate understanding of the dose-responsiveness of the genomic signature; and 3) unknown relevance of the current test system to informing benzene risk.

Without a scientifically substantiated link between the genomic signature and benzene-mediated leukemogenesis, it is unclear if indeed the identified genomic changes are a useful indicator of risk following benzene exposure.

IV. SPECIFIC COMMENTS ON SECTION 3.1.2: OZONE-INDUCED INFLAMMATION AND INJURY

Although the Draft NexGen Report indicates human studies are available to assess pulmonary inflammation due to ozone exposure, EPA only cites the single EPA-sponsored study by Devlin et al. Since the Devlin et al. study is used on a single high exposure concentration, it is

unclear how concentration response relationships at environmentally relevant concentrations will be evaluated.

An extensive review of the ozone case study has recently been published by Goodman et al. (2013),[31] and we refer EPA to this publication. Goodman and colleagues review the strengths and limitations of the ozone prototype and suggest recommendations for improvement. The key findings of this review, which ACC endorses, are paraphrased below:

- The ozone prototype does not provide a robust exploration of the literature or adequately discuss the limitations of the studies conducted, or those referenced to support its methods and choice of measured endpoints.
- The literature on ozone-induced lung inflammation contains several studies that have inconsistent results, use ozone levels that are not reflective of environmental exposures, and fail to measure endpoints that directly correlate to adverse health effects.
- The discussion of the relevance of the test systems used in the evaluated studies, as well as the environmental relevance of the tested ozone levels should be expanded.
- The scientific support for the inferences made concerning gene-environment interactions (e.g., GSTM1 null genotype) and other conditions (e.g., asthma) and their potential connection to increased susceptibility to ozone-induced lung injury should be cited and discussed in greater detail.
- The goal of proof-of-concept of applications of molecular biology data for ozone has not been attained; EPA needs to address missing information on the methods used and provide a discussion of uncertainties in the model before this tool can be used to accurately predict toxicity of ozone or any other pollutant.

V. SPECIFIC COMMENTS ON SECTION 3.1.3: BENZO[A]PYRENE (A POLYCYCLIC AROMATIC HYDROCARBON) AND CANCER

It is not clear that "proof of concept" was achieved with the benzo(a)pyrene (BaP) is prototype. This prototype originally focused on identifying whether human transcriptomics data from PAH mixtures found in cigarette smoke could be associated with lung cancer. However, even with 63 microarray studies located, for all the reasons listed in Box 7, EPA could not develop a suitable prototypical model. If this approach is meant to inform the hazard analysis for chemicals lacking data, then how can such an approach be relied upon if it could not be completed for a well-known hazard?

The microarray technology used in the case study was the transcriptomic data gathering platform employed in the prototype. Although Table 3 grouped all the microarrays, not all microarrays are equal. The difference between RNA and DNA microarrays might have different sensitivities to dosing schedules. A stringent set of methodological and reporting guidelines might be a requirement for standardizing hazard assessments using transcriptomic data. Such information might include whether the source material was DNA or RNA, whether the

[31] Goodman et al. (2014). Strengthening the foundation of next generation risk assessment Regulatory Toxicology and Pharmacology 68: 160-170.

appropriate protocols were followed (i.e., suggested replicate number used) and whether values were reported using standard units and methods. Lastly, microarrays will have an inherent selection bias and will only evaluate the genes selected by the researcher. This limitation could neglect important biological information which might be important for hazard assessment.

As noted (Box 7, page 34 of the Draft NexGen Report), a significant number of data limitations led EPA to modify the case study to instead focus on a single PAH, BaP, and liver cancer in mice. But serious data limitations are also present in the modified case study. First, the pathway model in Figure 14 (p. 35) is based on data from only two studies that used very high doses for short durations (e.g., 25 to 150 mg/kg BaP for up to three days) whereas human intake of BaP through the diet is estimated to be 0.16 – 3.3 ug/day or roughly 10,000 times less than the dose given to the mice.[32] Second, although the report states (p.38) that, "Due to the lack of data, speculating whether this system could be activated at low doses in the mouse is not possible," the report inexplicably, and without scientific justification, goes on to speculate that, "Due to genetic and epigenetic variability and potential species differences, these types of effects might occur at lower doses in humans than in mice."

Furthermore, it is not at all clear the extent to which the cellular systems and pathways based model for BaP liver cancer in mice are relevant to evaluating skin carcinogenesis in humans from PAH exposures. EPA should recognize that many lines of evidence suggest that a relevant organ system, human skin, is much less sensitive to the carcinogenic effects of PAHs than mouse skin. Differences between mouse and human skin include the following (from Crump, 2000)[33]:

- BaP and related PAHs induce the activity of the important, activating enzyme aryl hydrocarbon hydroxylase in both mice and humans, but the inducibility is some 40-75 times higher in mouse skin relative to human skin.

- BaP is transformed into the genotoxic metabolite 7,8-9,10-dihydrodiol epoxide — BPDE— in both mice and humans, but, at comparable BaP doses, mouse skin apparently produces considerably more of this metabolite than does human skin.

- BaP and related PAHs can induce DNA damage (via DNA-adducts) in both mice and humans, but these adducts are formed at higher rates in mouse skin relative to human skin; and the resulting damage is repaired at lower rates in mouse skin relative to human skin.

- Crump (2000) also provides extensive discussion of the published literature on metabolism of BaP, demonstrating that mouse skin produces at least 120 times more total BaP metabolites than human skin.

[32] Benford et al. (2012). Application of the margin-of-exposure (MOA) approach to substances in food that are genotoxic and carcinogenic e.g., benzo[a]pyrene and polycyclic aromatic hydrocarbons. Food Chem Toxicol 48(suppl 1) S42-S48. FAO/WHO (Food and Agriculture Organization of the United Nations, World Health Organization) 2006. Evaluation of Certain Food Contaminants. WHO Technical Report Series.

[33] Crump et al. (2000). The K.S. Crump Group, Inc. Estimation of lifetime skin cancer risk from the use of coal tar-containing shampoos. ICF Consulting, 602 East Georgia Street, Ruston, Louisiana 71270. (Available http://www.tera.org/ITER/Coal%20Tar/CoalTar.pdf. Accessed Nov. 13, 2013.

- BaP causes skin tumors in mice, but it does not cause tumors in human skin grafted onto mice.[34]

- In cell culture studies, BaP transforms mouse skin cells into malignant cells, but it does not transform human skin cells.[35]

Additionally, there is a significant literature on people who use coal tar pharmaceutical products for therapy of psoriasis and eczema (2 – 5% coal tar containing 0.5% BaP). Roelofzen et al. (2010)[36] evaluated 13,200 patients treated with coal tar for skin conditions. They were compared with patients treated with corticosteroids. The median exposure to coal tar ointments was 6 months (range 1–300 months) and the median duration of follow-up was 21 years. Coal tar did not increase the risk of non-skin malignancies (hazard ratio (HR) 0.92; 95% confidence interval (CI) 0.78–1.09), or the risk of skin cancer (HR 1.09; 95% CI 0.69–1.72). This study has sufficient power to show that coal tar treatment is not associated with an increased risk of cancer. In fact the FDA[37] concluded: "Upon reviewing the published studies, the agency does not find that there is evidence to implicate the use of OTC coal tar containing drug products as an independent risk factor for the development of skin cancer."

Considering these issues, although informative, the well documented differences between mouse and human skin suggest that the liver cancer model for BaP should be viewed as preliminary. While the pathways may prove to be fundamentally correct in both species, the modifying factors evident in BaP induced skin cancer demonstrate the need for extreme caution in using this model for human risk assessment of BaP or PAH-containing mixtures. The inability to predict the dermal carcinogenic potential for various complex substances based either on the BaP concentration or the BaP-equivalents content has been demonstrated in a number of studies.[38]

VI. SPECIFIC COMMENTS ON SECTION 3.1.4: RISK ASSESSMENT IMPLICATIONS ACROSS THE TIER 3 PROTOTYPES

With respect to the Tier 3 prototypes, it is not clear how the apparent overarching conclusions were arrived at, what processes were used, and whether independent external input

[34] Urano et al. (1995). Failure of genotoxic carcinogens to produce tumors in human skin xenografts transplanted to SCID mice. Carcinogenesis 16(9):2223–2226.

[35] Sala et al. (1987). Morphological transformation in three mammalian cell systems following treatment with 6-nitrochrysene and 6-nitrobenzo(a)pyrene. Careinogenesis, 8, 503-507; Fox et al. (1975). Metabolism of Benzo(a)pyrene by human epithelial cells in vitro. Cancer Research 35:3551-3557; Kuroki et al. (1989). Use of human epidermal cells in the study of carcinogenesis. J Invest Dermatol 92(5 Suppl):271S–274S.

[36] Roelofzen et al. (2010). No increased risk of cancer after coal tar treatment in patients with psoriasis or eczema. J Invest Dermatol. 130(4):953–961.

