National Toxicology Program Proposed Approach to Genomic Dose-Response Modeling

Introduction

Changes in the approach to toxicological assessment [1] and the advent of inexpensive, high-throughput transcriptomics data generation platforms have led to significant interest in the integration of genomic dose-response studies into the hazard characterization and risk assessment process [2]. Currently, many questions exist regarding how best to design, perform, and interpret genomic dose-response studies in a manner that most effectively facilitates integration of these types of assessments into the hazard characterization paradigm. Consensus or, at a minimum, guidance on how to carry out and analyze these types of studies would be helpful in advancing the use of their findings in toxicology and risk assessment decision making. This document describes a proposed framework for performing genomic dose-response analysis that is largely consistent with published approaches. The described approach is targeted at developing screening-level hazard assessments of test articles that can be used for prioritization and the setting of interim exposure limits, particularly for in vivo studies.

NTP is convening an expert panel on October 23–25, 2017, at the National Institute of Environmental Health Sciences, Research Triangle Park, NC to obtain input on specific details of its proposed approach to genomic dose-response modeling. NTP will carefully consider the panel's input and recommendations in determining what changes to the approach might be needed prior to finalization. NTP will also continue to monitor the scientific literature with regard to the development of improved approaches to data modeling and analysis. Importantly, in reviewing the proposed approach, we ask readers to keep in mind that NTP's goal in pursuing genomic dose-response studies is to quickly and cost effectively develop sensitive, screening-level potency estimates for test articles and provide a degree of contextualization to facilitate qualitative interpretation of observed genomic changes.

A. Overview of NTP's Proposed Approach

Approach: NTP's proposed approach is in large part consistent with an approach to genomic dose-response modeling outlined by Thomas et al. [3] and used extensively by other researchers with various slight modifications [4-13]. In short, this approach entails (1) filtering the measured features (genes/probe sets) to remove those not responding to chemical treatment; (2) fitting each of the filtered features to multiple, parametric, dose-response models, identifying the best-fit model, and deriving a potency value [benchmark dose (BMD)] from that model for the feature; (3) parsing the features into predefined gene sets (e.g., Gene Ontology Biological Processes); and (4) determining potency for each of the adequately populated gene sets by deriving the mean and median potency of the genes in each set. A software package referred to as BMDExpress was developed that facilitates data analysis in accordance with the outlined approach [14]. NTP recently modified the software to produce the new version BMDExpress 2.0, which is freely available.[1]

Two additional issues, which are not immediately central to the data modeling pipeline but are critical to overall success of the genomic dose-response approach, are study design and biological interpretation of findings. With this in mind, NTP proposes to employ the study-design approach proposed by Slob et al. [15] in which biological samples are distributed over a broad dose range, which allows for more accurate estimates of model parameters. This design contrasts with a more traditional toxicology study design that includes a limited number of dose levels in combination with a high level of biological replication. The addition of other study-design parameters (detailed below), such as study duration, organ/tissue/cell line selection, etc., will be based on published findings and consistent with guideline studies, where appropriate. To facilitate interpretation of results from genomic dose-response analysis, NTP proposes to employ the Hallmark gene sets [16] supplemented with peer-reviewed signatures of toxicity and additional gene sets derived from global mining of co-regulated gene space. Further, NTP plans to formulate toxicological interpretations of the Hallmark gene sets through literature and data-driven curation. Interpretation will be in collaboration with the toxicology community to facilitate extrapolation of findings from genomic dose-response studies to toxicological effects used in regulatory decision making. Finally, NTP proposes that quantitative interpretation use the genomic dose-response results from the most sensitive gene set (e.g., Hallmark gene set) based on median BMD when identifying the lowest dose at which biological changes occur [i.e., biological effect point of departure (BEPOD)].

