



HHS Public Access

Author manuscript

Regul Toxicol Pharmacol. Author manuscript; available in PMC 2021 March 18.

Published in final edited form as:

Regul Toxicol Pharmacol. 2015 July ; 72(2): 292–309. doi:10.1016/j.yrtph.2015.04.010.

Technical guide for applications of gene expression profiling in human health risk assessment of environmental chemicals

Julie A. Bourdon-Lacombe^a, Ivy D. Moffat^{a,*}, Michelle Deveau^a, Mainul Husain^b, Scott Auerbach^c, Daniel Krewski^d, Russell S. Thomas^e, Pierre R. Bushel^f, Andrew Williams^b, Carole L. Yauk^b

^aWater and Air Quality Bureau, Health Canada, Ottawa, ON, Canada

^bEnvironmental Health Science and Research Bureau, Health Canada, Ottawa, ON, Canada

^cBiomolecular Screening Branch, Division of the National Toxicology Program, National Institute of Environmental Health Sciences, Research Triangle Park, NC, United States

^dMcLaughlin Centre for Population Health Risk Assessment, University of Ottawa, Ottawa, ON, Canada

^eNational Centre for Computational Toxicology, U.S. Environmental Protection Agency, Research Triangle Park, NC, United States

^fBiostatistics and Computational Biology Branch, Division of Intramural Research, National Institute of Environmental Health Sciences, Research Triangle Park, NC, United States

Abstract

Toxicogenomics promises to be an important part of future human health risk assessment of environmental chemicals. The application of gene expression profiles (e.g., for hazard identification, chemical prioritization, chemical grouping, mode of action discovery, and quantitative analysis of response) is growing in the literature, but their use in formal risk assessment by regulatory agencies is relatively infrequent. Although additional validations for specific applications are required, gene expression data can be of immediate use for increasing confidence in chemical evaluations. We believe that a primary reason for the current lack of integration is the limited practical guidance available for risk assessment specialists with limited experience in genomics. The present manuscript provides basic information on gene expression profiling, along with guidance on evaluating the quality of genomic experiments and data, and interpretation of results presented in the form of heat maps, pathway analyses and other common approaches. Moreover, potential ways to integrate information from gene expression experiments

*Corresponding author at: Water and Air Quality Bureau, Health Canada, 269 Laurier Avenue, Ottawa, ON, Canada. Ivy.Moffat@hc-sc.gc.ca (I.D. Moffat).

Conflict of Interest

The authors declare no conflict of interest.

Disclaimer

The United States Environmental Protection Agency through its Office of Research and Development reviewed and approved this publication. However, it may not necessarily reflect official Agency policy, and reference to commercial products or services does not constitute endorsement.

Transparency Document

The Transparency document associated with this article can be found in the online version.

into current risk assessment are presented using published studies as examples. The primary objective of this work is to facilitate integration of gene expression data into human health risk assessments of environmental chemicals.

Keywords

Human health risk assessment; Toxicogenomics; Chemical mode of action; Hazard assessment; Risk characterization

1. Introduction

Chemical risk assessment agencies worldwide are facing challenges that require new toxicity testing approaches. Major limitations of current approaches include the high cost and length of time required for tests that rely on the observation of adverse clinical or pathological effects in whole animals. As a result, human health risk assessments have been performed for only a small fraction of chemicals in commerce. To date, chemical substances inventories in Canada, the United States, and Europe contain over 23,000 (Health Canada, 2003), 84,000 (U.S. EPA, 2013), and 107,000 (ECHA, 2011) compounds, respectively. In contrast, just over 1100 compounds are regulated under U.S. legal statutes (Dernbach, 1997), and occupational exposure limits from around the world have only been derived for approximately 6000 chemicals (Brandys and Brandys, 2008). Thus, there is an urgent need for faster and more cost-effective testing strategies capable of consistently predicting chemical toxicity, and the doses at which adverse effects occur in humans.

Gene expression profiling in the context of a toxicology study (also referred to as toxicogenomics) has been identified as a promising method to alleviate some of the current constraints on human health risk assessment of chemicals. Emerging science has demonstrated the utility of gene expression profiling in identifying likely health hazards and in deciphering chemical modes of action (U.S. EPA, 2009). In the long-term, gene expression profiling may be used in chemical screening to guide further testing approaches as well as to derive points of departure (PoDs) for chemicals with limited data (Thomas et al., 2013a). This is part of the larger vision for “toxicity testing in the 21st century”, in which recent advances in molecular biology are used to make more informed decisions relating to potential health risks of chemical exposures (Krewski et al., 2010, 2011; NRC, 2007). Toxicogenomics data have also been identified as occupying a prominent place in the next generation of risk science, as envisioned by the US Environmental Protection Agency (Krewski et al., 2014).

Toxicogenomics studies offer rich datasets that can provide valuable information on chemical toxicity relevant to human health risk assessment. Important applications to current risk assessment practices include: (1) improving confidence in selecting critical endpoints through building and supporting mechanistic information; (2) enhancing understanding of whether adverse effects observed in animals are likely to occur in humans via similar modes of action; (3) guidance in selecting appropriate risk assessment approaches (such as threshold or non-threshold approaches); and (4) supporting read-across for chemical groupings. However, applications of toxicogenomics data in risk assessment have been

limited, in part because there is very little information available that would allow a risk assessor with limited background in genomics to critically evaluate data quality and suitability for toxicological risk assessment.

The present manuscript provides an overview of criteria that can be used to assess the quality of toxicogenomics data, along with guidance on data interpretation. Recommendations for inclusion of these data in human health risk assessment are also provided, along with examples of their application. As this work is intended to guide non-specialists in using published genomics data to inform risk assessment, only high-level concepts have been presented, with references provided for those seeking more information on technical aspects. The manuscript is also divided into multiple sections to facilitate finding specific information when working through a toxicogenomics paper. The overarching objective of this work is to facilitate and promote the use of toxicogenomic data in human health risk assessment.

2. The basics: why gene expression profiling?

Every biological system (including cells, tissues, and whole organisms) must cope with changes in its environment, including exposure to toxic substances. A first line of defense in response to an environmental challenge can include alterations in gene expression, which generally translate into an increase or decrease in specific proteins required to carry out important tasks related to the maintenance of homeostasis. Gene expression profiles provide a snapshot of the system's overall response to a toxicant, which can be related to the mode of action (MoA) of the toxicant, and can be captured by measuring levels of messenger RNA (mRNA or protein coding RNA) in the system. These changes correspond to the molecular alterations that will give rise to phenotypic changes at higher levels of organization. Although recent research has demonstrated that non-coding RNAs are also important regulatory and structural molecules involved in biological responses (Bhan and Mandal, 2014; Cech and Steitz, 2014), the focus of this guide is primarily on mRNA.

In this article, we refer to gene expression profiling or toxicogenomics as the large scale measurement of changes in gene expression relative to control cells or tissues following a toxicological challenge. Gene expression profiling/toxicogenomics examines all of the genes in the system, or of a substantive portion of them, and takes into consideration that the human, rat and mouse genomes contain over 38,000, 29,000 and 33,000 genes, respectively (NCBI, 2011). Information on the identity of affected genes, the dose levels at which their function is altered, and the relationships among these genes is subsequently used to understand how a chemical is perturbing the system and to predict the adverse effects that may ensue.

It is important to note that toxicogenomics studies can vary considerably with respect to the biological questions under investigation as well as the technologies used to measure gene expression. Technical and biological considerations in toxicogenomics are briefly discussed below.

2.1. Technical considerations

Various technologies are used in gene expression profiling (McBride, 2015). Among those most frequently applied include DNA microarrays, large scale real-time quantitative polymerase chain reaction (RT-qPCR or qPCR) experiments, and, more recently, RNA sequencing (RNA-seq). These platforms each have their own inherent advantages and disadvantages (Table 1). In general, the data generated from these well-established toxicogenomic methodologies have been shown to be reproducible and concordant across platforms (Black et al., 2014; Shi, 2006; Wang et al., 2009; Yauk and Berndt, 2007; SEQC/MAQC-III Consortium, 2014). A brief overview of the concepts behind these technologies is presented in Fig. 1. Although the technical concepts underlying each method are different, the general concept is to identify changes in transcript abundance from exposed samples relative to controls.

2.2. Biological considerations

Experimental designs can differ markedly among toxicogenomic studies, depending on the biological question being addressed. Toxicogenomic studies have been conducted using various types of biological samples, including cells in culture, whole tissues of treated animals, or specific cell subtypes isolated from exposed humans or animals. Studies may involve single doses or seek to describe dose–response relationships using multiple doses. Moreover, the persistence of changes in gene expression responses are often investigated over time, using time course experiments. As is the case for all toxicity data, the study design will play an important role in the conclusions that can be drawn from the data. This will greatly affect how the data can be used in the context of human health risk assessment. Although the advantages and disadvantages of specific study designs will not be reviewed in this paper, examples of different types of toxicogenomics studies will be used to demonstrate their applications in risk assessment.

3. Determining the quality of toxicogenomics data

Data quality is critical to obtaining reproducible and reliable information from toxicogenomics studies. As noted above, significant improvements in toxicogenomics technologies and data analysis strategies have resulted in a high degree of concordance in data produced between different platforms and laboratories. However, there are several other considerations, such as experimental design, sample integrity, platform performance and analytical strategy, which should be assessed in order to ensure validity of the data. Additional approaches have been proposed to evaluate data quality using scoring criteria (e.g., Systematic Omics Analysis Review or SOAR tool) (McConnell et al., 2014). We believe it is in the best interest of the evaluator to clearly understand strengths and limitations of genomics studies, which will provide increased knowledge to enable the judgment calls often required in human health risk assessments. We proposed criteria in Tables 1–4 to provide a basis to evaluate whether or not a toxicogenomics experiment was designed, conducted and analyzed appropriately. The criteria ensure that the experiment has adequate power to measure gene expression changes while minimizing false-positive results. The criteria that have been deemed to be essential to experimental integrity are denoted with an asterisk (*). Studies that do not meet those criteria identified as critical should not be

considered for use in risk assessment. The additional criteria can serve as guidelines in assessing the overall data quality and in providing a rationale for inclusion or exclusion of specific studies.

3.1. Data normalization and annotation

An important consideration in handling global genomic data that warrants further discussion than can be provided in the tables is proper normalization (denoted as an essential criterion for microarray and RNA-seq data quality in Tables 2 and 4). Normalization ensures that each replicate is comparable to the others. For example, it removes differences in overall signal intensity across DNA microarrays resulting from variables such as RNA input, labeling, hybridization or other technical issues.

For microarrays, different commercial platforms are amenable to a variety of normalization approaches, which depend on both the type of array or the experimental design. The methodologies for normalizing DNA microarrays are considered mature and well established. We refer the reader to Zhang et al. (2009) and Welle (2013) for additional information on this topic. For RNA-seq, normalization will account for technical variables leading to differences in the number of reads per sample by expressing reads, for example, as a counts per million (number of times that site was counted per million reads). Normalization approaches for RNA-seq are established but are still being refined (Dillies et al., 2013). In addition to normalization, full transcriptome sequencing introduces a number of other computational challenges, including alignment of reads against the appropriate reference genome (i.e., how to figure out which site in the DNA sequence of an organism the small reads are complementary to), and proper annotation (assigning each 'read' to a gene or genomic location). We refer the reader to a review by Auer et al. (2012) on this topic.

Overall, it is important that the data be properly normalized (and annotated for RNA-seq data), particularly in older studies; normalization and annotation do not generally present an issue in more recent works. In the future of risk assessment, when toxicogenomics data may be considered more heavily in selections of PoDs, it would be suggested to consult with a biostatistician on topics related to normalization and annotation.

4. Understanding toxicogenomics studies

High content toxicogenomics studies can contain large amounts of data (i.e., thousands of measurements of gene transcript levels) and involve complex analyses of the relationships between affected genes and their association with potentially adverse health outcomes. As such information is not easily presentable in standard graphs and tables, approaches tailored to toxicogenomics data visualization have been developed and are used routinely in published studies. These analytical and data visualization approaches are largely unique to toxicogenomics studies. Below, we provide an overview of the concepts behind common data analysis approaches and guidelines on how these data can be interpreted, along with several examples taken from the literature. It should be noted that most toxicogenomics studies do not present all of these techniques within a report: although these techniques are generally used in combination to synthesize the results of an experiment, studies usually employ a selection of approaches that best represent the data.