[37] FDA. (2001). Letter from Dennis E. Baker, Associate Commissioner for Regulatory Affairs, FDA to Perry M. Gottesfeld, Occupational Knowledge International. Docket No. 00P-1210. February.

[38] Grimmer et al (1984). The contribution of polycyclic aromatic hydrocarbons to the carcinogenic impact of emission condensate from coal-fired residential furnaces evaluated by topical application to the skin of mice. Cancer Letters 23:167-176; Nesnow et al. (1982). Comparative tumor-initiating activity of complex mixtures from environmental particulate emissions on SENCAR mouse skin. J Natl Cancer Inst 68(5):829–834;Warshawsky et al. (1993). Factors affecting carcinogenic potential of mixtures. Fundamental & Applied Toxicology 20:376-382. Wright et al. (1985). Comparative chemical and biological analysis of coal tar-based therapeutic agents to other coal-derived materials. J Appl Toxicol. 5(2):80–88.

was involved. Our analyses of the Tier 3 prototypes raise issues that should be addressed in each case study. We recommend that EPA review stakeholder comments, revise the prototypes as warranted and then propose a revised set of overarching "implications."

Appendix A cites at least 7 reports that are used as the basis for the Draft NexGen Report, but those reports are "in preparation." A reviewer cannot assess the appropriateness of the interpretations based on those "in preparation" reports. Furthermore, it appears that not all of the reports in Appendix A are actually referenced within the Draft NexGen Report. If this is the case, we recommend that EPA clearly map the Appendix A reports to the relevant NexGen report text (i.e., ensure the Appendix A references are noted within the body of the text).

VII. SPECIFIC COMMENTS ON SECTION 3.2.1: KNOWLEDGE MINING

In general, the Draft NexGen Report discussion of the many challenges in studying the genomic variability in human cells or humans is not as balanced as the subject warrants. It leaves the incorrect impression that in the near future, results from these studies are sufficiently reliable or directly relevant for risk assessment purposes. This optimism should be tempered by the National Cancer Institute National Genome Research Institute and Genome-Environment Think Tank (NCI–NHGRI) report assessment that "despite many years of candidate gene studies testing for gene-environment interactions in cancer, there are only a few notable replicated and widely agreed-upon examples of successes".[39]

In the discussion of epidemiology studies evaluating variability in human response to chemical exposures, there is no discussion of many of the problems that have plagued candidate gene studies testing for gene-environment interactions according to the NCI-NHGRI.[40] For example, there is no mention of the large sample size (i.e., >1000+ cases) that would be needed to accommodate the large number of variants (e.g. SNPs) or patterns (e.g., gene expression levels). The multiple testing penalty is extremely high for SNPs; if 1 million SNPs are examined to determine if the allele frequencies of each SNP is different between the cases and controls, then a p-value of 10^{-7} or better is required to be certain that the association is not a false positive result.[41] Many epidemiology studies testing for gene-environment interactions suffer from marginal association, small sample sizes, insufficiently stringent thresholds for statistical significance, incomplete genetic coverage, and publication bias. Consequently, NCI convened workshops to develop criteria for evaluating the soundness of gene-environment interaction studies.[42] These criteria should be discussed in the report and used to analyze genome-wide association epidemiologic studies in the prototype examples.

In EPA's 2012 Framework for Human Health Risk Assessment to Inform Decision Making,[43] EPA highlights "…the importance of transparency of its human health risk assessment

[39] Hutter et al. (2013). Gene-environment interactions in cancer epidemiology: a National Cancer Institute Think Tank report. Genet Epidemiol. 37(7):643-57.
[40] Op. cit footnote 39 and Hutter et al. (2007), National Cancer Institute National Human Genome Research Institute (NHGRI) Working Group on Replication in Association Studies, Replicating genotype-phenotype associations. Nature 447 (7145):655-60.
[41] Ibid.
[42] Ibid.
[43] EPA. (2012). EPA External Review Draft Framework for Human Health Risk Assessment to Inform Decision Making July 12, 2012). http://www.epa.gov/raf/files/framework-document-7-13-12.pdf.

and decision-making processes." Yet the section on high-content knowledge mining is replete with jargon that is not familiar to most in the risk assessment community. Further, the tools described as integral components of high-content knowledge mining have been used by only a small number of researchers. Thus, it is fair to state that the approach is at an early exploratory phase and has yet to be vetted by the broader scientific community. For example, Patel et al. (2013)[44] noted that their "…top findings are ideal candidates for extensive validation through replication in higher-powered investigations." On a related note, although several topics in the document include descriptions of the uncertainties and limitations associated with the risk assessment approach, and despite the exploratory nature of high-content knowledge mining, the epidemiology section does not contain such a description. Despite the fact that the numerous shortcomings and issues related to the use of cross sectional databases for drawing inferences about associations between exposure and disease have been well documented, the Draft NexGen Report does not clearly describe these.

An example of the type of information that is needed in the section on high-content knowledge mining, in order to highlight the cautions needed in doing this type of explanation is found on page 16, Box 4. Although this example addresses omics results, a similar set of caveats are required for the data mining described in this section. The quote from page 16, "Although results presented here are promising, robust understanding and full implementation of new methods in general practice, and might take years, subject to the resources available for data generation and evaluation," applies equally to high-knowledge content mining, particularly when using cross-sectional databases.

Specific comments on Section 3.2.1 follow:

- Section 3.2 should have a section on Challenges such as the one included on pg. 27.

- Pg. 46, Table 4: This table includes the following language related to use of existing epidemiology databases: "Relationships generally associative; might be causal in certain circumstances (depending on data quality and amount of underlying evidence)." However, many of the large existing datasets such as NHANES are cross-sectional and cannot be used for assessing causality. This should be made clear.

- Pg. 48, table 5: It is unclear why the BPA/obesity relationship was omitted, as this outcome was noted for phthalates in the same table. Thayer et al. (2012)[45] concluded that "It was not possible to reach clear conclusions about BPA and obesity from the existing animal data." In terms of the need for databases for use in high content knowledge mining, this is an important conclusion as BPA is one of the most extensively studied environmental chemicals. If there are insufficient data with which to draw conclusions regarding associations between obesity and BPA, a legitimate question can be raised about the number of chemicals for which there are sufficient data to explore the knowledge mining approach.

[44] Patel (2013). Systematic identification of interaction effects between genome- and environment-wide associations in type 2 diabetes mellitus. Hum Genet 132:495–508.
[45] Thayer et al. (2012). Role of environmental chemicals in diabetes and obesity: A National Toxicology Program Workshop review. Environ Health Perspect 120: 779-789.

- Pg. 49, line 1: The Environment-Wide Association Study (EWAS) approach is an interesting and potentially exciting approach to utilizing available data in a more extensive fashion and for addressing the difficult issue of mixtures. However, it should be recognized in the report that the approach has only been used by a very limited number of investigators and requires further development and evaluation.

- Pg. 49, line 21: "Itemset associations": This approach is apparently used by businesses (as evidenced on pg. 49). However, issues are raised by the following statement: "Similarly, this technique can be used with the NHANES data to uncover a chemical or group of chemicals that tend to be associated with specific diseases." First, EPA has not described why this approach has advantages over more commonly used statistical approaches. Second, it is not clear that this approach is commonly used by the scientific community, raising questions about developing approaches that are transparent and that can be vetted by the scientific community. In fact, EPA does not provide any specific examples with citations for instances where this approach has been used successfully outside of the business world. While the mathematics underlying the approach may be valid across disciplines, does this approach address the complexities of human biology and environmental chemistry (e.g., confounders, variability in chemical concentrations)? This should be addressed in the report.

- Pg. 49, line 32: The Draft NexGen Report does not discuss NHANES survey years 2005/2006 and 2007/2008. Were there problems in conducting the analyses for these survey years? If so, could these problems illuminate broader issues that need to be addressed before this approach can be used?

- Pg. 50: EPA should provide a better and more transparent rationale for weighting the relative importance of lift, support and confidence than "When interpreting lift, support, and confidence, *we believe* lift is the most informative to start with" (emphasis added).

- Pg. 49-52: Overall, the approach is difficult to evaluate or even understand as the report does not include basic information such as how "high" levels of the chemical are defined. The report should define the terms used in this section (another example occurs on pg. 52, line 5, where the definition of "elevated" is not given). Further, if the definition of "high" were to be modified (e.g., a change in the percentile cutoff), the report should describe the impact on the association (i.e., a sensitivity analysis is needed). Finally, the impact of examining diabetics only (rather than combining diabetics and prediabetics) should be included. More broadly, justifications for selection of outcomes should be provided.

- Pg. 49, line 14: "…Patel et al. (2013, 2012b) successfully identified association linking diabetes, genes, gene variants, and environmental factors. This approach demonstrates a knowledge mining method that can be applied broadly to any number of common diseases to identify possible interactions between genetic and environmental factors and risks of disease." EPA needs to carefully define "successfully." For example, is an association between gene expression and data on a short-lived chemical biologically meaningful? If samples are not collected simultaneously, what limitations does this place on EPA's ability to interpret the results? In examining associations between exposure

and outcomes for chronic disease and chemicals with short half-lives, EPA should define the minimum requirements for intraclass correlation in order for the exposure information can be used to relate exposures to chronic outcomes.