Justification for the approach: The primary justifications for the overall approach are historical precedent and peer review, as several research groups in government, academic, and private sectors have used the general approach and it has been peer reviewed. Further, the proposed approach to dose-response modeling is largely consistent with standards used by US EPA to evaluate data from guideline toxicological assessments and make regulatory decisions. US EPA methods have been extensively documented and reviewed. NTP's mission includes providing data for regulatory decision making; thus, we hope that employing methods already accepted by much of the regulatory community will remove one major barrier to the use of genomic data in a regulatory context. Finally, results derived from genomic dose-response studies of multiple chemicals, which used the proposed approach, yielded estimated potency values that are similar to potency estimates derived from apical toxicological endpoints [17]. This finding is critical to the proposed approach, as it relates to the translation of findings from genomic dose-response studies to more traditional hazard characterization [18].

NTP's approach to study design focuses on obtaining the best data to determine accurate estimates of biological potency using the proposed modeling approach. Use of an extended version of the Hallmark gene sets is based on several factors, described in more detail below

[1] https://github.com/auerbachs/BMDExpress-2.0/releases.

including congruency with gene expression data, generalizability, and reduction in redundancy. Finally, reporting of the most sensitive gene-set BMD as the BEPOD is based on observations that the BMD for the most sensitive gene set from short-term, in vivo, genomic dose-response studies closely approximates the most sensitive BMD values from guideline toxicological assessments of the same test article.

Despite the relatively broad use of the methods in the proposed NTP analysis pipeline, questions remain about specific aspects of the analysis performed inconsistently across the published literature. We hope that consensus can be reached on the best practices for implementing the proposed analysis approach to genomic dose-response modeling. The following is a detailed description of the steps in the analysis.

B. A Detailed Description of the Steps in the Genomic Dose-Response Analysis

1. Filtering of Measured Features

Approach: In the first step of the genomic dose-response analysis protocol, we propose to apply a statistical (one-way ANOVA, $p < 0.05$, no multiple testing correction) and fold-change (threshold to be empirically determined for each technology) filter to each data set (one data set per chemical) to remove measured features (i.e., genes/transcripts/probe sets) from subsequent analysis that do not demonstrate a response to test article treatment.

Justification for approach. The basis for selecting this approach is the observation from MicroArray Quality Control studies that a nominal p-value combined with the fold-change filter yields the highest cross-laboratory and cross-platform reproducibility when evaluating transcriptomic data [19]. A multiple testing correction was not included, in part, because of our finding that it lowered the overall reproducibility of the results as compared to the proposed filtering process. A trend test-based filter (e.g., Williams Test) has been considered to select features with a monotonic trend; however, due to the complexity of the genomic response, biologically meaningful, non-monotonic responses can be observed that would be removed by a simple trend test.

2. Fitting Features to Dose-Response Models

Approach: In the second step of the genomic dose-response analysis, we propose to fit dose-response curves to each measured feature that exhibits a response to chemical treatment as determined by the filtering approach described above. Dose-response modeling will be performed as described previously, with minor modifications [3, 14]. To model the data, Hill, power, linear, polynomial 2°, polynomial 3°, exponential 2, exponential 3, exponential 4, and exponential 5 dose-response models will be fitted to the measured features. Equations describing each model are given below. The BMDExpress 2.0 software includes the US EPA model executables used in the BMDS software.[2] The specific details on each model are described in the BMDS software guidance document.[3]

Model equations

Equations for the models are taken from the BMDS software guidance document .[4] In all equations, μ is the mean response predicted by the model.

Polynomial model

$$\mu(dose) = \beta_0 + \beta_1\ dose + \ \beta_2\ dose^2 + \ \cdots\ + \ \beta_n\ dose^n$$

where *n* is the degree of the polynomial.

Linear model

The linear model is a special case of the polynomial model with *n* fixed at 1.

Power model

$$\mu(dose) = \gamma + \ \beta\ dose^{\delta}$$

where $0 < \gamma < 1$, $\beta \geq 0$, and $18 \geq \delta > 0$.