4.1. Approaches for assessing treatment effects and general trends

The following approaches are frequently used to determine the presence or absence of a treatment effect and to broadly examine similarities and differences in gene expression across individual samples and experimental groups. These types of analyses are useful in determining the robustness of a treatment effect and in evaluating data quality. These techniques can also be used to compare various datasets from different experiments.

4.1.1. Reporting differentially expressed genes

4.1.1.1. Key information.: The number of differentially expressed genes provides information on the extent of a transcriptional response with regards to the number of genes affected and the extent to which these genes are affected. In some cases, the extent of the response across chemicals or animals can be compared to examine chemical potency or species susceptibilities.

4.1.1.2. Concept.: When a gene is referred to as being “differentially expressed”, it means that the authors consider that the number of copies of that gene’s transcript is either significantly increased (up-regulated) or decreased (down-regulated) as a result of an exposure or treatment following thorough statistical considerations. Standard practice is to report the number of differentially expressed genes for each treatment group relative to control samples. Two important criteria are used in establishing a list of differentially expressed genes: fold-change and statistical significance.

Fold-change refers to the ratio of expression change in the treated sample in comparison with the control sample reference or baseline. For each gene, this fold-change can represent an increase or a decrease in gene expression. Generally, fold-changes exceeding 2.0 or below -2.0 (representing a doubling or halving of the number of transcripts in the exposed sample vs. control) are considered to represent a significant change. However, the choice of a cut-off to determine fold-change significance is arbitrary and other values are often used (± 1.5 -fold is fairly common) (Shi et al., 2008; St. Laurent et al., 2013). Higher fold-changes are generally more robust and reproducible (Shi et al., 2008). It should be noted that some studies use the ratio of the transcript number in the treated sample to the transcript number in the control sample to indicate a down-regulation instead of negative values; in this event, a -2 -fold-change in expression would be denoted by a fold-change of 0.5. It should be noted that applying a fold-change cut-off does not necessarily relate to the biological importance of the genes identified.

Significance refers to statistically calculated p -values generated for each gene. Given that the statistical evaluation of gene expression profiles must take into consideration thousands of endpoints, some false-positive genes can be expected. For this reason, p -values are sometimes adjusted using a false-discovery rate (FDR) approach (Tusher et al., 2001) or corrected in some way for multiple testing (Benjamini and Hochberg, 1995; Storey and Tibshirani, 2003). Usually, an FDR adjusted p -value cut-off of less than 5% ($p \leq 0.05$ FDR) is considered to be significant, although this criterion is sometimes relaxed to allow more comprehensive pathways searches (see pathways and networks section below). The Microarray Quality Control Consortium, a US Food and Drug Administration led group, has

conducted a number of comprehensive studies on the use of fold-changes and statistical significance. These studies demonstrate that use of fold-change ranking coupled with a p -value cutoff leads to differentially expressed gene lists that are more reproducible across laboratories and different technologies (Shi, 2006; Shi et al., 2008).

4.1.1.3. Interpretation.: In principle, a chemical that induces expression changes across numerous genes is implicitly perturbing normal cellular functions. Although the number of genes that are differentially expressed does not provide information on specific biological responses in exposed cells, it does indicate the transcriptional response by the chemical in terms of the number of genes affected and the extent to which these are affected. Information on these genes can be mined to determine biological processes or molecular functions that are affected by the exposure (discussed further in subsequent sections).

We caution that when reporting and interpreting differential gene expression, it is important to examine the cut-off values employed by the investigators. A list of affected genes based on broader cut-offs (smaller fold-changes and larger p -values or p -values uncorrected for multiple testing) will be longer than a list based on more stringent cut-offs and can contain false positives. Furthermore, in principle, due to redundancy on gene expression arrays, several reports for a gene should be differentially expressed. If not, then further investigation as to disagreements is necessary. Provided that similar statistical analyses and cut-offs are used, comparing the number of differentially expressed genes across various experimental contexts (e.g., different chemicals, species, tissues, or cell lines), can provide information related to, for example, chemical potencies and differences in susceptibility across tissues, species (or genotypes), sex or life stages.

4.1.1.4. Examples.: The number of differentially expressed genes observed can be indicative of the potential for a chemical to cause an adverse effect. For example, the magnitude of the transcriptional response to naphthalene, as shown through the number of differentially expressed genes in rat nasal epithelium following subchronic inhalation, is highly correlated with tumor outcome in the two year cancer bioassay (Clewel et al., 2014). Although difficult to implement in current risk assessment practices, such information can be helpful in prioritizing chemicals for further toxicological investigation.

Nesnow et al. (2009) compared differentially expressed genes between phenobarbital and conazole fungicides propiconazole and triadimefon at their respective carcinogenic doses (850 ppm phenobarbital, 2500 ppm propiconazole and 1800 ppm triadimefon for 4 and 30 days). Using an FDR corrected p -value of 0.05 as a cut-off with no restrictions on fold-change, overlap between the gene lists was examined. Analysis revealed that many of the differentially expressed genes were unique to phenobarbital, with more commonalities observed between the two conazoles. This analysis suggests that the responses leading to liver tumors in rodents for conazoles is different from phenobarbital. This observation has important risk assessment implications, as a phenobarbital MoA involves hyperplastic and anti-apoptotic responses that do not appear to cause liver tumors in humans even at plasma phenobarbital concentrations that lead to hepatocarcinogenesis in rats (Holsapple et al., 2006).

Yao et al. (2012) used the number of differentially expressed genes to examine inter-strain hepatic responses to 2,3,7,8-tetra chlorodibenzo-p-dioxin (TCDD) in rats. The analysis revealed a significant difference in responses across six different rat strains/lines that are known to exhibit different sensitivities to TCDD. In the analysis, the most sensitive strains (determined according to LD₅₀) exhibited the greatest number of differentially expressed genes. This example shows how information on the number of perturbed genes can be useful in deciphering species and strain sensitivities that can potentially impact the dose at which toxicity is observed in a risk assessment context.

The number of differentially expressed genes can also be used to examine potency of similar substances. The potency of habitual tobacco and marijuana smoking was examined *in vitro* using smoke condensates (Maertens et al., 2013). The analysis showed there was a very high degree of concordance between the differentially expressed genes in both exposure scenarios. However, marijuana smoke condensate was shown to perturb many more genes than tobacco smoke condensate at the same concentration, indicating that it has a higher potency than tobacco smoke. This is consistent with potency based on genotoxicity and cytotoxicity. In another example, fold-changes of genes known to be involved in male reproductive toxicity were ranked to evaluate the potency of various phthalates (Hannas et al., 2012). Ranking based on gene expression were consistent with potency rankings based on testosterone production, further illustrating the utility of gene expression information in comparing the potency of chemicals.

4.1.2. Heat maps and hierarchical clustering

4.1.2.1. Key information.: Heat maps provide an informative visual representation of gene expression responses using color. Hierarchical (unsupervised) clustering is used to compare gene expression data across individual samples, treatment groups, or experiments. Together, these methods provide a global overview of the raw data, highlighting key differences, trends or patterns between groups and, in some cases, can narrow down subsets of affected genes that are critical in the response. Another type of clustering is known as biclustering (Hartigan, 1972; Chou and Bushel, 2009). Here, a subset of genes and treatment conditions are grouped by common expression patterns. In addition, supervised clustering (Han and Kamber, 2001) can be applied to the data. During supervised clustering, an external phenotype is used to assign samples to classes and guide the grouping of the gene expression data. In general, there are numerous types of clustering approaches that can be applied to identify relationships across gene expression patterns and treatment conditions.

4.1.2.2. Concept.: A typical heat map of the type shown in Fig. 2 uses two colors (usually, red and green or in some cases, blue and yellow) to represent increased regulation (red or blue) and decreased regulation (green or yellow). Differential gene expression may be gauged relative to control samples, a reference sample or a median expression value. The intensity of the color is associated with the magnitude of differential expression. Brighter colors represent greater differences in either direction. In a fairly typical heat map, each column within the heat map represents a specific dataset; columns may reflect the expression of genes from an individual sample within an experiment or pooled data for an entire treatment group, depending on the objectives of the analysis. Rows within the heat

map represent individual genes (or probes for a non-annotated mRNA). In most cases, the genes plotted in a heat map include those identified as differentially expressed by the treatment or genes from another pre-defined gene list (e.g., examining changes in genes associated with an important toxicological pathway or a set of genes considered to be predictive of an adverse effect). As such, each square within a heat map represents the expression of a single gene in a defined dataset.

Hierarchical clustering refers to the associated tree structure (dendrogram) on the *X*-and/or *Y*-axis of the heat map, which depicts the groupings of the expression profiles across each sample or groupings between genes based on some type of dissimilarity metric (i.e., Pearson correlation, Euclidian distance, etc.). Branching on the *X*-axis (i.e., the 'tree' associated with the figure) reveals which datasets are most similar in terms of expression, and which are further apart from one another. When treatment effects are observed, it is expected that samples from a group (e.g., biological replicates within a treatment group) will be highly similar and will thus 'cluster' or group together. For example, the shorter the branches in the branching structure, the more correlated two gene expression profiles are. In a heat map comparing datasets of exposure to different chemicals, those with similar MoAs would be expected to be highly correlated and thus cluster more tightly. It should be noted that the number of clusters or groups is dependent on the cut point of the dendrogram (i.e., how far down the branching structure you go to assign the groups), which can be subjective. However, this cut-point can also be determined objectively using specialized software (Tibshirani et al., 2000).

Hierarchical clustering is also conducted for genes on the *Y*-axis. This groups genes by the degree of correlation in their expression levels across samples and enables visualization of patterns within the heat map (although the tree itself is not always shown). Gene trees facilitate the identification of genes that respond similarly to a treatment and can provide important information on the underlying MoA.

4.1.2.3. Interpretation.: The primary objective of heat maps and hierarchical clustering is to explore the experimental, biological and technical factors that have the greatest influence on gene expression changes. Interpretation should not be focused on individual data points, but on a more general assessment of the overall response. In most toxicity studies, heat maps and clustering can be used to evaluate the strength of the response. For example, if all replicates within a treatment group cluster closely together, this suggests that response to the treatment is robust and reproducible across samples. However, it should be noted that given the biological variability that arises *in vivo*, as well as random correlations that can occur in large datasets, clustering within treatment groups is not always perfect. This is especially true when treatment effects are modest. To reduce the effect of random correlation, it is preferable to work with subsets of differentially expressed genes, which is generally what is done in most published works. In addition, assigning *p*-values to clusters to denote the significance of the groupings is ideal for risk assessment purposes. Heat maps and hierarchical clustering can also be used to compare data with other published experiments, such as those examining other chemicals, different species or responses in other tissues. Heat maps can also be used to compare toxicogenomics profiles to disease profiles, in order to obtain additional insight into MoAs for toxic substances. In any heat map in which a gene

tree is depicted, the patterns or patches of color that vary substantially between groups (i.e., showing increases or decreases in mRNA levels from one cluster to another) can be used to narrow down subsets of genes that may be important in activating specific pathways. Thus, heat maps and hierarchical clustering can serve many purposes.

4.1.2.4. Examples.: Heat maps and hierarchical clustering may be used to evaluate data consistency (e.g., to verify that treatment groups cluster together) or to assess overall effects of specific conditions/biological states on expression profiles. Other more specific applications can also be useful in risk assessment. In one study, gene expression profiles from the livers of rats treated with various genotoxic and non-genotoxic hepatocarcinogens at their respective carcinogenic doses were compared using heat maps (Ellinger-Ziegelbauer et al., 2009). Using genes involved in oxidative stress and DNA damage response, as well as cell cycle progression, obvious distinctions were made between expression patterns for the genotoxic chemicals (specifically, 2-nitrofluorene, dimethylnitrosamine, 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanol and aflatoxin) and the non-genotoxic chemicals (i.e., methapyrilene HCl, diethylstilbestrol, Wy-14643 and piperonyl butoxide). Thus, the gene expression profiles were useful in elucidating potential MoAs related to carcinogenesis. This approach was used successfully by Thompson et al. (2012) to support chromium's MoA in small intestinal tumorigenesis. Heat maps using genes related to various MoAs associated with carcinogenicity (e.g., cell cycle progression, proliferation, oxidative stress and regeneration) clearly show that chromium clusters with other non-genotoxic carcinogens and not with chemicals that are genotoxic (specifically, the same chemicals as in Ellinger-Ziegelbauer et al., 2009). This information is useful in risk assessments where the mode of carcinogenicity will guide interpretation, such as the adoption of threshold and non-threshold risk assessment approaches for carcinogens considered to be non-genotoxic and genotoxic, respectively.