- Pg. 49, line 18: The document should be more explicit about the limitations of the use of large-scale cross-sectional databases. For example, they are useful for hypothesis-generation, but cannot be used for causal inference (e.g., "risk of disease").

- Pg. 49, line 22: It is unclear why EPA chose to combine prediabetes and diabetes. Issues associated with this approach have been described previously (e.g., Wei 2009; LaKind et al. 2012).[46,47] These kinds of decisions will require careful scrutiny by those attempting to use high-content knowledge-mining for decision-making. At a minimum, the impact of this kind of decision on reported associations will need to be explored.

- Pg. 51, line 4: "This rule is present in 11% of the 2009–2010 NHANES cohort, suggesting it might be true for 11% of the U.S. population at the time of study, assuming NHANES is a good random sample." As noted previously, this also assumes that the levels measured are representative of long-term exposure.

- Pg. 52, Table 7: Issues associated with reports of associations between chemicals with variable concentrations and chronic outcomes have been described previously (LaKind et al. 2012). However, the limitations of approaches based on cross-sectional databases cannot be ignored. This is especially true for chemicals with demonstrated high temporal variability in human exposures. And as noted by Barr et al. (2006)[48]: "Multiple measurements over the duration of exposure-related activities must be taken to determine peak exposures. To relate the biomonitoring measurement to a health outcome, other required information may include population susceptibility factors, plausible mechanism of toxicity, and information on whether the exposure evaluated preceded the development of the health effect in question." This type of discussion, along with an assessment of how these problems will be addressed, should be included in the report.

- Pg. 59, line 5: EPA states that "…many biological functions and disease pathways are conserved across species…" While there are some pathways conserved across species, this statement needs to be qualified, as many pathways are not conserved across species.

- Pg. 59, lines 12-19: EPA is correct in its support of further research to explore the approach for prioritization and safety assessment proposed by Thomas et al. (2013a,b).[49] This approach holds considerable promise, especially for data-poor substances. This approach does not predict an apical effect based on the affected pathways, but simply capitalizes on the correlation between the lowest transcriptional BMD based on a select list of tissues and the BMD identified for cancer and non-cancer endpoints from

[46] Wei. (2009). Association of bisphenol A with diabetes and other abnormalities. JAMA 301(7):720.
[47] LaKind et al. (2012). Use of NHANES data to link chemical exposures to chronic diseases: a cautionary tale. PLoS One. 2012;7(12):e51086.
[48] Barr et al. (2006). Biomonitoring of exposure in farmworker studies. Environ Health Perspect 114(6):936–942.
[49] Op cit. footnote 16.

conventional toxicity studies. The most sensitive BMD is compared with human exposures to determine if the MOE is acceptable.

VIII. SPECIFIC COMMENTS ON SECTION 3.2.2: SHORT-TERM *IN VIVO* MODELS – ALTERNATIVE SPECIES

Overall, this section falls short in addressing the challenges of developing, standardizing and validating models employing alternative species to assess chemical risks to humans and other species. While several methods and promising approaches are cited, a number of critical aspects are underemphasized or not addressed. For example, species differences in absorption, distribution and metabolism and excretion (ADME) are not mentioned, even though it's well-known that failure to consider ADME differences across species and routes of exposure can lead to erroneous conclusions regarding potential hazards and risks.

The statement that "Both the European Chemicals Agency (ECHA) and EPA use alternative species tests as part of required tests for endocrine disruptors (EPA 2012e, 2009a)" is incorrect. Currently there are no specific required tests for endocrine disruption under REACH regulation; scientific criteria (and tests) suitable to identify and characterize endocrine disruptors for regulatory purposes within EU pieces of legislation have yet to established and are under development. In EPA's current EDSP, ecotoxicology tests using frogs, fish, birds, etc. are being used, but these are not the same as "alternative species." These EDSP test systems are specifically intended to be used in screening and testing for ecological risk assessment purposes. ToxCast and related approaches are included in EPA's EDSP21 plan, but EDSP21[50] is planned to be phased in and used only after scientific confidence in such methods for their intended use have been established.

VIII.A. Specific Comments on Tier 2 Prototype: Using Alternative Species to Identify Thyroid Hormone Disruption

It is well established that physiological differences in thyroid hormone systems among alternative species, rodents and humans are key factors in determining potential responses to thyroid active substances. Yet these are not described in the Draft NexGen Report. This is a major shortcoming. The significant differences in thyroid hormone regulation between rats and humans, let alone alternative species and humans, limit the scientific confidence in applying the thyroid pathway prototype. For example, thyroxine-binding globulin is the predominant plasma protein in humans and non-human primates, but not rodents. The half-life of T4 is 12 hours in the rat versus 5–9 days in humans, and the serum level of TSH is 25 or more times higher in the rodent than in humans.[51] The rat also exhibits enhanced thyroid hormone elimination with less efficient enterohepatic recirculation compared to humans. The histology of the non-stimulated rodent thyroid is similar to that of the stimulated human gland, with small follicles lined by tall follicular cells. Thus, the physiological parameters and the histological appearance indicate that the rodent thyroid gland is more active and operates at a higher level with respect to thyroid

50 EPA. (2011). Endocrine Disruptor Screening Program for the 21st Century: (EDSP21 Work Plan). http://www.epa.gov/endo/pubs/edsp21_work_plan_summary%20_overview_final.pdf,

51 Dohler et al. (1979). The rat as model for the study of drug effects on thyroid function: consideration of methodological problems. Pharmacol Ther B. 5(1-3):305-18; McClain. (1992). Thyroid gland neoplasia: non-genotoxic mechanisms. .Toxicol Lett. 64-65 Spec No:397-408.

hormone turnover as compared to the human gland.[52] Even though pathways in the hypothalamus-pituitary-gonad axis are highly conserved among vertebrates, the implication that thyroid effects in fathead minnows are predictive of endocrine disrupting effects in humans is called into question because of the significant differences in thyroid hormone regulation in humans. It is incorrect to imply that reduction in circulating thyroid hormones in animal test systems, often transiently, is an indicator of thyroid endocrine disruption potential in humans or that this response poses a neurodevelopmental risk to humans.

EPA should not rely on the WHO/UNEP endocrine report which was issued in February 2013[53] incidence of thyroid disorders and alleged causation by environmental chemicals. This report failed to use best practices of systematic evidence-based review; specifically, the report authors did not employ objective criteria for determining data quality and study reliability, and the previous objective and transparent weight of evidence approach established by WHO [54] was inexplicably abandoned and substituted with the judgment of the authors.

IX. SPECIFIC COMMENTS ON SECTION 3.2.3: SHORT-TERM *IN VIVO* MODELS – MAMMALIAN SPECIES

Overall, this section falls short in describing the current state of the science with respect to understanding and using transcriptomics in risk assessment. While there is great interest and promise in using transcriptomics from *in vivo* exposures to predict adverse effects, it's a misstatement to claim that, "In addition, transcriptional changes in the most sensitive pathway were also highly correlated with the apical responses (see Figure 21)." The most "sensitive" pathway is described in the Thomas et al. papers as the pathway that exhibits change at the lowest dose. But there is no attempt to correlate the most sensitive pathway as a key event or causal step leading to the apical outcome. It is acknowledged by the investigators[55] that such pathways could be compensatory mechanisms, but it is not discussed in any detail.

Furthermore, on page 61, lines 26-29, EPA states that, "For hazard identification, a host of previous studies has demonstrated that transcriptomic signatures from short-term *in vivo* studies can be used to predict both subchronic and chronic toxic responses." At a minimum, the literature supporting such a statement should be cited. In particular, other investigators have concluded that there is an inconsistent correlation between transcription and histopathology.[56] A more accurate description is that applications of transcriptomics in risk assessment are still under investigation and will require additional development and validation through collaborative

[52] Dellarco et al. (2006). Thiazopyr and thyroid disruption: Case study within the context of the 2006 IPCS human relevance framework for analysis of a cancer mode of action. Crit. Rev. Toxicol.36:793-801.
[53] WHO/UNEP (2013). State of the Science of Endocrine Disrupting Chemicals. Retrieved from http://apps.who.int/iris/bitstream/10665/78102/1/WHO_HSE_PHE_IHE_2013.1_eng.pdf .
[54] WHO/IPCS. (2002). Global assessment of the state-of-the-science of endocrine disruptors. http://www.who.int/ipcs/publications/new_issues/endocrine_disruptors/en/,
[55] Thomas et. al. (2007). A method to integrate benchmark dose estimates with genomic data to assess the functional effects of chemical exposure. Toxicol Sci. 8(1):240-248.
[56] Foster et al. (2007). A retrospective analysis of toxicogenomics in the safety assessment of drug candidates. Toxicol Pathol. 2007 Aug;35(5):621-35.

science across a number of disciplines, including molecular biology, engineering, mathematics, chemistry, physics, statistics, and computer science.[57]

We agree that the most promising approaches are those described in the Thomas et al. publications, in which a tiered approach, based on pathway responses (ToxCast and transcriptomics), are interpreted in conjunction with exposure information to support a margin of exposure decision as to whether further evaluation in traditional toxicity tests is warranted. EPA should develop a research plan which outlines activities for expanding the domain of applicability and extending the correlational analysis in order to attain the degree of scientific confidence needed for regulatory acceptance of this method.