Hill model

$$\mu(dose) = \gamma + \frac{v\ dose^n}{k^n + \ dose^n}$$

Exponential model*

The four exponential models are:

$$\mu(dose) = a * \exp(sign * \ b * dose)$$

$$\mu(dose) = a * \exp(sign * \ (b * dose)^d)$$

$$\mu(dose) = a * (c - (c - 1) * \exp(-1 * \ b * dose))$$

$$\mu(dose) = a * (c - (c - 1) * \exp(-1 * \ (b * dose)^d))$$

*For the first two exponential models, 'sign' is the adverse direction.

The data will be log2 adjusted and presumed to have a normal distribution, and each model will be run assuming constant variance. In the future, a test for dose-related variance will be

[2] https://www.epa.gov/bmds/download-benchmark-dose-software-bmds-model-executables and https://www.epa.gov/bmds.

[3] https://www.epa.gov/sites/production/files/2015-01/documents/benchmark_dose_guidance.pdf.

[4] *ibid.*

implemented for each feature, and non-constant variance will be used in the modeling when appropriate.

The potency value derived from each model is the benchmark dose [20].

- Benchmark dose is defined as the estimated dose or concentration that produces a predetermined change in the response rate of a biological response (called the benchmark response or BMR) compared to background [20].
- The BMR for each feature will be set to 1.349 multiplied by the estimated standard deviation (SD) at zero dose. SD is estimated using the entire fitted curve, not just the control data. Assuming a normal distribution, a BMR of 1.349 × SD is the amount required to shift the mean transcriptional response of the control group distribution (as estimated by the model) such that the treatment group's distribution contains 11% in a single tail, that is, a 10 % increase or decrease in transcript abundance compared to control.
- An adverse direction (i.e., the direction of response indicating up- or down-regulation) is not selected *a priori*, but determined by the software based on the shape of the dose-response curve. The selection of an adverse direction is based on which direction gives the best fit to the specific model. In the case of polynomial 2° and polynomial 3° models, the models are run twice, once in each adverse direction, and the model with the lowest BMD is selected.

To identify the best-fit model for each feature, a two-step selection process is followed. First, a nested likelihood ratio test is performed on the linear, polynomial models. If the more complex model provides a significantly improved fit ($p < 0.05$), the more complex model will be selected. If the more complex model does not provide a significantly improved fit ($p \geq 0.05$), the simpler model will be selected [21]. Second, the Akaike information criterion (AIC) for the selected polynomial model is compared with the AIC for the power, Hill, and exponential models. The model with the lowest AIC [22] will be selected as the final model and used to calculate the probe set BMD, BMD lower confidence limit (BMD_L), and BMD upper bound (BMD_U) [23]. However, if the Hill model is identified as the best-fit model, but the "k" parameter is <1/3 of the lowest positive dose, the next best model with a goodness-of-fit p-value >0.05 will be selected. In the case where no model has a p-value >0.05, the feature will be assigned a BMD value equal to the lowest BMD from the probe set with a usable Hill model. This assigned value for the feature will then be used in the subsequent gene-set analysis described below. In a limited number of cases, a model will not be identified because of failed parameter convergence (i.e., BMD, BMD_L, or BMD_U). In such cases, the feature will not be considered in the gene-set analysis described below.

Justification for approach: The choice of using the parametric models specified in the US EPA BMDS software is based, in part, on validation of the model algorithms for using in regulatory data modeling and the greater simplicity of the parametric models. With the scale of data that

NTP will be generating, parametric models provide a considerable gain in computational efficiency as compared to alternative approaches [24, 25].

All continuous models currently available from US EPA are proposed for use in the modeling, in part, because a prior hypothesis for the behavior of any given transcript is unknown due to the complexity of the biological response. Further, use of multiple models maximizes the likelihood that a feature with dose-related response will fit the model well enough to be considered further in gene-set analysis.