Heat maps and hierarchical clustering can also be used to elucidate additional differences in experimental data. For example, Yauk et al. (2012) compared three tobacco smoke condensates from different brands to determine whether a brand marketed as 'reduced harm' was in fact less toxic. Cluster analysis demonstrated that the brand of condensate was the least important factor associated with gene expression; samples clustered first by concentration and then by time of sampling. These results corroborated those of mutagenicity and clastogenicity tests, which indicated that there were not significant differences in toxicity across brands, and that the brand marketed as 'reduced harm' was equally toxic to full flavor varieties.

4.1.3. Principal component analysis

4.1.3.1. Key information.: Visualization of toxicogenomics data can be difficult both because of the large number of genes included in the analysis and the need to consider the effects of multiple variables that can affect the results, such as dose and time. Principal components analysis (PCA) takes into consideration all of the study variables, yet allows simple data representation in a two or three dimensional plot. PCA makes it possible to visually assess how the gene expression profiles of individual samples are grouped across treatment conditions and other experimental variables.

4.1.3.2. Concept: PCA is generally done using a subset of genes determined to be differentially expressed. PCA is also used to compare gene expression data across individual samples, treatment groups or different experiments. It is a mathematical algorithm used to identify dimensions that capture the greatest variability in the data. These dimensions (termed principal components) encompass multiple variables, allowing PCA to reduce the dimensionality of the data by using a small number of principal components to describe variation in the data. In practice, it is usually possible to describe most ($\geq 70\%$) of the variation in the data using two or three principal components, which are used as the axes of a PCA plot describing variability among experimental samples, treatment groups or experiments. However, it should be noted that some data sets may require more than the top 3 principal components to capture the majority of the variability.

4.1.3.3. Interpretation.: PCA enables a visual representation of similarities and differences in the data elements according to their proximity within the PCA plot. PCA plots are often used to group data (Fig. 3), and are commonly used to determine if there is a specific treatment, time or dose effect. For example, clear delineation between control and treated groups would indicate an obvious treatment effect. Similarly, the data can be used to determine whether the biological responses change or remain consistent across time-points or other conditions. As with hierarchical cluster analysis, similarities to disease signatures or gene expression profiles for other chemicals can also be explored. Groupings are often made more apparent by ellipsoids (done either manually or as a statistical representation of the confidence regions encompassing similar groupings) around the data points.

4.1.3.4. Examples.: PCA can be used to examine whether or not samples subjected to the same experimental condition (e.g., dose, time) demonstrate similar responses. For example, Jackson et al. (2014, Fig. 1C) used PCA to demonstrate that gene expression patterns in rat liver samples following a 21 day exposure to carcinogenic and non-carcinogenic doses of the hepatocarcinogen furan cluster together according to dose, with clear separation from control samples. PCA can also be used to categorize different datasets. As described earlier, Thompson et al. (2012) compared gene expression profiles in mouse duodenum following treatment with chromium [Cr(VI)] to rat liver expression profiles from rodents treated with four mutagenic and non-mutagenic carcinogens in order to elucidate on the mechanisms of chromium-induced carcinogenesis. PCA analysis revealed that chromium-induced gene expression clustered most closely with non-mutagenic carcinogens, providing insight into the chemical's MoA.

4.2. Approaches for assessing gene functionality and interactions

It is important to understand that genes generally do not operate in isolation, but rather function together with other genes to carry out specific cellular functions. Consequently, the effects observed for a specific gene must also be considered within the context of the function or process of that specific gene in the cell (a concept often termed gene annotation, which is continuously updated as knowledge of each gene increases), how it interacts with other genes (e.g., in a network of genes), and how the other genes involved in those functions or networks also behave. The following approaches are useful in describing gene function and interactions and can help to elucidate the MoA for a chemical using modified

Bradford-Hill criteria (Meek and Klaunig, 2010; Meek et al., 2014) by identifying putative molecular initiating events (i.e., the initial interaction of the chemical with cellular molecules that triggers the downstream effects), key events within the MoA, and potentially adverse effects of exposure.

4.2.1. Gene functionality analysis

4.2.1.1. Key information.: Knowledge of the function of individual genes, especially the most affected ones, can provide information on potentially important biological perturbations occurring within biological systems. Larger groupings of differentially expressed genes that have similar functions provide evidence of specific biological responses potentially associated with adverse health effects.

4.2.1.2. Concept.: The process of assigning a gene to a functional group can be done by both literature searches and by applying computational approaches to retrieve pertinent information in bulk from databases. Thus, the type of analysis applied can vary from one study to another. One example is the Gene Ontology (GO) database (<http://www.geneontology.org>). Using this database, each gene (if it has been characterized) is assigned 'GO terms' that describe the molecular functions of the gene, the biological processes in which it is most likely to be involved, and the cellular location of the protein. GO terms use consistent terminology to describe a gene and can be used in analyses to determine which functional groups are over-represented within the study dataset. The Database for Annotation, Visualization and Integrated Discovery (DAVID) cluster analysis is one tool that uses GO terminology to do this (<http://david.abcc.ncifcrf.gov/>) (Huang et al., 2008). Other software can be used to group genes according to their functionality, including Metacore (<http://thomson-reuters.com/metacore/>) and Ingenuity Pathway Analysis (IPA) (<http://www.ingenuity.com/>), which have their own curated data annotations derived from information in the literature. Gene set enrichment analysis (GSEA) is a variant of the aforementioned gene functionality analysis approaches. GSEA identifies over-represented biological categories based on the ranking of the genes according to an enrichment score. In all cases, statistical tests are applied to determine whether the number of differentially expressed genes (given the number of genes annotated to a particular functional category) or the ranking of enrichment scores with a particular biological category or function is significantly greater or distributed differently than what would be expected by chance. GSEA has emerged as a powerful alternative to individual-gene analyses to reflect the functional relationship between genes in a set. Mootha et al. (2003) initially demonstrated the power of using pre-defined gene sets in a case where no individual gene's expression was significantly different between normal and diabetic patients. The goal of all gene set enrichment analysis (GSEA) methods is to identify functionally related genes that display coordinated expression changes.

4.2.1.3. Interpretation.: Irrespective of the method used to categorize gene functionality, the premise is that a larger number of differentially expressed genes related to a specific function or process is generally associated with a high probability that this function or process is perturbed in the system. When available, *p*-values may be useful in interpreting the strength of the functional association. Caution is required in the interpretation of these

data, as many genes can have multiple functions and play important roles in more than one pathway. Moreover, some of the terms used to describe gene function, such as metabolism, can be broad, and may be less informative than others that describe specific biological processes, such as DNA repair or apoptotic pathways.

4.2.1.4. Example.: This method is used routinely in toxicogenomics studies to determine the predominantly perturbed functions within a system. For example, Genter et al. (2002) examined gene expression profiles of alachlor, an agricultural chemical known to cause nasal turbinate tumors, similar to acetochlor and butachlor. Assessment of gene functionality of alachlor in rat olfactory mucosa revealed perturbations of genes involved in extracellular matrices, immune function, cellular proliferation, and apoptosis-related genes. Assuming that the MoA is the same for acetochlor- and alachlor-induced nasal turbinate tumors, the gene expression information was used by the EPA's Office of Pesticide Programs Cancer Assessment Review Committee to support cytotoxicity with regenerative cellular proliferation as the MoA for acetochlor, using read-across from gene expression profiles of alachlor (U.S. EPA, 2004).

4.2.2. Pathway and gene interaction network analysis

4.2.2.1. Key information.: Pathway and gene interaction networks examine interactions between gene products (i.e., proteins). When the probability that the relationships between genes is significantly greater than what would be expected by chance, the pathway or network is considered to be affected. This provides insight into the specific biological functions perturbed as a result of exposure that can be used in deciphering a chemical's MoA.

4.2.2.2. Concept.: Pathway and network approaches are used to examine gene interactions linked to specific biological processes and potential adverse health effects (Fig. 4). Pathways refer to known interactions between proteins and associated genes that perform specific cellular functions or lead to specific cellular endpoints. Networks are also based on small molecule/gene/protein interactions but with a much broader perspective; they encompass all potential (known) interactions among differentially expressed genes and pathways, without considering a specific endpoint (Fig. 4). Molecular pathways are generally well-established and supported in the literature, whereas interactions in networks can be more loosely defined.

A standard tool for pathway analysis is the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database (<http://www.genome.jp/kegg/pathway.html>), which contains hundreds of pathways involved in metabolism, genetic and environmental information processing, cellular processes, organismal systems and even human diseases. Additional sources include DAVID, Metacore (<http://thomsonreuters.com/metacore>), Ingenuity Pathway Analysis (<http://www.ingenuity.com/>), Panther (<http://www.pantherdb.org/>), Biocarta (<http://www.biocarta.com/>) and GenMAPP (<http://www.genmapp.org/>). Pathways that are most affected (often termed 'enriched') are determined by analyses of the percentage of differentially expressed genes within each pathway (i.e., are there significantly more genes differentially expressed within a pathway than would be expected by chance), in

combination with an evaluation of the nature of the changes in gene expression (i.e., fold-change and direction of change). As pathways are involved in a very specific response, they are useful in narrowing down potential toxic endpoints of concern following exposure. Specific examples include the p53 pathway involved in response to DNA damage and downstream cell cycle control, and the mevalonate pathway involved in biosynthesis of cholesterol.

With regards to networks, software tools (such as cytoscape (<http://www.cytoscape.org>), Metacore or Ingenuity Pathway Analysis) can be used to link large networks of molecules/genes/proteins/pathways to general biological responses (e.g., inflammatory responses and DNA damage response). Network analysis has proven to be of great value in determining the primary changes in a biological system and for identifying which genes and pathways are central to the overall biological response.

4.2.2.3. Interpretation.: Pathways and gene interaction networks provide an overview of the potential adverse effects that may result from exposure to toxic substances, and can be used in identifying MoAs for those substances. Because these analyses take into account multiple genes, pathway and network analyses may be more useful in elucidating MoA than analyses focused on individual genes (Pennie et al., 2004). Furthermore, concordance across laboratories and platforms increases when pathways and networks are considered instead of individual genes (Zhang et al., 2013). In order to confirm the perturbation of pathways/networks or hypothetical MoAs based on these, it is common to examine markers of the predicted toxicities directly in the test cells or animals. More weight can be placed on studies that employ confirmatory assays to validate their findings.

Networks can also aid in identifying the specific genes or pathways that are the most important to the overall response. For example, gene A in Fig. 4 is depicted as having multiple downstream effects. Thus, gene A may be considered as an important mediator of the overall biological response. Network nodes (central hubs that appear to drive many of the downstream consequences) can provide information on initiating and key molecular events involved in the MoA for the agent of interest. It is important to note that these central nodes (i.e., transcription factors or signaling molecule that appear to be connected and potentially 'driving' the perturbations) need not themselves be differentially expressed; for example, the p53 protein is activated by phosphorylation in order to function in regulating transcription.

4.2.2.4. Examples.: Pathway and network analyses represent one of the most common approaches for analyzing gene expression data. Pathway approaches specifically have been identified as the strategy that is most likely to prove to be both practical and useful for risk assessment purposes moving forward (U.S. EPA, 2014).

An example of the use of pathway analysis is provided by a case study examining the utility of genomics data for the qualitative human health risk assessment of dibutyl phthalate (Euling et al., 2013; U.S. EPA, 2009). Dibutyl phthalate is known to cause adverse male reproductive effects that are thought to be linked to reduced testosterone production (Makris et al., 2013). Analyses revealed perturbation of several pathways within testicular tissue that

support this mechanism of action (steroidogenesis, followed by lipid, sterol, and cholesterol transport). However, several pathways involved in other mechanisms were also identified, including perturbations in cellular growth and differentiation, peroxisome proliferation, and cell death. Thus, genomics data were able to inform risk assessment by identifying data gaps potentially indicating alternative or complementary MoAs and/or additional effects linked to exposure. Knowledge of pathways can be especially useful in general weight of evidence analyses for critical toxicological endpoints.