X. SPECIFIC COMMENTS ON SECTION 3.3: TIER 1 SCREENING AND PRIORITIZATION

Section 3.3 summarizes approaches which characterize "screening and prioritization" under Tier 1. In the Executive Summary, the impression is that only HTS approaches are being considered as part of Tier 1, when in reality both QSAR and HTS methods are implied. This could be made more explicit in the Executive Summary and in the start of Section 3.3 itself – "new approaches" is too generic a term.

The focus of Tier 1, as is noted in the Executive Summary, is to "explore entirely HT approaches that could be applied to thousands or tens of thousands of chemicals, are the least resource intensive and are likely to have the greatest uncertainty." We agree that, in the short-term, the most promising potential application for these methods is in priority setting and screening. While the degree of scientific confidence may be less for priority setting applications than for hazard and risk determinations, there is still a need document the validity of the assays, methods and prediction models used for priority setting and screening. This section does not adequately address this.

The "Qualification" step outlined in Patlewicz et al, (2013) is similar to the statement in the Draft NexGen Report that "The analyses provide the anchoring information critical to characterize relevance of a "hit" in an assay. Documenting the linkage from assay endpoint to MIE [molecular initiating event] to potential for adversity is key to evaluating the relevance of each assay that might be used as part of a Tier 1 risk assessment." What is not clear from these sentences is specifically what type of analyses will be conducted and how an evaluation will be made for the suitability and relevance of using a HTS assay relative to an MIE within an AOP. The text alludes to statistical modeling using *in vivo* and *in vitro* data but arguably this is only one approach that could be informative. The authors could cite the work under the OECD AOP Work Programme or examples that are being elucidated within the EU SEURAT programme to highlight other approaches that are being investigated to develop AOPs and their application role. Further the statement "linkage from assay to MIE to adverse endpoint" implies that a single assay could be relied upon to make this inference which is unlikely; a battery of assays characterizing different key events within an AOP and integrated in some fashion in a prediction model is more likely an approach.

[57] Afshari et al. (2011). The evolution of bioinformatics in toxicology: advancing toxicogenomics. Toxicol Sci. 120 Suppl 1:S225-37.

In Section 3.3 the scope of the Tier 1 is seemingly broadened from screening and prioritizing large numbers of chemicals to identify those for subsequent more in-depth testing and evaluation to also include use in "…immediate regulatory decision" (page 64 of the Draft NexGen Report). Given, as noted in the Draft NexGen Report, that the approaches under Tier 1 are likely to have the greatest uncertainty, it is not clear which Tier 1 approaches are sufficiently scientifically robust for immediate use in regulatory decision-making. EPA should consider providing examples along with citations which support scientific confidence in, and regulatory acceptance of, such methods for specific regulatory decisions.

Overall Table 8 ("Summary of Tier 1 NexGen Approaches, Including Weight of Evidence, Pros, and Cons") provides a generalized view of the status of QSARs that is not well balanced. There are endpoints for which QSARs are very well developed e.g., skin sensitization, Ames mutagenicity and others where the models are lacking or should not really be attempted (e.g., developmental toxicity). Specific comments on Table 8 include:

- Under the "pros" section, the statement "If the basis for the QSAR model(s) matches the physical chemistry of the evaluated chemicals, the model(s) generally predicts potency within a factor of 100." should be substantiated with citations.

- It is unclear want is meant by the statement "Harmonized international approaches are available" There are OECD Validation Principles for (Q)SARs to facilitate the consideration of (Q)SARs for regulatory purposes but it is not clear whether this is what is being implied.

- Under the "cons" section, the draft NexGen Report states that "if the models do not match the physical chemistry of the evaluated chemicals, results are unreliable." Again this appears to be a very specific statement. It is unclear if the statement is implying that MOA is known and is related to inherent reactivity of the parent or metabolite.

- Reference should be made to the OECD principles for QSARs-- specifically the need to characterize the applicability domain of the model of interest in order to facilitate an evaluation of whether a prediction for a given substance could be made with any degree of reliability. The Draft NexGen Report also states that "Models do not predict active metabolites." This is an over generalization. Some models are able to simulate likely metabolites and make predictions for both parents and metabolites. Expert systems are available to predict likely metabolites. While this is an evolving and challenging area, efforts are being made.

- Under the "pros" section related to validated HT *in vitro* assays, the Draft NexGen Report states that "False negatives and positives for ToxCast™ evaluated assays are low." As noted previously, this statement is at odds with the conclusions drawn by Thomas et al. 2012[58] and the preliminary analysis by CLA/EPF[59] of data generated under EPA's EDSP. At a minimum, references or more detail should be provided to substantiate this claim.

[58] Op. cit. footnote 10.
[59] Op. cit. footnote 11.

- We note the header of Table 8 includes the word "validated" which is encouraging. While some notion of validation is alluded to early in Section 3.3, no specifics are provided to qualify what validation represents or how it will be undertaken. The Agency should adopt the framework developed by Patlewicz et al. (2013), or a similar approach, for evaluating and documenting scientific confidence in HT/HC methods and their prediction models.

XI. SPECIFIC COMMENTS ON SECTION 3.3.1: QSAR AND HIGH-THROUGHPUT VIRTUAL MOLECULAR DOCKING (HTVMD)

- Page 67, lines 1-10: (Q)SAR models are described as regression or pattern recognition models. This is a narrow definition of what (Q)SAR models represent or how they are developed. The reference cited, Hansen et al. (2011), which discusses "increased methylation variation in epigenetic domains across cancer types," does not provide the best definition for QSAR. There are many QSAR practitioners and papers in this field, as well as OECD guidance and documentation that could have been cited instead to provide a less biased perspective.

- Page 67, lines 11-13: The Draft NexGen Report states: "QSAR models have been most commonly used in classification of chemicals by comparing the "query" chemical's inherent properties with similar properties for a set of chemicals (the "training set")." "Again this is a very limited perspective of how QSARs are developed and applied. Some of the references appear to have been taken out of context, e.g., citing the OECD Toolbox appears to be erroneous in this case. Moreover, it is unclear what is meant by "classification." It could be interpreted to mean classification and labeling obligations which is likely not the intended meaning in this context.

- Page 67, lines 16-18: TOPKAT is only one example of a commercial expert system; there are a number of others such as Derek, Leadscope, MCASE, TIMES etc. that could have been referenced as well. In fact, EPA's own suite of tools is not explicitly mentioned. The OECD Toolbox is not a QSAR tool per se, hence should not be referenced as a "suite of models," which implies it houses a suite of statistically based QSAR models. The OECD Toolbox is a tool that facilitates the development, evaluation, justification and documentation of chemical categories. It contains some QSAR models such as those developed by the Danish EPA and those by the US EPA to facilitate data gap filling within analogue and category approaches and certain third party expert systems such as that of TIMES or Catalogic can be docked with the OECD Toolbox to enable data gap filling.

- Page 67, lines 18-19: The Draft NexGen Report states that in the EC, (Q)SAR results are used to prioritize chemicals for additional toxicity testing. This is not wholly correct. REACH permits QSARs to be used to satisfy information requirements without the need for additional testing if a number of conditions are met. This sentence is incorrect. Specific examples of who/where/how QSARs are being used in other regions outside of the US would be instructive.

- Page 67, lines 30-32: The Draft NexGen Report states: "These models are increasingly being used in risk assessment…" Examples would be helpful to clarify and specify existing uses. It might be reasonable to note that such models have been "proposed" in the EDSP. The language in this paragraph appears to be more aspirational than reflective of reality.

- Page 68, lines 6-19: The discussion should be revised to more accurately describe the OECD Toolbox. The OECD Toolbox is not a QSAR software package, nor does it merely permit the construction of chemical categories for screening purposes. The OECD Toolbox is a tool that facilitates the development, evaluation, justification and documentation of chemical categories. It contains some QSAR models such as those developed by the Danish EPA, and those by the US EPA to facilitate data gap filling within analogue and category approaches. Certain third party expert systems such as that of TIMES or Catalogic can be docked with the OECD Toolbox to enable data gap filling, and is not restricted to screening. The approach to develop categories is also not limited to structurally similar substances. Indeed the technical guidance associated with the Toolbox promotes the use of endpoint specific read-across and highlights how structural similarity might not be a good tenet in every case. The Toolbox does not estimate potential toxicity per se – it provides a workflow to facilitate the development of categories, data gaps can then be filled either by using qualitative or quantitative read-across, trend analysis or external QSARs. Read-across is a data gap filling technique used with analogue and category approaches and is not a means to construct categories. Similarly trend analysis is a data gap filling technique not an approach to construct categories.