NTP selected BMD as the potency metric because it is consistent with common practice in regulatory toxicology. The BMD potency metric is used in regulatory toxicology because it provides a model-based determination of the minimum dose level expected to have a significant biological effect. The BMD often occurs between the no observed effect level and the lowest observed effect level, both of which are determined by more traditional, pairwise statistical analysis.

A BMR of 1.349 × SD of the control samples is used because it approximates a 10 % increase in the number of extreme responses of treated groups relative to the response of controls (i.e., in the tails of the control distribution). This is consistent with the common practice of using a BMD_{10} (i.e., a BMR of 10% change in incidence). An NTP goal for performing genomic dose-response analysis is to estimate a biological effect level that aligns with the toxicological effect levels observed in guideline studies where using the 10% response threshold for genomic analysis has been deemed appropriate.

Prior to data analysis and after data normalization, the counts (in the case of RNA-seq) or intensities (in the case of microarray) from the measured features are log transformed. Log transformation of data is generally agreed to produce data that approximate a normal distribution with predominantly constant variance [26]; therefore, the data are assumed to exhibit constant variance.

For model selection, nested chi square is used to select the best poly model followed by AIC to compare the best poly model with other model types. When possible, NTP prefers to use a statistical test (i.e., nest chi square test) for making comparison among the nested models (e.g., poly models). However, when this is not possible, such as with non-nested models, the AIC metric is employed.

3. Determining Gene-Set Level Potencies

Approach: After selecting the best model, the measured features and their associated $BMD/BMD_L/BMD_U$ values are parsed into predefined gene sets (the specific gene sets are discussed below). For features to be passed into the gene-set analysis, the best-fit model must: (1) demonstrate convergent BMD, BMD_L and BMD_U values; (2) have a BMD less than the highest positive dose used in the study; (3) not map to more than one gene; (4) have a model fit p-value >0.0001; and (5) have a BMD_U to BMD_L ratio <40. All features passing these selection criteria are converted to their corresponding NCBI Entrez Gene ID and then parsed into pre-defined gene sets. Gene sets that contain at least three genes, are at least 5% populated (based on total annotated gene number), and are significantly enriched (Fisher's exact test, $p<0.05$) with genes from the study are declared "active," and BMD, BMD_L and BMD_U are determined by calculating the mean and median BMD, BMD_L, and BMD_U for each "active" gene set.

Justification: Features that do not exhibit convergent BMDs or exhibit BMDs greater than the highest dose are removed because in both cases there is significant uncertainty in the model and/or the derived potency values. Features that map to more than one gene are removed because which gene the signal represents is uncertain. A fit p-value >0.0001 is a more liberal threshold than US EPA's recommended threshold of 0.1 and, therefore, allows more features into the gene-set analysis. This threshold has been empirically evaluated using positive control genes with a dose-related response. A fit p-value threshold >0.0001 often excludes clear dose-related responses for those features representing the positive control genes (e.g., *CYP3A4* in response to rifampicin and *CYP2C9* in response to phenobarbital in studies using HepaRGs). Finally, a BMD_U to BMD_L threshold <40 removes features with high uncertainty in their fit to the model, which is often related to noisy data that should not be considered in the gene-set analysis.

The thresholds for calling a gene set "active" have been subject to significant debate as most notably indicated by all of the different thresholds used in the peer-reviewed literature [27]. Selection of the 3-gene, 5% populated, and Fisher exact test ($p < 0.05$) thresholds has been derived primarily through empirical assessment of which thresholds plausibly yield genomic BMD values that are comparable to BMD values derived from apical endpoints, such as organ weight change or pathology from longer-term, guideline, toxicological assessments.

The selection of mean and median BMD, BMD_L, and BMD_U values as representative gene-set potency values is based on estimating the central tendency of the gene BMDs in a gene set. This relatively simple approach to estimating gene-set level potency has been effective in limited empirical assessments. Specifically, NTP has observed mechanistically congruent, pathway-level potency estimates for certain drugs at known pharmacologically active dose levels (e.g., fatty acid metabolism gene set and fenofibric acid).