Network analyses, on the other hand, are generally used to identify larger-scale gene interactions affected by exposure. These can reveal relationships between the regulators and the different pathways that were affected, potentially elucidating a molecular sequence of events involved in the response. For example, networks were used to identify genotoxicity as a potential concern for exposure to carbon black nanoparticles (Bourdon et al., 2012a). However, delving further into the specific genes and pathway within the network, alongside other tests, helped to elucidate oxidative stress as the likely cause of genotoxicity (Bourdon et al., 2012a,b). In the context of risk assessment, the information provided by network analyses can be especially valuable for chemicals with limited toxicological information as it provides insight into the most likely outcome of exposure.

4.2.3. Analysis of upstream regulators

4.2.3.1. Key information.: Upstream regulators can include transcription factors as well as an array of small molecules capable of initiating specific gene expression responses. Signals transmitted from outside the cell are communicated to transcription factors, which then regulate gene expression. Transcription factors are critical regulatory proteins that control gene expression through DNA binding. Alone, or in combination with other proteins, transcription factors bind to DNA and promote or inhibit transcription of specific genes. Small molecules can include reactive oxygen species, hormones, and chemicals that initiate very specific downstream responses. Various databases are used to mine expression patterns induced by chemicals and effectively ‘work backwards’ to identify possible regulatory molecules responsible for the measured downstream changes in gene expression. Knowledge of the key regulators causing changes in transcription is directly relevant to risk assessment, as it provides information on the specific biological effect that initiated alterations in gene expression profiles; this knowledge can also be used in the elucidation of MoAs for environmental agents, and the identification of similar hazards using read-across approaches. An excellent depiction of an upstream regulatory analysis demonstrating the key regulatory molecules involved in tumorigenesis mediated through the TNF- α pathway is shown in Fig. 1 of (Breslin et al., 2005).

4.2.3.2. Concept.: Upstream regulators and master regulators at the top of the regulatory hierarchy can activate or suppress gene transcription during specific cellular events. For example, nuclear factor kappa B (NF- κ B) binds to DNA under conditions of inflammation, oxidative stress, and infection, and has been implicated in the transcription of over 150 genes involved in immune and inflammatory responses (Pahl, 1999). Similarly, the p53 transcription factor is activated when DNA damage occurs, allowing cell cycle changes and activation of DNA repair or apoptotic pathways. There are over a thousand known human

transcription factors with defined position weight binding matrices that are experimentally validated to be associated with particular biological effects (Vaquerizas et al., 2009) and various computational models and databases that can be used to identify transcription factors that are modulating observed gene expression changes (e.g., TRANSPATH and TRANSFAC; Krull et al., 2003; Wingender et al., 2001). Identifying which transcription factors are activated upon exposure can help to identify the MoA of chemical toxicants.

One approach to identifying upstream regulators that have been affected by exogenous exposures is to carefully examine the set of differentially expressed genes to determine if known targets of specific transcription factors or other signaling molecules possess the regulator's binding motif and are enriched in the dataset. The direction of the change in expression change (increase or decrease in transcript levels) can be used to predict whether a certain chemical activates or suppresses the activity of a transcription factor or signaling pathway. Given that many transcription factors are activated by specific ligands, the data can also be used to predict whether a chemical is an agonist or antagonist with respect to a specific transcription factor. As in any analysis of large datasets, statistical justification should be included when upstream regulators are presented and the biological plausibility for associations with transcription factors should be discussed.

4.2.3.3. Interpretation.: Identification of transcription factors using the entire gene list can help identify the major cellular processes affected by initial exposure to the agent of interest. In some cases, investigators may offer more substantive proof of activation by conducting *in vitro* reporter assays, such as luciferase assays, using the agent of interest. Identification of other upstream regulators, such as hormones, can be quantified within the biological system under study to lend further support to the relative importance of these regulators in the overall biological response. Identification of what may be referred to as a "master regulator" (i.e., the top of a regulatory hierarchy (Chan and Kyba, 2013)) can be especially beneficial to aid in the development of a MoA for risk assessment. These master regulators may be the specific molecules that control multiple downstream networks and pathways to lead to the associated adverse effect. Thus, master regulators in a gene expression analysis can be used to hypothesize the molecular initiating or key events in a MoA.

4.2.3.4. Examples.: Upstream regulator analysis is commonly used to predict initiating molecular events or other key events involved in chemical toxicity. This was done for the carcinogen furan described by Jackson et al. (2014). Upstream regulator analysis identified activation of reactive oxygen species, inflammation and growth factor regulation as critical events, which is consistent with the known MoA for furan-induced carcinogenesis (specifically, chronic cytotoxicity followed by sustained regenerative proliferation) (Jackson et al., 2014). In another example, interferon γ (IFN γ), which is associated with cytokine activity and immunoregulatory properties, was suggested to be a key molecular regulator of forestomach responses following oral exposure to benzo[a]pyrene in mice (Labib et al., 2013). There is increasing evidence indicating immunosuppressive properties of benzo[a]pyrene and this has important implications for carcinogenicity (Zaccaria and McClure, 2013). Such information can be used to narrow down the possible mechanisms of

chemical-induced toxicity and is thus useful for both MoA discovery and weight of evidence evaluation in human health risk assessment.

4.3. Approaches specific to human health risk assessment

The utility of gene expression profiling in human health risk assessment of chemicals is becoming increasingly apparent. Various practical demonstrations of utility and specific analytical approaches that have been developed are discussed below.

4.3.1. Comparison of effects observed in animals with responses in human cells

4.3.1.1. Key information.: When it is unclear whether or not an adverse effect observed in animals also applies to humans, studies in human cells can provide some evidence that pathway perturbations associated with an event in animals are likely, or unlikely, to occur in humans.

4.3.1.2. Concept.: Determining human relevance of effects observed in experimental animals can be challenging, particularly when epidemiological evidence is limited or unavailable. Studies using immortalized or primary human cells may provide supporting evidence for human relevance in these circumstances. For example, gene expression analysis can be used to demonstrate perturbations to the same pathways in human cells in culture as in an animal model, or support the hypothesis that activation/repression of key transcription factors or signaling molecules are similar across species. Likewise, predictive models built in animal model systems and based on gene expression may be used to extrapolate cause and effect in humans. Having a training gene expression data set from a rodent model system, one can use a variety of classifiers (i.e., k-nearest neighbor, support vector machines, neural networks) to identify a subset of genes that are highly predictive of a particular phenotype or endpoint. Using genomics and sequence similarity between species, these predictor genes can be “mapped” to the human species. With those predictors, one can then apply the classifier model to a test human gene expression data set to predict outcome. These findings can be used to determine if similar processes are operating in different species and provide guidance on the probability that downstream consequences (i.e., adverse health effects) observed in animals represent a relevant apical endpoint in humans.

4.3.1.3. Interpretation.: Several factors must be taken into consideration when interpreting studies using human cells for purposes of human health risk assessment. First, the cell type used will play an important role in the conclusions that can be drawn. It is important to select human cells that are relevant to the adverse effects under investigation (e.g., renal cells should be used as a basis for inferences about the risk of nephropathy). Greater confidence can generally be placed on studies employing primary untransformed cells as compared to transformed cells lines. Although the latter cell lines are used in most *in vitro* human studies and can provide evidence of human relevance, these genetically transformed cells may not accurately represent the biological response occurring in normal somatic cells (Milo et al., 1995).

An important factor to be taken into consideration is the dosing used *in vitro*. Dosing should reflect actual tissue concentrations expected or observed in the *in vivo* test system. Thus, pharmacokinetic data should be considered when comparisons between profiles of cells in culture and tissues of live animals are made.

The association between responses in human cells *in vitro* and *in vivo* will vary greatly depending on the nature of the agent under investigation. Confidence in human relevance is enhanced when the MoA is understood and the genes that are changing are known to be specific to the adverse health outcome (for example, specific genes are involved in repairing damaged DNA, whereas generalized metabolic changes may induce a less specific gene expression response).

In general, it is preferable to compare pathways instead of gene subsets. Specific genes involved in biological responses can vary from one species to another; however, the biological processes and their corresponding pathways are similar across species. Thus, pathway analyses can be more useful and more accurate than examining individual genes.

4.3.1.4. Example: Many studies have investigated gene expression differences between species to examine the plausibility of occurrence in humans (Dere et al., 2011; Vanden Heuvel et al., 2006; Richert et al., 2003; Martignoti et al., 2006). An excellent example is provided by the risk assessment of dimethyl arsenic acid (DMA) conducted by the US EPA Office of Prevention, Pesticides and Toxic Substances (U.S. EPA, 2006; Wilson et al., 2013). DMA is known to cause bladder tumors in rats. To examine the plausibility of this outcome in humans, gene expression profiles of rat urothelium (*in vitro* and *in vivo*) and human urothelial cells were compared (Sen et al., 2005). Overall, 12 and 49 pathways were common between the human cells and the *in vivo* and *in vitro* rat cells respectively, with only one pathway observed uniquely in human cells but not in the rat cells or tissue (Sen et al., 2007). This provided sufficient evidence to determine that bladder cancer is an outcome of relevance to humans. The information was used in conjunction with other toxicological data to recommend a reduction of the uncertainty factor for inter-species variability from 10 to 3. Additionally, blood gene expression acquired from rats exposed to subtoxic and toxic doses of acetaminophen for various durations were used to build classifiers to select genes highly predictive of exposure level. Expression profiles of 66 human orthologs classified the exposure status of acetaminophen-overdosed patients from normal individuals with a probability of 0.0061 (Bushel et al., 2007).

4.3.2. Comparisons with publicly available gene expression profiles

4.3.2.1. Key information: Various gene expression profiles covering a large array of biological conditions (including chemical toxicities, diseases and other health conditions in humans, animals and cells) are readily available from genomics repositories such as the Gene Expression Omnibus (<http://www.ncbi.nlm.nih.gov/geo/>), ArrayExpress (<http://www.ebi.ac.uk/arrayexpress/>) and Chemical Effects in Biological Systems (<http://www.niehs.nih.gov/research/resources/databases/cebs/>). The Comparative Toxicogenomics Database (<http://ctdbase.org/>) links biological conditions and diseases with chemical-gene

interactions. These gene expression profiles can be compared to toxicity profiles to provide further insight into the MoA for toxic chemicals and to elucidate potential adverse effects.

4.3.2.2. Concept.: Comparison of a toxicogenomics profile against publicly available expression profiles in order to decipher chemical toxicity is a strategy that is increasingly reported in the scientific literature. These analyses can be used to ask a variety of questions, including: (1) whether the chemical acts in a similar way as other well-studied chemicals (by comparing changes induced in expression for one chemical to another); (2) if a specific MoA or transcription factor has been induced (through comparison of chemically-induced profiles with those of model agents or knock-out/knock-in samples); (3) if the chemical induces expression profiles that are consistent with an adverse effect or disease (by comparing the chemically-induced expression changes to the gene expression profiles of diseased tissues or other chemicals that induce that particular adverse effect); and (4) if an animal MoA may also occur in humans (by comparing exposed animal model samples to human *in vitro* and/or *in vivo* data).

This concept has briefly been discussed in preceding sections in which heat maps and principal component analysis are discussed. These common analytical methods for gene expression data can also be used to compare different studies. An additional approach is the Prediction Analysis for Microarray (PAM) method (Tibshirani et al., 2002). This approach relies on the user to identify and compile gene expression profiles available either in public repositories or produced by the investigators. PAM effectively identifies a subset of genes that best characterize each gene expression profile.

Another approach involves deriving a gene expression profile for chemicals that act as positive and negative controls (e.g., presence or absence of liver tumors following chronic exposure) for a particular effect of interest to obtain a ‘training set’ against which expression profiles for other chemicals can be compared (Waters et al., 2010). The training set chemicals are used to identify the genes that accurately identify chemical MoAs involved in the development of the outcome, known as the ‘gene set’. Thus, chemical gene expression profiles can be compared to these profiles to elucidate whether or not the adverse effect is expected to occur. The results of such comparisons are generally presented using probability, summary statistics and heat maps. Recent efforts in the field have focused on whether or not gene expression profiles can be used to guide the necessity of the costly and time-consuming two year cancer bioassay, by prediction of cancer outcomes using short-term exposure gene expression profiles of exposed animals. The method appears to be promising for rat liver tumors (Ellinger-Ziegelbauer et al., 2008; Fielden et al., 2011, 2007; Nie et al., 2006; Uehara et al., 2011), mouse lung tumors (Thomas et al., 2009) and renal tubular toxicity (Fielden et al., 2005). Although more work will be necessary to fully implement this concept (i.e., guide toxicity testing), comparisons to training sets can be very useful in deciphering chemical MoA in the current risk assessment context.