- Page 68, lines 12-14: The Draft NexGen Report states, "The popularity of read-across is driven by its relative simplicity and the availability of the QSAR Toolbox online." This statement is not an accurate perspective on read across. Read-across is not simple. Nominally, read across it is "accepted" by regulatory authorities but the practical reality of documenting and justifying read across is quite a different story. The report should note some of the challenges in read across as discussed in Patlewicz et al. (2013b,c).[60]

- Page 68, lines 13-51: The OECD ratified principles for the validation of QSARs, to facilitate their consideration for regulatory purposes in 2004. These were first proposed in Setubal at an ICCA-LRI workshop in 2002. Guidance to clarify these principles was subsequently developed first by the JRC and then refined and published by the OECD in 2007. Lines 13-15 do not convey this. OECD 2004 refers to the principles alone and not the accompanying guidance.

XII. SPECIFIC COMMENTS ON SECTION 3.3.2: HIGH-THROUGHPUT AND HIGH-CONTENT ASSAYS

- There is nothing in this short section on HT/HC that specifically describes the "validation and evaluation of HT/HC assays" prior to their use. EPA needs to propose a framework

[60] Patlewicz G, Roberts DW, Aptula A, Blackburn K, Hubesch B. Workshop: Use of 'read-across' for chemical safety assessment under REACH. Reg. Toxicol. Pharmacol. 2013, 65(2): 226-228.; Patlewicz G, Ball N, Booth ED, Hulzebos E, Zvinavashee E, Hennes C. Use of Category approaches, Read-across and (Q)SAR: General considerations. Reg. Toxicol. Pharmacol. 2013, 67(1): 1-12.; ECETOC Technical Report 116.

for developing scientific confidence in the high throughput and high content methods, proposed in the Draft NexGen Report. The framework developed by Patlewicz et al. (2013a) should be considered by the Agency for validating these HT/HC methods.

- Page 68, line 2: HT/HC have potential to be major tools in the early evaluation of chemicals but in a commodity chemical setting, they are not yet routinely used.

- Page 68, lines 40-41: It is unclear whether the categorization of HTS assays into 2 types is a general characterization or an EPA definition.

- Page 69, line 4: Dick et al. (2010) is not the best reference in support of the varying types of cell-based assays, since this reference is just one example of an assay. Other examples or a broad review article would be more appropriate.

- Page 69, lines 10-12: There are other methods being used that should be considered for discussion in this section, such as use of varying culture physical configuration (e.g., use of a collagen or extracellular matrix sandwich model) to enhance differentiation capability of primary hepatocytes. It is not clear if this is amenable to robotic level analyses, but it might still fit the definition of HTS if organism level screens are being annotated as HT.

XIII. SPECIFIC COMMENTS ON SECTION 3.3.3: TOXICOKINETICS

- Page 69, line 19: Potency values among *in vitro* assays have indeed been proposed for use in hazard identification but much more work needs to be done to evaluate whether the associated prediction models being developed are truly predictive.

XIV. SPECIFIC COMMENTS ON SECTION 3.3.4: HIGH-THROUGHPUT EXPOSURE ESTIMATION: EXPOCAST PRIORITIZATIONS

The Draft NexGen Report is generally lacking information on how to incorporate exposure into the next generation of risk assessment. To date, ExpoCast has directed most of the effort to far field environmental models. These models are appropriate tools for estimating exposure to various environmental compartments and to human exposure from the environment. However, as shown in Wambaugh et al. (2013)[61] focusing on far field only results in few cases of exposure concern; the highest exposure prediction was 1E-04 mg/kg/day for the almost 2,000 chemicals ranked. One of the challenges for high throughput exposure estimation is addressing all the relevant sources of human exposure.

More relevant to human exposures are the near field exposures from occupational exposure, use of consumer products and indoor environment exposures, and we recommend EPA focus on these. There are many models that can predict such exposures based on physical chemical properties and type of use (EFAST, ECETOC TRA, AISE model, EGRET model, CONSEXPO, etc). In the REACH program of the EU, the Chemical Safety Assessment and

[61] Wambaugh et al. (2013). High-throughput models for exposure-based chemical prioritization in the ExpoCast project. Environ Sci Technol.6;47(15):8479-88.

Reporting tool (Chesar)[62] uses ECETOC TRA models to estimate worker and consumer exposures. The consumer model in ECETOC TRA (AISE and EGRET models are based on ECETOC TRA with additional refinements) is an extremely conservative model that has 46 consumer use categories designed for the REACH regulation implementation which required exposure assessments for the life cycle of the substance. The model has defined sentinel exposures for each use.

The ILSI-HESI RISK 21 project has been looking at tiered risk assessment approaches and for prioritization or Tier 0 suggests using the banding approach. The consumer exposure model is based on banding and there are 4 vapor pressure bins which determine total exposure (oral, dermal and inhalation). Tier 0 banding tables were developed for both the ECETOC TRA and the EGRET model (example shown below for ECETOC TRA version 3 – the saturated vapor pressure concentration is not invoked as an upper bound, publication in preparation). The same approach was developed for worker exposures. Substances could easily be assigned a vapor pressure bin based on known vapor pressure or predicted vapor pressure. This would lead to a ranking of external exposures which could be compared to external health benchmarks. For comparison to validated HTS *in vitro* assays the external exposure would need to be converted to an internal exposure and this could be based on a generic PBPK approach even though this would add additional uncertainty. This would enable the identification of the most hazardous substances with the highest likelihood of exposure. These substances and their uses could then be looked at more explicitly in a Tier 1 or 2 approach depending on the available data. The same models can be used for Tier 1 when more information such as weight percent (wt. %) in product is available.

Descriptor	Product Subcategory	Total Predicted Exposure (mg/kg/d) - day of use			
		< 0.1 Pa	0.1 - <1 Pa	1 - < 10 Pa	> = 10 Pa
PC1:Adhesives, sealants	Glues, hobby use	1.8	1.8	1.8	5.4
	Glues DIY-use (carpet glue, tile glue, wood parquet glue)	28	28	88	6727
	Glue from spray	105	105	105	105
	Sealants	1.9	1.9	3.4	159
PC3:Air care products	Aircare, instant action (aerosol sprays)	5.0	5.0	5.0	5.0
	Aircare, continuous action (solid & liquid)	0.1	0.1	0.1	7.9
PC9a: Coatings, paints, thinners, removers	Waterborne latex wall paint	38	38	56	2067
	Solvent rich, high solid, water borne paint	36	36	43	740
	Aerosol spray can	47	47	47	47
	Removers (paint-, glue-, wall paper-, sealant-remover)	131	131	153	2548

[62] ECHA. (2013). http://chesar.echa.europa.eu/documents/2326902/2424432/Chesar+in+a+nutshell

PC9b: Fillers, putties, plasters, modeling clay	Fillers and putty	7	7	19	1350
	Plasters and floor equalizers	169	169	403	26106
	Modeling clay	35	35	35	35
PC9c: Finger paints	Finger paints	195	195	195	195
PC12:Fertilizers	Lawn and garden preparations	86	86	86	86
PC13:Fuels	Liquids	75	75	105	3431
PC24: Lubricants, greases, and release products	Liquids	75	75	105	3431
	Pastes	29	29	29	29
	Sprays	237	237	237	237
PC31:Polishes and wax blends	Polishes, wax / cream (floor, furniture, shoes)	72	72	75	441
	Polishes, spray (furniture, shoes)	162	162	162	162
PC35:Washing and cleaning products (including solvent based products)	Laundry and dish washing products	86	86	86	107
	Cleaners, liquids (all purpose cleaners, sanitary products, floor cleaners, glass cleaners, carpet cleaners, metal cleaners)	71	71	72	111
	Cleaners, trigger sprays (all purpose cleaners, sanitary products, glass cleaners)	38	38	38	38
AC5:Fabrics, textiles and apparel	Clothing (all kind of materials), towel	1031	1031	1034	1369
	Bedding, mattress	28	28	63	3963
	Toys (cuddly toy)	57	57	57	57
	Car seat, chair, flooring	148	148	171	2667
AC6: Leather articles	Purse, wallet, covering steering wheel (car)	0.7	0.7	1.0	27.6
	Footwear (shoes, boots)	3.6	3.6	4.7	129.5
	Furniture (sofa)	16	16	28	1359
AC8: Paper articles	Diapers	56	56	56	56
	Sanitary towels	7.1	7.1	7.1	7.1
	Tissues, paper towels, wet tissues, toilet paper	29	29	29	29
	Printed paper (papers, magazines, books)	4.2	4.2	8.4	476
AC10: Rubber articles	Rubber handles, tires	6.1	6.1	54.5	5377
	Flooring	6.0	6.0	28.7	2525
	Footwear (shoes, boots)	3.6	3.6	4.7	130
	Rubber toys	2.3	2.3	2.3	2.3