4. Addressing the Study Design

Approach: Two paradigms are related to study design in toxicology studies [28]. Paradigm 1 is the standard study design prescribed by the toxicology testing guidelines of the Organisation for Economic Co-operation and Development, for which the goal is to maximize the power for performing pairwise comparisons between dose groups to identify a no observed effect level. This design is highly dependent on prior knowledge of the biological potency, particularly if applying dose-response, model-based approaches to the data analysis, such as those proposed herein. In paradigm 2, the study design focuses on creating a data set on which dose-response modeling is performed on the entire data set. The goal of this type of analysis is to identify a model-based estimate of the minimal dose that produces an effect (i.e., a BMD). The number of biological replicates for paradigm 2 is less than for paradigm 1, and more dose levels are employed. NTP proposes to use paradigm 2 for in vivo and in vitro genomic dose-response studies.

In addition to the two paradigms described above, several, critical, study design parameters require discussion, including the species/sex of the model system, age of the animals at study start (in vivo), duration of exposure, organ (in vivo studies) or cell type (in vitro studies) selection for genomic analysis, and high-dose selection.

For in vivo studies, NTP proposes the following study parameters:

- Using 6- to 8-week-old male Sprague Dawley rats unless otherwise justified through literature review.
- Setting exposure duration to 5 days (i.e., 5 repeated doses, 1 per day, followed by euthanasia 24 hours after the last dose).
- Selecting organs/tissues for evaluation based on route of proposed exposure and review of the literature with a specific focus on target organs from guideline toxicity studies of the test article or structurally related test articles. In addition, liver will be evaluated in all studies as it is commonly affected by chemical challenge and often serves as a biosensor of systemic toxicological effects.
- Setting the highest dose for a test article based on determining the 5-day maximum tolerated dose (MTD; i.e., the highest dose that produces less than 20% decrease in body weight gain after 5 days of repeated dosing).

For in vitro studies, NTP proposes the following study parameters:

- Using human cell lines/types with sex as male or female depending on availability.
- Setting exposure duration based on expert review and results from range-finding studies.
- Selecting cell type based on several variables including culturability in 384-well format, patterns of use in the field (i.e., cell types that are more commonly used by the testing community will be given higher priority), and representation of common tissues of concern (i.e., organotypic representation of in vivo target tissues such as liver, e.g. HepaRG).
- Setting the highest dose for the test article to target the lethal concentration 20 (LC20; 20% reduction in cell viability relative to control). LC20 values will be determined by a range-finding study. Alternatively, some chemicals will not reach an LC20, and the high dose will be set at the solubility limit.

Justification: With most test articles, prior knowledge of their biological (toxicogenomic) potency is limited; therefore, NTP proposes to employ a broad dose range to capture the entire dose-response space adequately for all measured features. This proposed study design is consistent with recommendations for the design of studies for which the goal is deriving a BMD [15].

For in vivo study parameters, 6- to 8-week-old male Sprague Dawley rats were chosen, in part, to be congruent with subchronic toxicological assessments, which commonly employ this strain in this age range. Rats in this age range grow rapidly and, therefore, are often more susceptible to toxicological challenge. Male rats were chosen due to their legacy use in toxicogenomic studies. Legacy toxicogenomics data are often used for interpreting results obtained from new test articles. The duration of exposure selected was based on findings by Thomas et al. [17], which demonstrated that 5 days of exposure is long enough for non-bioaccumulative chemicals to achieve a genomic point of departure similar to that observed for apical endpoints such as cancer. Selecting organs/tissues for gene expression studies based on anticipated target organ is common practice in targeted toxicological assessments. Further, analyzing effects on the liver for every test article, particularly if administered by the oral route, is justified primarily because liver is the most frequently affected toxicological target. Second, liver responds to significant disease-related changes in other organ systems and, therefore, can often serve as a sensor for systemic toxicological effects [29]. Using a 5-day MTD as the highest dose helps ensure a clear toxicogenomic response at the top dose level that can be effectively modeled into lower dose levels. A strong response at the top dose is critical because without a clear toxicogenomic response, the data often yield highly uncertain estimates of potency.