Commercial software has recently been developed for comparing gene expression datasets. NextBio is the largest repository of publicly available gene expression profiles generated in various fields of study, including toxicology, pharmacology, and medicine. The software works by ranking the genes by responsiveness relative to a control sample (e.g., disease

versus control, or exposed versus control) and can be used to search for commonalities among datasets to reveal the most highly correlated profiles (Kupersmidt et al., 2010). This analytical technique can be used to identify toxicants that induce similar profiles (e.g., compound X is similar to compound Y), the disease profile that is most similar to the profile (compound X induces genes involved in disease Y), as well as provide other insights into chemical toxicity. Disease and exposure profiles can be mined to determine whether similar mechanisms are involved in disease development across species, and if the gene expression changes observed in animals would also be expected in humans.

Another smaller, but more targeted and integrated resource for comparing toxicogenomic gene expression data is the DrugMatrix Database and Analysis Tool (<https://ntp.niehs.nih.gov/drugmatrix/index.html>) that is freely available through the National Toxicology Program at NIEHS. DrugMatrix is a rat-centric toxicogenomics database that contains gene expression profiles for 638 different compounds; these compounds include FDA approved drugs, drugs approved in Europe and Japan, withdrawn drugs, drugs in preclinical and clinical studies, biochemical standards, and industrial and environmental toxicants. Nearly all of the gene expression data in DrugMatrix is paired with histopathology, clinical chemistry and hematology performed in the same animals from which the gene expression data were taken. The integration of the different data types in DrugMatrix allows for the identification gene expression signatures related to a number of organ-specific pathologies and modes of toxicological action. Notably, the toxicogenomic profiles from DrugMatrix are available in NextBio; however, the additional data (e.g., integrated histopathology) associated with the samples are not.

4.3.2.3. Interpretation.: Overall, the details relating to methodology and purpose of the comparison, as well as the strength of the association should be clearly presented. The results of such analyses can be used in conjunction with other data used in risk assessment to support a variety of possible associations including: MoA, groupings of chemicals or read-across between chemicals, human relevance, and onset of adverse health outcomes or disease.

4.3.2.4. Example.: In order to elucidate the potential health implications of exposure to carbon black nanoparticles (CBNPs), Bourdon et al. (2013) compared pulmonary gene expression profiles of CBNP exposed mice that exhibited inflammation to 13 profiles of various mouse inflammatory lung disease models. These included models of allergic airway inflammation, bacterial infection, and tissue injury and fibrosis. Using PAM, the authors determined that CBNP-induced mouse pulmonary gene expression profiles are correlated with profiles of lung injury and fibrosis. Using NextBio, they then determined that the molecular pathways in mice exhibiting fibrosis are similar to those seen in people with fibrosis. These analyses suggest that fibrosis is a likely outcome of exposure to CBNPs and that the response is likely to occur in humans.

In another example, daily oral gavage doses of hepatocarcinogenic and non-hepatocarcinogenic doses of furan (based on cancer occurrence patterns in rodents at two years of age) was administered to mice for three weeks, followed by hepatic gene expression analyses (Jackson et al., 2014). The authors used NextBio analysis of gene expression

profiles to predict that the most relevant health impacts of carcinogenic doses were liver injury, followed by liver regeneration, hepatic fibrosis, hepatocellular dysplasia, liver cancer, inflammation of the liver, and cirrhosis of the liver. These health effects are consistent with chronic cytotoxicity and sustained regenerative proliferation as the postulated MoA of furan-induced hepatic carcinomas, a hypothesis that is supported by the extensive literature on furan. This information can thus be used to support furan's MoA and the endpoint of concern in the context of human health risk assessment. Moreover, comparisons with other chemical profiles found that hepatic gene expression profiles were most consistent with malathion, chloroform, hydrogen peroxide, and acetaminophen. Were the MoA of furan unknown, this would have aided in deciphering potential mechanisms of action.

These examples suggest that comparisons to select profiles of disease models (such as in the CBNP example) or to a wide array all disease and toxicant-induced profiles (as in the furan example) can be useful in predicting potential health outcomes. Further comparisons of this type will help in elucidating the likelihood of a response observed in *in vitro* and animal test systems also occurring in exposed humans. In the current risk assessment paradigm, this can be useful for supporting the choice of a key endpoint of concern or determination of chemical MoA.

4.3.3. Dose–Response relationships and identification of biologically relevant doses

4.3.3.1. Key information.: The same techniques used to examine dose- or concentration–response relationships in traditional toxicity assays can also be used in gene expression studies to determine the levels of exposure associated with specific biological effects. These include the no-observed-adverse-effect level (NOAEL) and lowest-observed-adverse-effect level (LOAEL), as well as the benchmark dose (BMD) or benchmark concentration (BMC).

4.3.3.2. Concept.: It is becoming increasingly common to examine dose- or concentration–response relationships for specific genes or pathways that are determined to be biologically significant with respect to the overall response, or are thought to be linked to an adverse health outcome or disease. Exposure–response relationships can be examined by determining if fold-changes in specific genes increase or decrease significantly with dose, or if the number of differentially expressed genes in a pathway increases with dose. PoDs, such as the NOAEL and LOAEL, that are widely applied with traditional toxicological data also apply to toxicogenomics data. For example, the lowest dose corresponding to a significant effect on a specific gene or pathway corresponds to the LOAEL, and the dose below that at which no significant changes are observed (assuming the experimental design produced such a dose) would be the NOAEL. In certain cases, a no-observed-transcriptional-effect level (NOTEL) is assigned for an entire experiment and denotes the dose at which mRNAs are not significantly increased or decreased as a result of exposure (Lobenhofer et al., 2004).

BMD or BMC modeling is increasingly used in human health risk assessment, and is often preferred over the traditional NOAEL and LOAEL approaches because it makes use of all of the experimental data to produce a harmonized description of the entire exposure- or dose–response relationship (Sand et al., 2011) These approaches can be applied directly to specific

genes and pathways of interest. Recently, bioinformatics tools such as BMD Express (downloadable from <http://sourceforge.net/projects/bmdexpress/>) have been developed to do simultaneous modeling of dose–response across all differentially expressed genes or across all significant pathways, using the same continuous models used by the U.S. EPA Benchmark Dose Modeling Software (BMDs) to calculate BMDs and BMCs for standard toxicological endpoints (<http://www.epa.gov/ncea/bmds/index.html>). The number of studies reporting BMDs or BMCs is expected to increase as researchers continue to explore the potential utility of these as PODs for future human health risk assessments of chemicals (Thomas et al., 2011, 2012)

4.3.3.3. Interpretation.: As with traditional toxicological data, observation of a dose- or concentration–response relationship increases the weight-of-evidence that the effects on transcription, and the potentially related downstream effects of transcription (including both key events and adverse health outcomes), are related to exposure. Reference doses derived from the NOAEL/LOAEL or BMD/BMC may also be established (Sand et al., 2011). As is the case with traditional toxicological studies, lower reference doses or concentrations reflect an increased likelihood of chemically-induced toxic effects at low levels of exposure.

4.3.3.4. Examples.: There are many promising applications of dose–response assessments of toxicogenomics data that are relevant to the future of quantitative risk assessment. As the objective of this manuscript is to discuss topics related to current uses of toxicogenomics in risk assessment, issues pertaining to dose–response assessment are only briefly discussed here. However, it should be emphasized that as in all toxicological studies, evidence of dose–response trends increases the likelihood of the treatment being responsible for the effect. This information can be useful within the current risk assessment context.

It has been shown for five carcinogenic chemicals (i.e., 1,4-dichlorobenzene, 1,2,3-trichloropropane, propylene glycol, mono-*t*-butyl ether and methylene chloride) that BMDs and BMDLs calculated for standard cancer and non-cancer endpoints (e.g., tumors, organ weight, histopathology) in mice exposed for two years correlates well with the BMD/BMDL of the most sensitive GO category (i.e., lowest GO BMD) of mice exposed for only 13 weeks (Thomas et al., 2011, 2012). This concept was confirmed using an additional six chemicals with gene expression analysis of known target organs in rats following exposure for 5 days, and 2, 4, and 13 weeks (Thomas et al., 2013b). Comparison of BMDs for the most sensitive pathways to traditional toxicological endpoints showed a high degree of correlation at all time points.

Case studies of benzo[a]pyrene, conazoles, carbon black nanoparticles, and furan have demonstrated similar points of departure and/or margins of exposure (MOEs) using both traditional toxicological endpoints and genomics data (Bourdon et al., 2013; Jackson et al., 2014; Moffat et al., 2015; Nesnow et al., 2009). For example, a drinking water risk assessment was conducted on benzo[a]pyrene using: (a) traditional data alone; (b) genomics data alone; and (c) a combination of genomics and traditional data (Moffat et al., 2015; Chepelev et al., 2015). These three risk analytic scenarios resulted in very similar PoDs for risk assessment, using the most biologically relevant pathways (i.e., pathway perturbations

relating to the hypothesized MoA). These examples provide evidence of the expanded applications of genomics data in quantitative risk assessment in the future.

5. Overview of applications of toxicogenomics in human health risk assessment

Collectively, the examples presented in this paper document numerous practical applications of toxicogenomics data in risk assessment that extend beyond basic reporting. This includes the risk assessment of DMA by the U.S. EPA Office of Prevention, Pesticides and Toxic Substances, in which human relevance was determined by examining commonalities between pathways observed in treated rat and human urothelial cells (U.S. EPA, 2006; Sen et al., 2005, 2007). In another example from the U.S. EPA's Office of Pesticides Programs Cancer Assessment Review Committee, gene expression profiles for alachlor were used in the development of a MoA for acetochlor induced nasal turbinate tumors in rats (U.S. EPA, 2004; Genter et al., 2002). Toxicogenomics information was also considered in detail in the Integrated Risk Information System (IRIS) assessment of trichloroethylene (U.S. EPA, 2011). However, due to several questions surrounding the validity of toxicogenomics data, many of which have been elaborated upon in this paper, the information was not considered further in that evaluation.

In order to better understand how risk assessment agencies are incorporating gene expression data in their evaluations, we conducted an investigation of documents within Health Canada's Existing Substances Risk Assessment Bureau (ESRAB), Health Canada's Guidelines for Canadian Drinking Water Quality program, and EPA's IRIS program spanning the period January 1, 2000 to February 2, 2013. These programs were selected as their evaluations are readily available online and include important risk assessment initiatives within Canada and the United States. It was found that EPA IRIS assessments included gene expression information in 20% (12 of 59) of the evaluations conducted during this time period. Only 2% of evaluations within ESRAB (5 of 209) contained genomics information. Gene expression information was not used in any of the drinking water evaluations at Health Canada.

When gene expression information was used in IRIS and ESRAB assessments, it tended to be discussed qualitatively, adding to the database of studies of a chemical's toxicity or MoA. In fact, 82% and 42% of IRIS assessments employing gene expression data used genomics to support a specific MoA or health effect, respectively. In Health Canada ESRAB evaluations, 40% of the reports used gene expression information to support an MoA, and 20% to support determinations of adverse health effects. An additional 20% accounts for reports that cited gene expression studies, although the toxicogenomics data were not used within the document as part of the evaluation. Interestingly, though perhaps not surprisingly, reporting of gene expression data has increased significantly from 2000 to 2012, as shown in Fig. 5. It is unclear whether this increase is due to better understanding of gene expression data and its potential utility in human health risk assessment, to the increased number of published toxicogenomics studies available for use in risk assessment, or to the increased ease with which gene expression measures can be taken given advances in technology.

Consideration of all relevant toxicological information is important for ensuring the integrity of risk assessments. Given the broad utility of gene expression data as described in this article, we would expect a notable increase in the routine application of these types of data in human health risk assessment in the future.

Potential areas in which gene expression data can be especially useful in improving current risk assessments include increased confidence in selecting critical endpoints, determining human relevance of existent toxicological data, deciding on the most appropriate risk assessment approaches and strategies, and supporting read-across of toxicological data for relevant chemical groupings. For example, reporting of gene expression changes related to a specific adverse health effect can add to the overall weight of evidence for that critical endpoint. This can be especially useful for chemicals for which information on the critical effect is limited by a lack of strong candidate studies, such as traditional — albeit time-consuming and expensive — long-term rodent toxicity studies, which is often the case for difficult to assess endpoints such as neurotoxicity. Moreover, certain chemicals such as solvents can elicit non-specific toxicities in various tissues. Gene expression information can help strengthen the rationale for selection of the specific critical endpoint and target tissue on which to base risk assessment.