AC11: Wood articles	Furniture (chair)	14.8	14.8	16.3	189
	Walls and flooring (also applicable to non-wood materials)	5.9	5.9	27.2	2367
	Small toys (car, train)	2.3	2.3	2.3	2.3
	Toys, outdoor equipment	6.6	6.6	6.6	6.6
AC13: Plastic articles	Plastic, larger articles (plastic chair, PVC-flooring, lawn mower, PC)	68	68	117	5483
	Toys (doll, car, animals, teething rings)	24.4	24.4	24.4	24.4
	Plastic, small articles (ball pen, mobile phone)	1.0	1.0	1.5	51.8

XV. SPECIFIC COMMENTS ON SECTION 3.3.5: VIRTUAL TISSUE (VT) MODELING

The brief discussion in Section 3.3.5 seems to serve more as a placeholder, and appropriately notes that this area of research is in its early stages of development. References to seminal ongoing research are lacking, such as the Wyss Institute's (supported by DARPA) activities focused on building 10 different human organs-on-chips and to link these to mimic whole body physiology.[63]

XVI. SPECIFIC COMMENTS ON SECTION 3.3.6. EXAMPLE: THYROID PATHWAY DISRUPTING CHEMICALS AND HIGH-THROUGHPUT SYSTEMS

- Section 3.3.6 starts to allude to a validation framework for HT assays with reference to an example for thyroid pathway disrupting chemicals. It is encouraging to see statements about development of criteria to facilitate the translation of data across different types of testing. The issues identified mirror, to an extent, those discussed in Patlewicz et al. (2013) as follows:
    - Assay Identification and Refinement discusses what assays are available from the ToxCast suite and where certain data gaps are for research purposes. However there is no discussion of how these assays will be evaluated for their performance – reliability, sensitivity, specificity, etc.
    - Algorithm Development for Toxicity and Hazard Prediction essentially discusses how prediction models could be developed from the assays combined together. We agree that there is uncertainty in relation to the assay results themselves. The section raises questions for consideration rather than any practical recommendations. Key considerations that are missing in this section include the following: defining the algorithm for each prediction model; ensuring

[63] http://wyss.harvard.edu/viewpressrelease/91/wyss-institute-to-receive-up-to-37-million-from-darpa-to-integrate-multiple-organonchip-systems-to-mimic-the-whole-human-body

transparency and characterizing the model in sufficient detail to facilitate review; reconstruction and independent verification of results; a reviewer should not have to reverse engineer any step or process in the prediction model. All data sets – both training sets and test sets – should be available to promote independent verification, thus increasing confidence in the model. This will also facilitate model improvements by other investigators who can refine and extend the algorithm and extend the domain of applicability with data on additional substances. reporting appropriate measures of goodness-of-fit, robustness and predictivity of the prediction model; summarizing known limitations of each prediction model; and the need for contextual and weight-of-evidence analysis on (quantitative) use of the prediction model for each specific purpose. In addition, all data sets – both training sets and test sets – should be available to promote independent verification, thus increasing confidence in the model. This will also facilitate model improvements by other investigators who can refine and extend the algorithm and extend the domain of applicability with data on additional substances.

- Assay Conduct, Data Analysis and Data Reporting for Risk Assessment Needs notes that understanding the characteristics of the individual assays is critical. Some of the language in this section could be better accommodated under Assay Identification and Refinement. Characterizing the information requirements for a given assay and how it is anchored within a given AOP(s) should be part of an analytical validation of the assay itself. Efforts are underway within OECD in terms of outlining what the minimal information requirements for non-guideline assays (including HT/HC) should be. Missing from this section are considerations such as defining a chemical domain of applicability, as well as transparency in the training and test (reference) data sets.

The statement on page 73 that "Some advantages of the ToxCast data sets are the (1) availability of dose-response information for all assays, (2) availability of assay method details, and (3) availability of the source code for all computational models used in the data analyses" is not accurate. Most of the ToxCast assays are proprietary, and only summary descriptions of the methods are available. Furthermore, the source code and algorithms for EPA's ToxCast computational models are not publically available. Even though the ToxCast prediction models and results may be published in the scientific literature, sufficient details are not provided to readily enable independent replication. Instead, extensive reverse-engineering is needed to deduce the algorithms and replicate calculations.

XVII. SPECIFIC COMMENTS ON SECTION 4.1. HUMAN VARIABILITY AND SECTION 4.1.1- GENOMIC VARIABILITY

While informative, the discussion on the use of new *in vitro* high-throughput methods to estimate human variability and susceptibility for risk assessment purposes is overly simplistic and optimistic. While genetic variability can affect human susceptibility, focusing the analysis of genetic variants (such as SNPs) in particular genes provides only a partial picture of human variability, because of modifier genes (and variants in them) that are not well understood. The Draft NexGen Report cites new methods and data such as the 1000 Genomes project in which

human cells from various individuals can be evaluated using *in vitro* high-throughput models to understand responses across subsets of the human population. Yet, these human cells are immortalized lymphoblastoid cell lines that have been passaged many times, thereby greatly reducing the representativeness of a chemical response *in vitro* for predicting effects *in vivo*.

The discussion of approaches to address human variability in risk assessment is incomplete. In particular, publications which have investigated the adequacy of adjustment factors to account for human variability in toxicokinetics and toxicodynamics should be cited and discussed. For example, the analysis of human metabolism and excretion of pharmaceuticals across a wide variety of pathways led to the conclusion that monomorphic pathway-related adjustment factors that cover the 99th percentile of the healthy adult population phase I metabolism (CYP1A2, CYP2A6, CYP2E1, CYP3A4, ADH, and hydrolysis), phase II metabolism (glucuronidation, glycine, and sulfate conjugation) and renal excretion were all below the default toxicokinetic adjustment factor of 3.16. While polymorphisms can lead to greater variability, it is important to place this into the perspective of the overall population.[64] While the statement "Variability is not explicitly accounted for in cancer…" (page 73 of the Draft NexGen Report) may be technically correct, in a narrow sense, it creates the misperception that EPA's standard cancer risk assessment methods are not health protective. The highly conservative nature of the typical linear low dose extrapolation process results in upper bound estimates. Cancer potency factors calculated by the U.S. EPA are presented as the 95% upper confidence limit on the dose-response curve, not the maximum likelihood estimate. This combination of upper bound estimation and linear extrapolation does not result in derivation of true or actuarial risks. EPA's cancer risk estimates inherently account for human variability, in that true risks are not likely to exceed such estimates, are likely to be less, and may even be zero.[65] EPA's cancer risk guidelines (page A-9) indicates that the Agency's linear default procedure adequately accounts for human variation "unless there is case-specific information for a given agent or mode of action that indicates a particularly susceptible subpopulation or lifestage, in which case the special information will be used."[66]

Additional specific comments on this section are provided below:

- While Figure 22 illustrates how susceptibility may arise from variability, this theoretical graphic provides little or no insight into ways to improve risk assessment methods. For example, in this section it seems odd that probabilistic methods were not discussed, particularly in light of the data which show that the default uncertainty factor for human variability in toxicokinetics is protective for 99% of the health adult human population.

  It is unclear why the assertion on page 75 of the Draft NexGen Report that "The proportion of total variation explained by individual genome-wide-significant variants has reached 10%–20% for a number of diseases" cites a publication, Visscher et al. 2013, which deals with anthracycline-induced cardiotoxicity in children. It appears that the reference should instead be Am J Hum Genet. 2012 January 13; 90(1): 7–24.

[64] Dorne and Renwick. 2005. The Refinement of Uncertainty/Safety Factors in Risk Assessment by the Incorporation of Data on Toxicokinetic Variability in Humans. Toxicol Sci.86: 20–26.
[65] http://www.epa.gov/ttnatw01/urban/appx1011.pdf.
[66] EPA. (2005). Guidelines for Carcinogen Risk Assessment http://www.epa.gov/raf/publications/pdfs/CANCER_GUIDELINES_FINAL_3-25-05.PDF.

- The discussion of understanding human variability in pharmacokinetics and dynamics is important, but discussion of environmentally relevant exposures must also be included. As is clearly shown in the simulations in Figure 23 (page 75), the differences in internal dose and biological response are largely ameliorated at low, environmentally relevant dose levels. Thus, the relevance and impact of human variability (both dynamics and kinetics) in terms of risk assessment is highly dependent on the degree of exposure. Genetic polymorphisms in ethanol metabolism are well documented, and while such genetics impact individual susceptibilities to effects of ethanol ingestion, there is no evidence at all that such polymorphisms have any impact whatsoever following environmentally relevant levels of ethanol ingestion from consumption of fruits, breads and other products which contain low levels of naturally occurring ethanol.

- Overall, this section lacks a "fit for purpose" discussion of how emerging approaches to address the range of variability in subpopulations will be considered for use in refining variability parameters in different types of risk assessments.