For in vitro study parameters, the most critical decision point is likely the cell types to use. NTP has chosen to focus on using organotypic cell types as opposed to transformed cells because we believe that the data derived from organotypic cultures will have greater qualitative and quantitative biological relevance to the organs/tissues that we hope to model [30]. Further, using organotypic cell types allows for the evaluation of known, prototypical, expression-related changes that have been observed in intact organisms, thereby providing a level of model validation [30]. Similar to the MTD-based selection of the highest dose for in vivo studies, the highest dose in in vitro studies is meant to significantly challenge the cells to obtain a clear toxicogenomic response at the top dose level that can be modeled into lower dose levels.

5. Addressing Biological Interpretation

Approach: A major challenge for interpreting toxicogenomics data relates to linking the observations to higher-order phenotypes, such as pathology and other endpoints typically associated with adverse effects [31]. Although gene sets have been derived for a small subset of tissues [32], such an approach requires a large amount of combined pathology and gene expression data, which is both cost prohibitive and infeasible for many tissues due to the rarity of observed effects. An alternative approach to selecting gene sets is to focus solely on literature-curated information, such as the pathway sets within the KEGG (Kyoto Encyclopedia of Genes and Genomes) database [33]. These gene sets are often not derived from gene expression data, however, and instead come from a collection of mechanistic, molecular, and biological studies that consider many interrelationships between genes and proteins, only a fraction of which are related to changes in gene expression. The nature of the pathway curation, while useful for mechanistic insight, is often of limited utility for interpreting adversity in gene expression data. An integrated approach to gene-set derivation, which entails combining empirically derived (i.e., phenotypically anchored) signatures with curated pathways to identify limited redundancy "Hallmark" gene sets, has recently been published [16].

The Hallmark gene sets encompass a variety of biological processes from a gene expression-centric perspective (i.e., the curated genes in each set exhibit concordant changes in expression in response to specific stimuli). Further, due to the way that the gene sets were derived, they exhibit generalizability that tissue-specific signatures (i.e., those that are phenotypically anchored) often lack. Some challenges exist with the Hallmark gene sets in relation to their use for interpreting genomic dose-response data including (1) limited biological feature coverage (~4500 total genes curated to at least one Hallmark set out of approximately 21,000 genes in the genome); (2) limited association with toxicological/biological effects of concern; and (3) needed refinement for use in species other than humans.

Due to the limited redundancy and integrated selection of gene sets that is congruent with interpretation of the gene expression data, NTP proposes to use the Hallmark gene sets for interpreting genomic dose-response data. Further, NTP plans to expand the gene sets to cover a greater diversity of the biological space and in collaboration with the large toxicology community, provide toxicological contextualization of the gene sets. In addition, efforts are planned to create refined gene sets for species with toxicological utility (e.g., rat and mouse).

Expansion of the biological coverage of the Hallmark sets will consider several information resources. Where available, phenotypically anchored gene sets derived from toxicogenomic studies will be added to the Hallmark gene sets. To survey the totality of co-regulated gene space [as defined by data in the GEO (Gene Expression Omnibus) database], a co-regulation map will be generated and co-regulated gene sets not associated with a Hallmark gene set will be identified. These new Hallmark gene sets will be curated by determining their relationships with expert-curated gene sets (e.g., KEGG pathways) and through data-driven interpretation using experimental data contained in the Nextbio Correlation Engine [34]. Species-specific curation will entail integrating functional paralogs (e.g., cytochrome P450s) and when possible, mining species-specific, co-regulated, gene expression space.