Development of MoAs using gene expression data in combination with other mechanistic data can be of great value within the current risk assessment paradigm. The MoA can be used to evaluate the possibility of key events occurring in humans, thus strengthening the rationale for selection of a specific critical effect on which to base a risk assessment (Fig. 6). This can be useful, for example, when evaluating carcinogens such as benzo[a]pyrene (IARC, 2010) for which there exists substantial information in animals, but little evidence in humans. In some cases, MoA data can also be used to exclude the use of a specific endpoint; this can be especially useful in assessing endpoints such as kidney tumors in rats, as these can often occur due to an MoA that is not relevant to humans. More thorough comparisons with various disease models (as described in Section 4.3.2) can be especially useful in determining a chemical's potential to elicit similar responses and effects in humans as seen in *in vitro* and/or animal studies.

Current risk assessment practice is moving away from using default approaches, such as using a threshold approach exclusively for non-cancer effects and low dose linear extrapolations for all cancer endpoints. Gene expression data can guide the selection of appropriate approaches through increased understanding of the MoA. For example, an extensive MoA evaluation that considered gene expression information has been published for chromium (Thompson et al., 2012, 2013). These analyses point to a non-genotoxic MoA for chromium-induced intestinal tumors and can be used to argue for the use of a threshold approach over the traditional linear low dose extrapolation for cancer endpoints in this case. It is also common for some risk assessment agencies to use read-across for chemicals within a common chemical grouping (such as polycyclic aromatic hydrocarbons, arsenicals, or phthalates) when information on the compound to be evaluated is scarce, but a rich database of toxicological information for a similar compound within the group is available. Toxicogenomics can aid in ensuring that these compounds do in fact behave similarly within the biological system of interest (an essential component enabling read-across) and may

even provide insight into relative chemical potency (Hannas et al., 2011, 2012). Overall, gene expression data can contribute to current evaluations by providing data that strengthen toxicological evaluations and subsequent risk decisions.

Finally, it is worth considering the advantages and disadvantages of inclusion of genomics data in a risk assessment. If good strong datasets are available, we see no disadvantages in considering the data in a risk assessment. However, caution must be exercised to ensure that the data are of high quality, as there are certainly numerous DNA microarray publications that contain poorly designed experiments (e.g., insufficient replicates, inappropriate methodologies applied, etc.) that could provide false information about effects or mode of action. This is, in part, one of the reasons for the present document. For extensive discussions on the advantages and shortcomings of toxicogenomics datasets for the current risk assessment paradigm we point the reader to our recent case study comparing toxicogenomics and traditional approaches in the risk assessment of benzo[a]pyrene (Moffat et al., 2015) and the associated commentary to the case study (Chepelev et al., 2015).

6. Concluding remarks

Gene expression data tend to be presented and evaluated using approaches that are not generally applied in traditional toxicity studies. The information in Sections 2-4 of this document is meant to serve as a practical reference to aid in interpreting toxicogenomics data. As with all studies considered in risk assessment, some will carry more weight than others. Table 5 provides an overview of how each method can be used within the current risk assessment paradigm.

The utility of gene expression data to risk assessment will vary, depending on study quality and relevance of the available genomics information, as well as the specific data gaps to be addressed and objectives of the assessment. As such, the degree to which gene expression data may be incorporated and used in a toxicological evaluation will vary on a case by case basis. A general strategy for the inclusion of gene expression data in risk assessment is presented in Fig. 6. Although the specifics of such a strategy will undoubtedly vary both among assessments and across agencies, it demonstrates that toxicogenomics data could be effectively used in risk assessment. It also illustrates that toxicogenomics is highly relevant to current risk assessment practices. As an overarching goal of the NRC (2007) vision for *Toxicity Testing in the 21st Century* is to use genomics data quantitatively, it is anticipated that gene-based PODs may be used for risk assessment in the future (see Fig. 6).

This work documents a number of specific applications of analytical toxicogenomics approaches and how they can be used in risk assessment to provide insight into potential human health hazards associated with exposure to toxic chemicals, and to the MoAs of such agents. It is anticipated that the utility of toxicogenomics data to risk assessment will continue to increase, as knowledge in this area is expanding rapidly and major international efforts underway are now focused on increasing our knowledge of pathway perturbations and their associations with adverse health effects (Krewski et al., 2010; NRC, 2007). For example, in 2012 the Organisation for Economic Cooperation and Development released the Adverse Outcome Pathway (AOP) programme (<http://www.oecd.org/chemicalsafety/testing/>

[adverse-outcome-pathways-molecular-screening-and-toxicogenomics.htm#Documents](#)). AOPs map the progression of a toxicological response from the initial interaction of the chemical with cellular molecules through to an adverse effect at the individual or population level. In September 2014, the OECD officially released the AOP-wiki, which makes peer-reviewed AOPs publicly accessible in order to consolidate knowledge of toxicity pathways. The overarching goal is to produce a web-based application that consolidates knowledge of how toxicants cause adverse effects in human populations, and provide a weight of evidence summary supporting the associations within the pathway, across its components, and the adverse health outcome. Another highly relevant initiative is the mapping of the human toxome (<http://humantoxome.com/>) (Hartung and McBride, 2011; Hartung, 2015), which aims to create a public database of toxicity pathways, thus providing a means to comprehensively map human responses to toxicants (Bouhifd et al., 2014). Collectively, these initiatives will significantly increase the utility and integration of toxicogenomics changes in routine toxicity evaluations and regulatory risk assessments over the long-term.

Application of toxicogenomics to human health risk assessment of chemicals is a field that continues to expand and that is expected to eventually transform toxicological evaluation practices, thus allowing more efficient evaluations of the large numbers of chemicals for which there exist inadequate toxicity profiles. However, full implementation of the vision for *Toxicity Testing in the 21st Century* in the 21st century is a very large task that still requires some time to complete (Krewski, 2014). We hope that the present document will help to promote the implementation of the NRC vision, and, in the interim, serve to enrich the current human health risk assessment paradigm through expanded use of toxicogenomics data along the lines discussed in this paper.

Acknowledgments

We would like to thank Richard Charron and Errol Thomson of Health Canada, as well as Jennifer Fostel of NIEHS for providing comments. We would also like to acknowledge management at Health Canada's Water and Air Quality Bureau for their support in the development of these guidelines.

References

- Auer PL, Srivastava S, Doerge RW, 2012. Differential expression- the next generation and beyond. *Brief Funct. Genomics* 1 (1), 57–62.
- Benjamini Y, Hochberg Y, 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* 57 (1), 289–300.
- Bhan A, Mandal SS, 2014. Long noncoding RNAs: emerging stars in gene regulation, epigenetics and human disease. *Chem Med. Chem* 9 (9), 1932–1956. [PubMed: 24677606]
- Black MB, Parks BB, Pluta L, Chu TM, Allen BC, Wolfinger RD, Thomas RS, 2014. Comparison of microarrays and RNA-seq for gene expression analyses of dose-response experiments. *Toxicol. Sci* 137 (2), 385–403. [PubMed: 24194394]
- Bouhifd M, Hogberg HT, Kleensang A, Maertens A, Zhao L, Hartung T, 2014. Mapping the human toxome by systems toxicology. *Basic Clin. Pharmacol. Toxicol* 115 (1), 24–31. [PubMed: 24443875]
- Bourdon JA, Halappanavar S, Saber AT, Jacobsen NR, Williams A, Wallin H, Vogel U, Yauk CL, 2012a. Hepatic and pulmonary toxicogenomic profiles in mice intratracheally instilled with carbon black nanoparticles reveal pulmonary inflammation, acute phase response, and alterations in cholesterol homeostasis. *Toxicol. Sci* 127 (2), 474–484. [PubMed: 22461453]

- Bourdon JA, Saber AT, Jacobsen NR, Jensen KA, Madsen AM, Lamson JS, Wallin H, Møller P, Loft S, Yauk CL, Vogel U, 2012b. Carbon black nanoparticle instillation induces sustained inflammation and genotoxicity in mouse lung and liver. *Part. Fibre Toxicol* 9 (5), 1–14. [PubMed: 22239852]
- Bourdon JA, Williams A, Kuo B, Moffat I, White PA, Halappanavar S, Vogel U, Wallin H, Yauk C, 2013. Gene expression profiling to identify potentially relevant disease outcomes and support human health risk assessment for carbon black nanoparticle exposure. *Toxicology* 303, 83–93. [PubMed: 23146762]
- Brandys RLN, Brandys YM, 2008. *Global Occupational Exposure Limits for Over 6000 Specific Chemicals*, second ed. Occupational and Environmental Health Consulting Services, 635 Harding Road, Hingdale, IL.
- Breslin T, Krogh M, Peterson C, Troein C, 2005. Signal transduction pathway profiling of individual tumor samples. *BMC Bioinformatics* 29 (6), 163.
- Bushel PR, Heinloth AN, Li J, Huang L, Chou JW, Boorman GA, Malarkey DE, Houle CD, Ward SM, Wilson RE, Fannin RD, Russo MW, Watkins PB, Tennant RW, Paules RS, 2007. Blood gene expression signatures predict exposure levels. *Proc. Natl. Acad. Sci. U.S.A* 104 (46), 18211–18216. [PubMed: 17984051]
- Cech TR, Steitz JA, 2014. The noncoding RNA revolution- trashing old rules to forge new ones. *Cell* 157 (1), 77–94. [PubMed: 24679528]
- Chan SS-K, Kyba M, 2013. What is a master regulator? *J. Stem Cell Res. Ther* 4 (3), 114.
- Chepelev N, Moffat I, Labib S, Bourdon-Lacombe JA, Buick J, Lemieux F, Malik A, Halappanavar S, Williams A, Yauk C, 2015. Integrating toxicogenomics into human health risk assessment: lessons learned from the benzo[a]pyrene case study. *Crit. Rev. Toxicol* 45 (1), 44–52. [PubMed: 25605027]
- Chou JW, Bushel PR, 2009. Discernment of possible mechanisms of hepatotoxicity via biological processes over-represented by co-expressed genes. *BMC Genomics* 10, 272. [PubMed: 19538742]
- Clewell HJ, Efremenko A, Campbell DE, Thomas RS, 2014. Transcriptional responses in the rat nasal epithelium following subchronic inhalation of naphthalene vapor. *Toxicol. Appl. Pharmacol* 280, 78–85. [PubMed: 24976557]
- Dere E, Lee AW, Burgoon LD, Zacharewski TR, 2011. Differences in TCDD-elicited gene expression profiles in human HepG2, mouse Hepa1c1c7 and rat H4IIE hepatoma cells. *BMC Genomics* 12, 193. [PubMed: 21496263]
- Dernbach JC, 1997. The unfocused regulation of toxic and hazardous pollutants. *Harvard Environ. Law Rev* 21 (1), 1–82.
- Dillies MA, Rau A, Aubert J, Hennequet-Antier C, Jeanmougin M, Servant N, Keime C, Marot G, Castel D, Estelle J, Guernec G, Jagla B, Jouneau L, Laloë D, Le Gall C, Schaëffer B, Le Com S, Guedj M, Jaffrézic F, French StatOmique Consortium, 2013, 2013. A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. *Brief Bioinform.* 14 (6), 671–683. [PubMed: 22988256]
- ECHA, 2011. Classification and labelling notification report on January 4 2011. Retrieved from <http://echa.europa.eu/documents/10162/13585/clp_final_report_20110104_en.pdf>.
- Ellinger-Ziegelbauer H, Gmuender H, Bandenburg A, Ahr HJ, 2008. Prediction of a carcinogenic potential of rat hepatocarcinogens using toxicogenomics analysis of short-term in vivo studies. *Mutat. Res* 637, 23–39. [PubMed: 17689568]
- Ellinger-Ziegelbauer H, Aubrecht J, Kleinjans JC, Ahr H-J, 2009. Application of toxicogenomics to study mechanisms of genotoxicity and carcinogenicity. *Toxicol. Lett* 186, 36–44. [PubMed: 18822359]
- Euling SY, White LD, Kim A, Sen B, Wilson VS, Keshava C, Keshava N, Hester S, Ovacik MA, Ierapetritou MG, Androulakis IP, Gaido KG, 2013. Use of genomics data in risk assessment case study: evaluation of the dibutyl phthalate toxicogenomic data set. *Toxicol. Appl. Pharmacol* 271, 349–362. [PubMed: 21745491]
- Fielden MR, Eynon BP, Natsoulis G, Jarnagin K, Banas D, Kolaja KL, 2005. A gene expression signature that predicts the future onset of drug-induced renal tubular toxicity. *Toxicol. Pathol* 33 (6), 675–683. [PubMed: 16239200]