XVIII. SPECIFIC COMMENTS ON SECTION 4.1.2. EARLY-LIFE EXPOSURES

Overall, this section presents a cursory and unbalanced discussion of early life exposures and susceptibilities. Children may be less, more or equally susceptible compared to adults. For example, Charnley and Putzrah[67] document that "Experimental evidence clearly shows that young animals are not always more sensitive than older animals to chemically induced carcinogenesis (1) (sic). Because the biological modes of action involved in carcinogenesis are multiple and varied, many factors may account for differences in sensitivity. Rates of metabolism and clearance in the young are generally faster than those in adults, which may increase or decrease toxicity. For example, many cytostatic agents must be administered to children at up to five times greater doses on a per body weight basis than the dose to adults. Although it is true that developing organisms may be of special sensitivity simply because they are developing, whether environmental contaminants are having a disproportionate impact on the young is a matter primarily of conjecture (with some obvious exceptions, such as lead)."

For toxicokinetics, Dorne and Renwick (2005)[68] showed that pathway-related factors for the 99th percentile of children were less than the 3.16 default toxicokinetic factor except for the polymorphic CYP2D6 and CYP2C19 isoforms, suggesting a similar degree of susceptibility for children and adults except for toxicants metabolized to inactive moieties by CYP2D6 and CYP2C19. While neonates were shown to have the lowest enzyme activity of the whole human population in both in phase I and phase II metabolism, Dorne and Renwick conclude, "The low activity could translate into a greater susceptibility if the parent compound were the active toxicant, but less susceptibility if the compound underwent metabolic activation."

[67] Charnley G, Putzrah RM. 2001. Children's health, susceptibility, and regulatory approaches to reducing risks from chemical carcinogens. Environ Health Persp 109:187–192.
[68] Op. cit. footnote 64.

The discussion of epigenetics is also very cursory and also fails to discuss and cite important literature of direct relevance to considering epigenetics in risk assessment such as Goodman et al. (2010)[69] and Alyea et al. (2013).[70]

It is particularly concerning that the outcomes from the recent workshop on epigenetics and safety assessment (Goodman et al. 2010), which focused specifically on the four questions below, were not discussed in the Draft NexGen Report.

1. What model systems might be employed to evaluate the ability of a chemical to produce an epigenetic change (affecting the F1 and/or F3 generation);
2. What endpoints/targets might be evaluated;
3. What techniques might be employed; and
4. Regulatory perspective: when is it appropriate to incorporate "new" science, in this case epigenetics, into the regulatory process? What does one need to know, what are the pitfalls and how might these be overcome/avoided?

This represents an opportunity lost, as the Draft NexGen Report could have built upon the workshop recommendations to describe the Agency's perspectives for a path forward.

The characterization of the publication by Boekelheide et al. 2012 in the Draft NexGen Report as "A good example….that associated early-life exposure to arsenic and DNA hypomethylation with the development of arsenic-induced skin lesions" should note the following: 1) this reference is to a descriptive review paper, not a quantitative (meta) analysis; 2) there do not appear to be direct studies of DNA hypomethylation in early life exposed individuals with respect to the development of arsenic-induced skin lesions; and 3) the Boekelheide et al. 2012 publication refers to paper that reports adult exposures and DNA hypomethylation, not early life exposures. The Boekelheide et al. 2012 paper states, "In arsenic-exposed adults, DNA hypomethylation has been associated with the subsequent risk of developing arsenic-induced skin lesions (Pilsner et al. 2009)."

XIX. SPECIFIC COMMENTS ON SECTION 4.1.3. MIXTURES AND NONCHEMICAL STRESSORS

In ACC's view, cumulative risk assessments should be performed on a case by case basis, when available knowledge and data will allow for an informed scientifically based assessment. It is important to recognize that it is not necessary to conduct a cumulative assessment on all chemicals. Specifically, for chemicals to be considered for inclusion in a cumulative assessment, scientific evidence should first indicate a likelihood of both a co-exposure and some level of interaction of stressors. Furthermore, most, if not all biological processes, respond to or manage a plethora of mixtures of naturally occurring and endogenously formed substances, with these same biological processes responding to, or managing, man-made chemicals. Therefore, approaches to evaluating mixtures and nonchemical stressor section must account for the presence of naturally occurring and endogenous substances and mixtures and homeostatic

[69] Goodman et al. (2010). What do we need to know prior to thinking about incorporating an epigenetic evaluation into safety assessments? Toxicol Sci. 116(2):375-81.

[70] Gollapudi et al. (2013). Are we ready to consider transgenerational epigenetic effects in human health risk assessment? Environ Mol Mutagen. Nov 21. doi: 10.1002/em.21831. [Epub ahead of print]

mechanisms. Unfortunately, this section provides only a cursory discussion of cumulative risk, and fails to address this. Moreover, this section doesn't cite or discuss a number of publications which provide practical approaches to addressing simultaneous exposures to multiple agents. Such publications include:

- Han and Price. 2011. Determining the Maximum Cumulative Ratios for Mixtures Observed in Ground Water Wells Used as Drinking Water Supplies in the United States. Int. J. Environ. Res. Public Health 8: 4729-4745. http://www.mdpi.com/1660-4601/8/12/4729/pdf .

- ECETOC. 2011. Technical Report 115, Effects of Chemical Co-exposures at Doses Relevant for Human Safety Assessment.

- ECETOC. 2011. Workshop Report No. 22: Workshop on combined exposures to chemicals.

- Han and Price. 2012. Applying the maximum cumulative ratio methodology to biomonitoring data on dioxin-like compounds in the general public and two occupationally exposed populations. J Expo Sci Environ Epidemiol 23: 343-349 http://www.nature.com/jes/journal/vaop/ncurrent/full/jes201274a.html

- Meek ME et al. 2011. Risk assessment of combined exposure to multiple chemicals: A WHO/IPCS framework. Regulatory Toxicology and Pharmacology 60: S1- S-14. http://www.sciencedirect.com/science/article/pii/S0273230011000638.

- OECD. 2011. WHO OECD ILSI/HESI International Workshop on Risk Assessment of Combined Exposures to Multiple Chemicals, Workshop Report. http://www.oecd.org/officialdocuments/displaydocumentpdf/?cote=env/jm/mono%282011%2910&doclanguage=en

- Price PS and Chaisson CF. 2005. A Conceptual Framework for Modeling Aggregate and Cumulative Exposures to Chemicals. Journal of Exposure Analysis and Environmental Epidemiology 15: 473–48.

- Price et al 2012a. A decision tree for assessing effects from exposures to multiple substances. Environmental Sciences Europe 2012, 24:26. http://www.enveurope.com/content/24/1/26.

- Price et al. 2012b. An application of a decision tree for assessing effects from exposures to multiple substances to the assessment of human and ecological effects from combined exposures to chemicals observed in surface waters and waste water effluents. Environmental Sciences Europe 2012, 24:34. http://www.enveurope.com/content/24/1/34.

- Price and Han. 2011. Maximum Cumulative Ratio (MCR) as a Tool for Assessing the Value of Performing a Cumulative Risk Assessment. Int. J. Environ. Res. Public Health 8: 2212-2225. http://www.mdpi.com/1660-4601/8/6/2212/pdf.

EPA should have drawn upon these publications to describe how the Agency intends to proceed to incorporate 21st century methods into these tiered approaches. Clearly, the evolutionary process for these new tools is in its very early stages and must proceed with scientific rigor. In this regard, concerns will arise if cumulative assessments are based on multiple assumptions which combine conservative default assumptions, since such an approach would likely result in a significant over-estimation of risk. Thus, it is critical that sufficient data and robust models be available to inform a cumulative risk assessment.

Moreover, before cumulative risk can be assessed, the individual risks from both chemical and non-chemical stressors need to be quantified in terms of a "common metric." The common metric needs to be such that the basis for each stressor is comparable (e.g., central tendency estimates of risks) so that the "addition" of risks from each stressor is scientifically supportable. For example, a cumulative risk assessment would be skewed if it added the upper bound estimate of a chemical risk to a central tendency estimate of risk of a non-chemical stressor.

XX. SPECIFIC COMMENTS ON SECTION 4.2. INTER-SPECIES EXTRAPOLATION

It is unclear why this section fails to include and discuss PBPK modeling and internal dosimetry for improving inter-species extrapolation. The interspecies adjustment factor used in risk assessment includes both kinetic and dynamic components. Using methods which focus on internal dosimetry (e.g., blood concentrations) rather than administered dose, the toxicokinetic component of the interspecies uncertainty factor can be eliminated. It has been shown that when a parent compound in blood is the directly relevant dose measure, both the inter- and intra-species toxicokinetic components are replaced by direct measurement of blood concentration and the composite target MOE could be reduced to 10 to achieve an equivalently protective MOE.[71] Thus, improving methods for conducting internal dosimetry is an important component of 21st century risk assessment.