In addition to qualitative interpretation of the genomic dose-response data, NTP proposes to interpret the results quantitatively by identifying the BMD median from the most sensitive gene set (i.e., the BMD from the gene set with the lowest BMD). The BMD median from this gene set is described as the dose at which a test article begins to demonstrate an effect on a biological system or a BEPOD, as described above.

Justification: The Hallmark gene sets are the product of an integrated approach to gene-set identification that leverages both expert-curated information and data-derived signatures from the literature. Hallmark gene sets exhibit limited redundancy, an immediate congruency with gene expression data (as many of the genes in a set are co-regulated), a high degree of generalizability, and offer combined mechanistic and phenotypic interpretability. The Hallmark gene sets can be expanded to cover more biological space by integrating phenotypically anchored, toxicity signatures and by mining co-regulated gene space contained within public databases such as GEO. Finally, through the mining of publicly available, gene expression data related to disease and pathogenesis, formulating data-driven toxicological interpretations and identifying associated citations will be possible for most of the Hallmark gene sets.

Using the BMD median of the most sensitive gene set to determine the most sensitive biological effect level is based on the observation from in vivo genomic dose-response studies that this approach leads to identification of the biological effect level that approximates the point of departure from guideline toxicology studies [17].

C. References

1. Collins, F.S., G.M. Gray, and J.R. Bucher, *Transforming environmental health protection.* Science, 2008. **319**(5865): p. 906-907.
2. Thomas, R.S., et al., *Incorporating new technologies into toxicity testing and risk assessment: moving from 21st century vision to a data-driven framework.* Toxicol Sci, 2013. **136**(1): p. 4-18.
3. Thomas, R.S., et al., *A method to integrate benchmark dose estimates with genomic data to assess the functional effects of chemical exposure.* Toxicol Sci, 2007. **98**(1): p. 240-8.
4. Bhat, V.S., et al., *Concordance of transcriptional and apical benchmark dose levels for conazole-induced liver effects in mice.* Toxicological Sciences, 2013. **136**(1): p. 205-215.
5. Dunnick, J.K., et al., *Hepatic transcriptomic alterations for N,N-dimethyl-p-toluidine (DMPT) and p-toluidine after 5-day exposure in rats.* Archives of Toxicology, 2017. **91**(4): p. 1685-1696.
6. Fader, K.A., et al., *2,3,7,8-Tetrachlorodibenzo-p-dioxin (TCDD)-elicited effects on bile acid homeostasis: Alterations in biosynthesis, enterohepatic circulation, and microbial metabolism.* Scientific Reports, 2017. **7**(1).
7. Kawamoto, T., et al., *Mechanism-based risk assessment strategy for drug-induced cholestasis using the transcriptional benchmark dose derived by toxicogenomics.* Journal of Toxicological Sciences, 2017. **42**(4): p. 427-436.
8. Rager, J.E., et al., *High-throughput screening data interpretation in the context of in vivo transcriptomic responses to oral Cr(VI) exposure.* Toxicological Sciences, 2017. **158**(1): p. 199-212.