- Fielden MR, Brennan R, Gollub J, 2007. A gene expression biomarker provides early prediction and mechanistic assessment of hepatic tumor induction by nongenotoxic chemicals. *Toxicol. Sci* 99 (1), 90–100. [PubMed: 17557906]
- Fielden MR, Adai A, Dunn RT, Olaharski A, Searfoss G, Sina J, Aubrecht J, Boitier E, Nioi P, Auerbach S, Jacobsen-Kram D, Raghavan N, Yang Y, Kinaid A, Sherlock J, Chen S-J, Car B, 2011. Development and evaluation of a genomic signature for the prediction and mechanistic assessment of nongenotoxic hepatocarcinogens in the rat. *Toxicol. Sci* 124 (1), 54–74. [PubMed: 21813463]
- Genter MG, Burman DM, Vijayakumar S, Ebert CL, Aronow BJ, 2002. Genomic analysis of alachlor-induced oncogenesis in rat olfactory mucosa. *Physiol. Genomics* 12 (1), 35–45. [PubMed: 12419858]
- Han J, Kamber M, 2001. *Data Mining: Concepts and Techniques (Morgan-Kaufman Series of Data Management Systems)*. Academic Press, San Diego, ISBN 1-55860-489-8.
- Hannas BR, Lambright CS, Furr J, Howdeshell KL, Wilson VS, Gray LE Jr., 2011. Dose-response assessment of fetal testosterone production and gene expression levels in the testes following *in vitro* exposure to diethylhexyl phthalate, diisobutyl phthalate, diisoteptyl phthalate, and diisononyl phthalate. *Toxicol. Sci* 123 (1), 206–216. [PubMed: 21633115]
- Hannas BR, Lambright CS, Furr J, Evans N, Foster PM, Gray EL, Wilson VS, 2012. Genomic biomarker of phthalate-induced male reproductive developmental toxicity: a targeted RT-PCR array approach for defining relative potency. *Toxicol. Sci* 125 (2), 544–557. [PubMed: 22112501]
- Hartigan JA, 1972. Direct clustering of a data matrix. *J. Am. Stat. Assoc* 67 (337), 123–129.
- Hartung T, 2015. Evolution of toxicological science: the need for change. *J. Toxicol. Environ. Health* (Submitted).
- Hartung T, McBride, 2011. Food for thought...on mapping the human toxome. *ALTEX* 28, 83–93. [PubMed: 21625825]
- Health Canada, 2003. Proposal for priority setting for existing substances on the Domestic Substances List under the Canadian Environmental Protection Act, 1999: Greatest potential for human exposure. pp. 1–62.
- Holsapple MP, Pitot HC, Cohen SH, Boobis AR, Klauning JE, Pastoor T, Dellarco VL, Dragan YP, 2006. Mode of action in relevance of rodent liver tumors to human cancer risk. *Toxicol. Sci* 89 (1), 51–56. [PubMed: 16221960]
- Huang DW, Sherman BT, Lempicki, 2008. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc* 4, 44–57.
- IARC: International Agency for Research on Cancer, 2010. *Some Non-heterocyclic Polycyclic Aromatic Hydrocarbons and Some Related Exposures. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans, Vol. 92*. WHO Press, Geneva, pp. 1–853. [PubMed: 21141735]
- Jackson AF, Williams A, Recio L, Waters MD, Lambert I, Yauk CL, 2014. Case study on the utility of hepatic global gene expression profiling in the risk assessment of the carcinogen furan. *Toxicol. Appl. Pharmacol* 274, 63–77. [PubMed: 24183702]
- Krewski D, 2014. Progress Made on Tox21: A Framework for the Next Generation of Risk Science. Inaugural presentation to the Society of Toxicology Risk Assessment Lecture Series. <<http://www.toxicology.org/ms/presentPortal.asp#tox21>>.
- Krewski D, Acosta D, Andersen M, Anderson H, Bailar JC, Boekelheide K, Brent R, Charnley G, Cheung VG, Green S Jr., Kelsey KT, Kerkvliet NI, Li AA, McCray L, Meyer O, Patterson RD, Pennie W, Scala RA, Solomon GM, Stephens M, Yager J, Zeise L, Staff of Committee on Toxicity Testing and Assessment of Environmental Agents, 2010. Toxicity testing in the 21st century: a vision and a strategy. *J. Toxicol. Environ. Health B Crit. Rev* 13 (2–4), 51–138. [PubMed: 20574894]
- Krewski D, Westphal M, Al-Zoughool M, Croteau MC, Andersen ME, 2011. New directions in toxicity testing. *Annu. Rev. Public Health* 32, 161–178. [PubMed: 21219154]
- Krewski D, Westphal M, Andersen ME, Paoli GM, Chiu WA, Al-Zoughool M, Croteau MC, Burgoon LD, Cote I, 2014. A framework for the next generation of risk science. *Environ. Health Perspect* 122, 796–805. [PubMed: 24727499]

- Krull M, Voss N, Pistor S, Potapov A, Wingender E, 2003. TRANSPATH: an integrated database on signal transduction and a tool for array analysis. *Nucleic Acids Res.* 31, 97–100. [PubMed: 12519957]
- Kupersmidt I, Su QJ, Grewal A, Sundaresh S, Halperin I, Flynn J, Shekar M, Wang H, Park, Cui W, Wall GD, Wisotzkey R, Alag S, Akhtari S, Ronaghi M, 2010. Ontology-based meta-analysis of global collections of high-throughput public data. *PLoS ONE* 5 (9), e13066. 10.1371/journal.pone.0013066. [PubMed: 20927376]
- Labib S, Guo C, Williams A, Yauk CL, White PA, Halappanavar S, 2013. Toxicogenomic outcomes predictive of forestomach carcinogenesis following exposure to benzo(a)pyrene: relevance to human cancer risk. *Toxicol. Appl. Pharmacol* 273, 269–280. [PubMed: 23735875]
- Laurent G St., Shtokalo D, Tackett MR, Yang Z, Vyatkin Y, Milos PM, Seilheimer B, McCaffrey TA, Kapranov P, 2013. On the importance of small changes in RNA expression. *Methods* 63 (1), 18–24. [PubMed: 23563143]
- Lobenhofer EK, Cui X, Bennett L, Cable PL, Merrick BA, Churchill GA, Afshari CA, 2004. Exploration of low-dose estrogen effects: identification of no observed transcriptional effect level (NOTEL). *Toxicol. Pathol* 32 (4), 482–492. [PubMed: 15223774]
- Maertens RM, White PA, Williams A, Yauk CL, 2013. A global toxicogenomic analysis investigating the mechanistic differences between tobacco and marijuana smoke condensates in vitro. *Toxicology* 308, 60–73. [PubMed: 23542559]
- Makris S, Euling SY, Gray LE Jr., Benson R, Foster P, 2013. Use of genomics data in risk assessment case study: I. Evaluation of the dibutyl phthalate male reproductive development toxicity data set. *Toxicol. Appl. Pharmacol* 271 (3), 336–348. [PubMed: 20849870]
- Martignoti M, Groothuis GMM, Kanter R, 2006. Species differences between mouse, rat, dog, monkey and human CYP-mediated drug metabolism, inhibition and induction. *Expert Opin. Drug Metab. Toxicol* 2 (6), 875–894. [PubMed: 17125407]
- McBride MT, 2015. Future platforms for toxicity testing. *J. Toxicol. Environ. Health* (Submitted).
- McConnell ER, Bell SM, Cote I, Wang R-L, Perkins EJ, Garcia, Reyero N, Gong P, Burgoon L, 2014. Systematic Omics Analysis Review (SOAR) tool to support risk assessment. *PLoS ONE* 9 (12), e110379. 10.1371/journal.pone.0110379. [PubMed: 25531884]
- Meek ME, Klaunig JE, 2010. Proposed mode of action of benzene-induced leukemia: interpreting available data and identifying critical data gaps for risk assessment. *Chem. Biol. Interact* 184 (1–2), 279–285. [PubMed: 20153303]
- Meek ME, Palermo CM, Bachman AM, North CR, Lewis RJ, 2014. Mode of action human relevance (species concordance) framework: evolution of the Bradford Hill considerations and comparative analysis of weight of evidence. *J. Appl. Toxicol* 34 (6), 595–606. 10.1002/jat.2984. [PubMed: 24777878]
- Milo GE, Shuler CF, Lee H, Casto BC, 1995. A conundrum in molecular toxicology: molecular and biological changes during neoplastic transformation of human cells. *Cell Biol. Toxicol* 11 (6), 329–345. [PubMed: 8788209]
- Moffat I, Chepelev N, Labib S, Bourdon-Lacombe JA, Kuo B, Buick J, Lemieux F, Williams A, Halappanavar S, Malik A, Luijten M, Aubrecht J, Hydake DR, Fornace AJ, Swartz CD, Yauk CL, 2015. Comparison of toxicogenomics and traditional approaches in quantitative human health risk assessment of benzo(a)pyrene in drinking water. *Crit. Rev. Toxicol* 45 (1), 44–52. [PubMed: 25605027]
- Mootha VK, Lindgren CM, Eriksson KF, Subramanian A, Sihag S, Lehar J, Puigserver P, Carlsson E, Ridderstråle M, Laurila E, Houstis N, Daly MJ, Patterson N, Mesirov JP, Golub TR, Tamayo P, Spiegelman B, Lander ES, Hirschhorn JN, Altshuler D, Groop LC, 2003. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat. Genet* 34 (3), 267–273. [PubMed: 12808457]
- NCBI, 2011. Genome. Retrieved from <<http://www.ncbi.nlm.nih.gov/genome/>>.
- Nesnow S, Ward W, Moore T, Ren H, Hester S, 2009. Discrimination of tumorigenic triazole conazoles from phenobarbital by transcriptional analyses of mouse liver gene expression. *Toxicol. Sci* 110 (1), 68–83. [PubMed: 19363144]