The Draft NexGen Report alludes to AOPs, in stating "Regulatory toxicology as a whole will move toward increasing reliance on predictive approaches to assessing chemical risk, with a greater emphasis placed on understanding chemical perturbation(s) of conserved biological pathways at key junctures, including molecular initiating events (MIEs) (e.g., activation or inactivation of specific receptors, enzymes, or transport proteins)." As noted in our General Comments and in our comments on Section 3.3: Tier 1 Screening and Prioritization, the authors could cite the work under the OECD AOP Work Programme .[72]

The suggestion that extrapolation between species can occur at different levels of biological organization within an AOP does not appear to be well described. It's implied that hazard values can be readily extrapolated across species based on initial or key events within a pathway. Extrapolation in this sense is a prediction model, and as such, there needs to be scientific confidence in both the assays involved and the prediction model before employing such approaches in hazard characterization or risk assessment. In this regard, EPA should

[71] Aylward et al.(2011). Assessment of margin of exposure based on biomarkers in blood: an exploratory analysis. Regul Toxicol Pharmacol. 61(1):44-52.
[72] OECD AOP Work Programme http://www.oecd.org/env/ehs/testing/adverse-outcome-pathways-molecular-screening-and-toxicogenomics.htm

articulate the framework the Agency will employ, such as that of Patlewicz et al., 2013 or a similar method, for evaluating and documenting the scientific confidence in HT/HC methods and prediction models derived therefrom.

XXI. SPECIFIC COMMENTS ON SECTION 4.3. LOW DOSE-RESPONSE MODELING

ACC agrees that robust biologically-based dose response (BBDR) models can improve the accuracy of risk predictions. It's encouraging that EPA intends to make better use of BBDR models in the future. In moving forward, when evaluating BBDR models, EPA should take note of lessons learned and the recommendations of the NRC (2011),[73] which admonished EPA for not using the formaldehyde BBDR model in the draft IRIS assessment, emphasizing that EPA's evaluation involved extreme manipulations, were not scientifically justified and should not have been used as the basis of rejection of the use of the BBDR model in its assessment.

Characterizing the expected response at low exposure levels (i.e., those that the public is most likely to encounter) is a significant challenge. Typically, EPA risk assessment methods seek to model or estimate high end or upper bound risks, not maximum likelihood or expected risks. Clearly, improvements focused on developing realistic, expected risks are an important goal. But this section lacks detail on how EPA will go about achieving this. Similarly, it is not clear what EPA is proposing with respect to the automated dose-response modeling approach proposed by Burgoon and Zacharewski (2008) depicted in Figure 25.

The discussion of "virtual models" on page 82 is also informative, but again, it is not clear what actions EPA intends to take in terms of conducting additional developmental research, initiating translational research to standardize these methods or testing and analysis to establish scientific confidence in these methods for regulatory acceptance.

XXII. SPECIFIC COMMENTS ON SECTION 5. LESSONS LEARNED FROM DEVELOPING THE PROTOTYPES

- Page 84, line 13, begins by stating: "Thousands of chemicals to which humans are exposed have inadequate data for predicting their potential for toxicological effects." For a report that is overly confident in so many areas, it is surprising to see EPA take such a negative approach to the status of chemical testing. A more positive approach would begin by noting that if the tools mentioned in the report are appropriately developed and validated, thousands of chemicals can be successfully screened to understand where there may be potential risks.

- Page 84, line 21, provides a very appropriate caveat, citing Tice et al. 2013, to note that while progress is being made, "both reliability and relevance of the approach need to be demonstrated". EPA should take this to heart and ensure that the revised prototypes benefit from a more complete evaluation of the methodologies they rely upon.

- Page 85, line 6, refers to Tier 3 as those assessments where risks are generally considered "known." While Tier 3 is data rich, there is not full agreement that human risks are known and that comparisons to traditional assessments are always going to provide an appropriate

[73] NAS. (2011). Review of the Environmental Protection Agency's Draft IRIS Assessment of Formaldehyde. http://www.nap.edu/catalog.php?record_id=13142

baseline. EPA must ensure that animal data is not inappropriately treated as a gold standard because there are well documented species differences and dose-response factors which limit extrapolation to humans. In addition, traditional assessments, such as IRIS, are designed to be purposefully conservative so as to not underestimate risks.

- Page 86, line 32, appropriately talks of need for systematic review criteria for study selection. EPA should ensure that this includes clear criteria related to study quality, study reliability, and study relevance.[74]

- Page 86, line 36, cites Krauth et al. 2013. EPA should better clarify why this citation is appropriate and relevant in referring to using weight-of-evidence criteria and noting that "data from multiple, similar studies are preferred". Krauth et al. 2013 does not make this finding.

- Page 89, line 40, discusses variability and susceptibility. Please see our comments in section XVII regarding our concerns with the tools and approaches EPA has identified.

- Page 91, Table 10, provides the various approaches that could provide data to support each type of decision context. For all tools, EPA must first ensure that they have sufficient scientific confidence for the intended use. As noted previously, many of the NextGen methods described in the Draft NextGen Report are not yet ready for use by the Agency.

- Page 91, line 6, notes that Tier 3 "could be" augmented with new tools if of sufficient quality. If EPA takes the steps to appropriately test and validate new screening tools, once they are deemed reliable, EPA should indeed rely on this information in Tier 3 assessments. IRIS assessments can certainly be improved by relying on other streams of evidence if they are sufficiently reliable.

XXIII. SPECIFIC COMMENTS ON SECTION 6. CONCLUSIONS

In discussing the challenges that remain, while the four key challenges identified by EPA are indeed challenges, EPA must not forget that an overarching framework to document the scientific confidence that exists for each of the new methodologies is necessary and still remains a key challenge for the Agency. Section 6 later (at page 94) mentions the need to validate high-throughput assays, including the need for the development of a framework for validation. ACC supports this approach and encourages EPA to engage stakeholders early in the process of the development of decision considerations, criteria for systematic review, and weight of evidence approaches.

In regards to next steps, EPA notes that the Agency will develop toxicity values in each tier. It is not clear how EPA plans to do this for Tier 1 screening and problem formulation assessments. Further elaboration by EPA of the planned approaches would be useful. EPA should also ensure appropriate opportunities for stakeholder and peer review comment on these approaches.

[74] Bevan and Strother. (2012). Best Practices for Evaluating Method Validity, Data Quality and Study Reliability of Toxicity Studies for Chemical Hazard and Risk Assessments. http://arasp.americanchemistry.com/Data-Quality-Evaluation.

XXIV. ACC'S OVERARCHING RECOMMENDATIONS

We support continued investment of research resources to improve development and application of the new technologies to enable faster, less expensive and less laboratory animal-intensive hazard, exposure and risk estimations, and to more accurately determine the probability of adverse health outcomes at environmentally relevant exposure levels. To assure that these 21st century technologies and approaches are scientifically robust, and optimized as fit for purpose for regulatory use in EPA's OCSPP and other regulatory offices, ACC makes the following recommendations:

1) EPA should develop an integrated plan, led by the OSCPP, in coordination with National Center for Computational Toxicology, which:

    a. Groups the technologies and approaches into three categories: 1) ready now, 2) undergoing scientific evaluation for relevance, reliability and fitness for purpose, and 3) under development. Such a categorization will provide a clearer picture of the state of the science, the current confidence in use of data from these methods/approaches for specific purposes, and the path forward to improve scientific confidence in the methods.

    b. Defines the additional datasets and analyses needed to expand the tiered, HTS/transcriptomics margin of exposure method to achieve regulatory acceptance for the full range of chemical domains needed by EPA.

    c. Focuses additional effort on incorporating exposure into the next generation of risk assessment approaches, by building from the consumer model in the ECETOC Targeted Risk Assessment exposure tool and the tiered exposure assessment approaches developed by the ILSI-HESI RISK 21 project

    d. Develops the framework the Agency will use to establish and document the scientific confidence that is needed for these methods and prediction models to be used for regulatory purposes.

    e. Employs a more open stakeholder engagement process which improves upon the past practices used in generating the Draft NexGen Report by maximizing open meetings, broadening consultations and collaborations and conducting peer review in accordance with Agency procedures for influential risk assessment guidance and policies.

2) EPA should revise the benzene, ozone and benzo(a)pyrene prototypes, taking into consideration public comments, and then submit the revised drafts for independent peer review before using these methods in EPA's IRIS or other EPA programs.

3) EPA should build from the NexGen work and the OECD AOP activities to develop draft guidance which addresses the development, evaluation and use of AOPs for defined purposes such as prioritization, formation of categories[75] for read across, integrated testing, and screening level hazard/risk assessment.

[75] EPA. 1999. Development of Chemical Categories in the HPV Challenge Program http://www.epa.gov/hpv/pubs/general/categuid.htm. EPA. Supplemental Guidance for Final Category Analysis http://www.epa.gov/hpv/pubs/general/hpvsuplm.pdf. OECD (2014). Grouping of Chemicals: Chemical Categories and Read-Across http://www.oecd.org/env/ehs/risk-assessment/groupingofchemicalschemicalcategoriesandread-across.htm.