9. Robinson, J.F., et al., *Triazole induced concentration-related gene signatures in rat whole embryo culture.* Reproductive Toxicology, 2012. **34**(2): p. 275-283.
10. Rowlands, J.C., et al., *A genomics-based analysis of relative potencies of dioxin-like compounds in primary rat hepatocytes.* Toxicological Sciences, 2013. **136**(2): p. 595-604.
11. Thompson, C.M., et al., *Transcriptomic responses in the oral cavity of F344 rats and B6C3F1 mice following exposure to Cr(VI): Implications for risk assessment.* Environmental and Molecular Mutagenesis, 2016. **57**(9): p. 706-716.
12. Yang, Y., et al., *Differential reconstructed gene interaction networks for deriving toxicity threshold in chemical risk assessment.* BMC Bioinformatics, 2013. **14**(SUPPL.14).
13. Dean, J.L., et al., *Application of gene set enrichment analysis for identification of chemically induced, biologically relevant transcriptomic networks and potential utilization in human health risk assessment.* Toxicological Sciences, 2017. **157**(1): p. 85-99.
14. Yang, L., B.C. Allen, and R.S. Thomas, *BMDExpress: a software tool for the benchmark dose analyses of genomic data.* BMC Genomics, 2007. **8**: p. 387.
15. Slob, W., et al., *A statistical evaluation of toxicity study designs for the estimation of the benchmark dose in continuous endpoints.* Toxicol Sci, 2005. **84**(1): p. 167-85.
16. Liberzon, A., et al., *The Molecular Signatures Database (MSigDB) hallmark gene set collection.* Cell Syst, 2015. **1**(6): p. 417-425.
17. Thomas, S.R., et al., *Temporal concordance between apical and transcriptional points of departure for chemical risk assessment.* Toxicological Sciences, 2013. **134**(1): p. 180-194.
18. Thomas, R.S., et al., *Incorporating new technologies into toxicity testing and risk assessment: Moving from 21st century vision to a data-driven framework.* Toxicological Sciences, 2013. **136**(1): p. 4-18.
19. Shi, L., *The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements.* Nature Biotechnology, 2006. **24**(9): p. 1151-1161.
20. Crump, K.S., *A new method for determining allowable daily intakes.* Fundamental and Applied Toxicology, 1984. **4**(5): p. 854-871.
21. Posada, D. and T.R. Buckley, *Model selection and model averaging in phylogenetics: Advantages of akaike information criterion and bayesian approaches over likelihood ratio tests.* Systematic Biology, 2004. **53**(5): p. 793-808.
22. Akaike, H., *Information theory and an extension of the maximum likelihood principle.* Second International Symposium on Information Theory, 1973: p. 267-281.
23. Crump, K.S., *Calculation of Benchmark Doses from Continuous Data.* Risk Analysis, 1995. **15**(1): p. 79-89.
24. Shockley, K.R., *Estimating Potency in High-Throughput Screening Experiments by Maximizing the Rate of Change in Weighted Shannon Entropy.* Sci Rep, 2016. **6**: p. 27897.
25. Budtz-Jorgensen, E., N. Keiding, and P. Grandjean, *Benchmark dose calculation from epidemiological data.* Biometrics, 2001. **57**(3): p. 698-706.
26. Hoyle, D.C., et al., *Making sense of microarray data distributions.* Bioinformatics, 2002. **18**(4): p. 576-84.

27. Farmahin, R., et al., *Recommended approaches in the application of toxicogenomics to derive points of departure for chemical risk assessment.* Archives of Toxicology, 2017. **91**(5): p. 2045-2065.
28. Holland-Letz, T. and A. Kopp-Schneider, *Optimal experimental designs for dose-response studies with continuous endpoints.* Arch Toxicol, 2015. **89**(11): p. 2059-68.
29. Edwards, L. and I.R. Wanless, *Mechanisms of liver involvement in systemic disease.* Best Pract Res Clin Gastroenterol, 2013. **27**(4): p. 471-83.
30. Gerets, H.H., et al., *Characterization of primary human hepatocytes, HepG2 cells, and HepaRG cells at the mRNA level and CYP activity in response to inducers and their predictivity for the detection of human hepatotoxins.* Cell Biol Toxicol, 2012. **28**(2): p. 69-87.
31. Paules, R., *Phenotypic anchoring: linking cause and effect.* Environ Health Perspect, 2003. **111**(6): p. A338-9.
32. Natsoulis, G., et al., *The liver pharmacological and xenobiotic gene response repertoire.* Molecular Systems Biology, 2008. **4**: p. 175.
33. Kanehisa, M., et al., *KEGG: new perspectives on genomes, pathways, diseases and drugs.* Nucleic Acids Res, 2017. **45**(D1): p. D353-D361.
34. Kupershmidt, I., et al., *Ontology-based meta-analysis of global collections of high-throughput public data.* PLoS One, 2010. **5**(9).