- Nie AY, McMillan M, Parker JB, Leone A, Bryant S, Yieh L, Bittner A, Carmen A, Wan J, Lord PG, 2006. Predictive toxicogenomics approaches reveal underlying molecular mechanisms of nongenotoxic carcinogenicity. *Mol. Carcinog* 45, 914–933. [PubMed: 16921489]
- NRC (National Research Council of the national Academies), 2007. *Toxicity Testing in the 21st Century: a Vision and Strategy*. The National Academies Press, 500 Fifth St. N.W., Washington, D.C.
- Pahl HL, 1999. Activators and target genes of Rel/NF- κ B transcription factors. *Oncogene* 18 (49), 6853–6866. [PubMed: 10602461]
- Pennie W, Pettit SD, Lord PG, 2004. Toxicogenomics in risk assessment: an overview of an HESI collaborative research program. *Environ. Health Perspect* 112(4), 417–419. [PubMed: 15033589]
- Richert L, Lamboley C, Viollon-Abadie C, Grass P, Hartmann N, Laurent S, Heyd B, Manton G, Chibout S-D, Staedtler F, 2003. Effects of clofibric acid on mRNA expression profiles in primary cultures of rat, mouse and human hepatocytes. *Toxicol. Appl. Pharmacol* 191 (2), 130–146. [PubMed: 12946649]
- Sand S, Portier CJ, Krewski D, 2011. A signal-to-noise crossover dose as the point of departure for health risk assessment. *Environ. Health Perspect* 19, 1766–1774.
- Sen B, Wang A, Hester SD, Robertson JL, Wolf DC, 2005. Gene expression profiling of responses to dimethylarsinic acid in female F344 rat urothelium. *Toxicology* 215, 214–226. [PubMed: 16122865]
- Sen B, Wolf DC, Turpaz Y, Bugrim A, Retief A, Hester SD, 2007. Identification of interspecies concordance of mechanism of arsenic-induced bladder cancer. *Toxicol. In Vitro* 21, 1513–1529. [PubMed: 17720352]
- SEQC/MAC-III Consortium, 2014. A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the Sequencing Quality Control Consortium. *Nat. Biotechnol* 32 (9), 903–914. [PubMed: 25150838]
- Shi L, 2006. The MicroArray quality control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. *Nat. Biotechnol* 24 (9), 1151–1161. [PubMed: 16964229]
- Shi L, Jones WD, Jensen RV, Harris SC, Perkins RG, Goodsaid FM, Guo L, Croner LJ, Boysen C, Fang H, Gian F, Amur S, Bao W, Barbacioru CC, Bertholet V, Cao XM, Chu T-M, Collins PJ, Fan X, Frueh FW, Fuscoe JC, Guo X, Han J, Herman D, Hong H, Kawasaki ES, Li Q-Z, Luo Y, Ma Y, Mei N, Peterson RL, Puri RK, Shippy R, Su Z, Sun YA, Sun H, Thorn B, Turpaz Y, Wang C, Wang SJ, Warrington JA, Willey JC, Wu J, Zhang L, Zhang L, Zhong S, Wolfinger RD, Tong W, 2008. The balance of reproducibility, sensitivity, and specificity of lists of differentially expressed genes in microarray studies. *BMC Bioinformatics* 9 (Suppl. 9), S10.
- Storey JD, Tibshirani R, 2003. Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. U.S.A* 100 (16), 9440–9445. [PubMed: 12883005]
- Thomas RS, Bao W, Chu T-M, Bessarabova M, Nikolskaya T, Nikolsky Y, Andersen M, Wolfinger RD, 2009. Use of short-term transcriptional profiles to assess the long-term cancer-related safety of environmental and industrial chemicals. *Toxicol. Sci* 112 (2), 311–321. [PubMed: 19776212]
- Thomas RS, Clewell HJ III, Allen BC, Wesselkamper SC, Wang NCY, Lambert JC, Hess-Wilson JK, Zhao QJ, Andersen ME, 2011. Application of transcriptional benchmark dose values in quantitative cancer and noncancer risk assessment. *Toxicol. Sci* 120 (1), 194–205. [PubMed: 21097997]
- Thomas RS, Clewell HJ, Allen BC, Yang L, Healy E, Andersen ME, 2012. Integrating pathway-based transcriptomic data into quantitative chemical risk assessment: a five chemical case study. *Mutat. Res. Genet. Toxicol. Environ. Mutagen* 746 (2), 135–143.
- Thomas RS, Philbert MA, Auerbach SS, Wetmore BA, Devito MJ, Cote I, Rowlands JC, Whelan MP, Hays SM, Andersen ME, Meek ME, Reiter LW, Lambert JC, Clewell HJ III, Stephens ML, Zhao QJ, Wesselkamper SC, Flowers L, Carney EW, Pastoor TP, Petersen DD, Yauk CL, Nong A, 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. [PubMed: 23958734]
- Thomas RS, Wesselkamper SC, Wang NCY, Zhao QJ, Petersen DD, Lambert JC, Cote I, Yang L, Healy E, Black MB, Clewell HJ III, Allen BC, Andersen ME, 2013b. Temporal concordance

between apical and transcriptional points of departure for chemical risk assessment. *Toxicol. Sci* 134 (1), 180–194. [PubMed: 23596260]

- Thompson CM, Hixon JG, Proctor DM, Haws L, Suh M, Urban JD, Makris MA, 2012. Assessment of genotoxic potential of Cr (VI) in the mouse duodenum: an in silico comparison with mutagenic and nonmutagenic carcinogens across tissues. *Regul. Toxicol. Pharmacol* 64, 68–76. [PubMed: 22705708]
- Thompson CM, Proctor DM, Suh M, Haws LC, Kirman CR, Harris MA, 2013. Assessment of mode of action underlying development of rodent small intestinal tumors following oral exposure to hexavalent chromium and relevance the humans. *Crit. Rev. Toxicol* 43 (3), 244–274. [PubMed: 23445218]
- Tibshirani R, Walther G, & Hastie T, 2000. Estimating the number of clusters in a dataset via the Gap statistic. In Technical Report 208, Department of Statistics, Stanford University.
- Tibshirani R, Hastie T, Narasimhan B, Chu G, 2002. Diagnosis of multiple cancer types by shrunken centroids of gene expression. *Proc. Natl. Acad. Sci. U.S.A* 99 (10), 6567–6572. [PubMed: 12011421]
- Tusher VG, Tibshirani R, Chu G, 2001. Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl. Acad. Sci. U.S.A* 98 (9), 5116–5121. [PubMed: 11309499]
- U.S. EPA, 2004. Acetochlor report of the cancer assessment review committee (CARC). Fourth Evaluation. Environmental Protection Agency, Office of Pesticides Programs, Washington, D.C., <<http://www.regulations.gov/fdmspublic-rell1/component/main>> (search for Docket number OPP-2005-0227).
- U.S. EPA, 2006. Revised reregistration eligibility decision document (RED) for MSMA, DSMA, CAMA, and Cacodylic Acid. U.S., Environmental Protection Agency, Washington, D.C., <http://www.epa.gov/oppsrrd1/REDs/organic_arsenicals_red.pdf>.
- U.S. EPA, 2009. An approach to using toxicogenomics data in U.S. EPA human health risk assessments: a dibutyl phthalate case study. U.S. Environmental protection Agency, National Center for Environmental Assessment, Office of Research and Development, Washington, D.C., Agency. <<http://cfpub.epa.gov/ncea/cfm/recorddisplay.cfm?deid=213405#Download>>.
- U.S. EPA, 2011. Toxicological review of trichloroethylene (CAS No 79-01-6): in support of information on the Integrated Risk Information System (IRIS). U.S. Environmental Protection Agency, Washington, D.C., EPA/635/R-09/011F, <<http://www.epa.gov/iris/toxreviews/0199tr/0199tr.pdf>>.
- U.S. EPA, 2013. TSCA chemical substance inventory: basic information. Retrieved from <http://echa.europa.eu/documents/10162/13585/clp_final_report_20110104_en.pdf>.
- U.S. EPA, 2014. Next generation risk assessment: recent advances in molecular, computational, and systems biology. U.S. Environmental Protection Agency, National Center for Environmental Assessment, Office of Research and Development, Washington, DC., EPA/600/R-14/004, <http://cfpub.epa.gov/si/si_public_record_report.cfm?dirEntryId=286690>.
- Uehara T, Minowa Y, Morikawa Y, Kondo C, Maruyama T, Kato I, Nakatsu N, Igarashi Y, Ono A, Hayashi H, Mitsumori K, Yamada H, Ohno Y, Urushidani T, 2011. Prediction model of potential hepatocarcinogenicity of rat hepatocarcinogens using a large-scale toxicogenomics database. *Toxicol. Appl. Pharmacol* 255, 297–306. [PubMed: 21784091]
- Vanden Heuvel JP, Thompson JT, Frame SR, Gillies PJ, 2006. Differential activation of nuclear receptors by perfluorinated fatty acid analogs and natural fatty acids: a comparison of human, mouse, and rat peroxisome proliferator-activated receptor- α , - β and - γ , liver X receptor- β , and retinoid X receptor- α . *Toxicol. Sci* 92 (2), 476–489. [PubMed: 16731579]
- Vaquerez JM, Kummerfeld SK, Teichmann SA, Luscombe NM, 2009. A census of human transcription factors: function, expression and evolution. *Nat. Rev. Genet* 10(4), 252–263. [PubMed: 19274049]
- Wang Z, Gerstein M, Snyder M, 2009. RNA-seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet* 10 (1), 57–63. [PubMed: 19015660]
- Waters MD, Jackson M, Lea I, 2010. Characterizing and predicting carcinogenicity and mode of action using conventional and toxicogenomics methods. *Mutat. Res* 705 (3), 184–200. [PubMed: 20399889]

- Welle S, 2013. What statisticians should know about microarray gene expression technology. *Methods Mol. Biol* 972, 1–13. [PubMed: 23385528]
- Wilson VS, Keshava N, Hester S, Segal D, Chiu W, Thompson CM, Euling SY, 2013. Utilizing toxicogenomic data to understand chemical mechanism of action in risk assessment. *Toxicol. Appl. Pharmacol* 271, 299–308. [PubMed: 21295051]
- Wingender E, Chen X, Fricke E, Geffers R, Hehl R, Liebich I, Krull M, Matys V, Michael H, Ohnhausser R, Pruss M, Schacherer F, Thiele S, Urbach S, 2001. *Nucl. Acids Res* 29, 281–283. [PubMed: 11125113]
- Yao CQ, Prokopec SD, Watson JD, Pang R, Png C, Chong LC, Harding NJ, Pohjanvirta R, Okey A, Boustros PC, 2012. Inter-strain heterogeneity in rat hepatic transcriptomic responses to 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD). *Toxicol. Appl. Pharmacol* 260, 135–145.
- Yauk CL, Berndt ML, 2007. Review of the literature examining the correlation among DNA microarray technologies. *Environ. Mol. Mutagen* 48 (5), 380–394. [PubMed: 17370338]
- Yauk CL, Williams A, Buick JK, Chen G, Maertens RM, Halappanavar S, White P, 2012. Genetic toxicology and toxicogenomic analysis of three cigarette smoke condensates in vitro reveals few differences among full-flavor, blonde, and light products. *Environ. Mol. Mutagen* 53 (4), 281–296. [PubMed: 22431010]
- Zaccaria KJ, McClure PR, 2013. Using immunotoxicity information to improve cancer risk assessment for polycyclic aromatic hydrocarbon mixtures. *Int. J. Toxicol* 32 (4), 236–250. [PubMed: 23922326]
- Zhang Y, Szustakowski J, Schinke M, 2009. Bioinformatics analysis of microarray data. *Methods Mol. Biol* 573, 259–289. 10.1007/978-1-60327-337-4_1. [PubMed: 19763933]
- Zhang L, Zhang J, Yang G, Wu D, Jiang L, Wen Z, Li M, 2013. Investigating the concordance of gene ontology terms reveals the intra- and inter-platform reproducibility of enrichment analysis. *BMC Bioinformatics* 14, 143. [PubMed: 23627640]

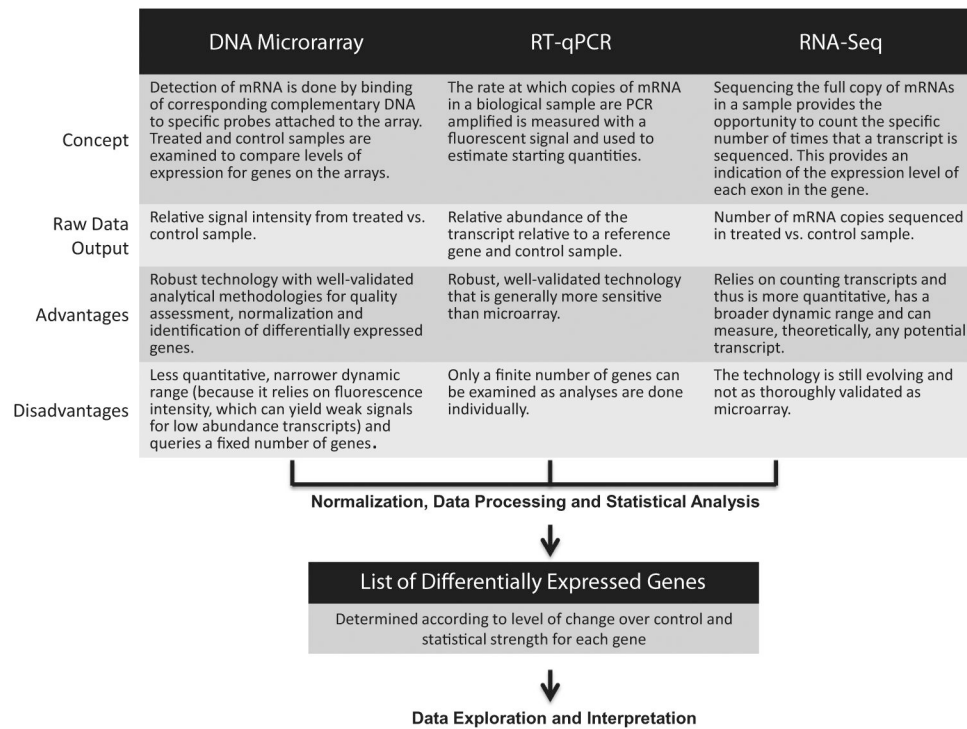


Fig. 1. Brief overview of technologies used for gene expression profiling.

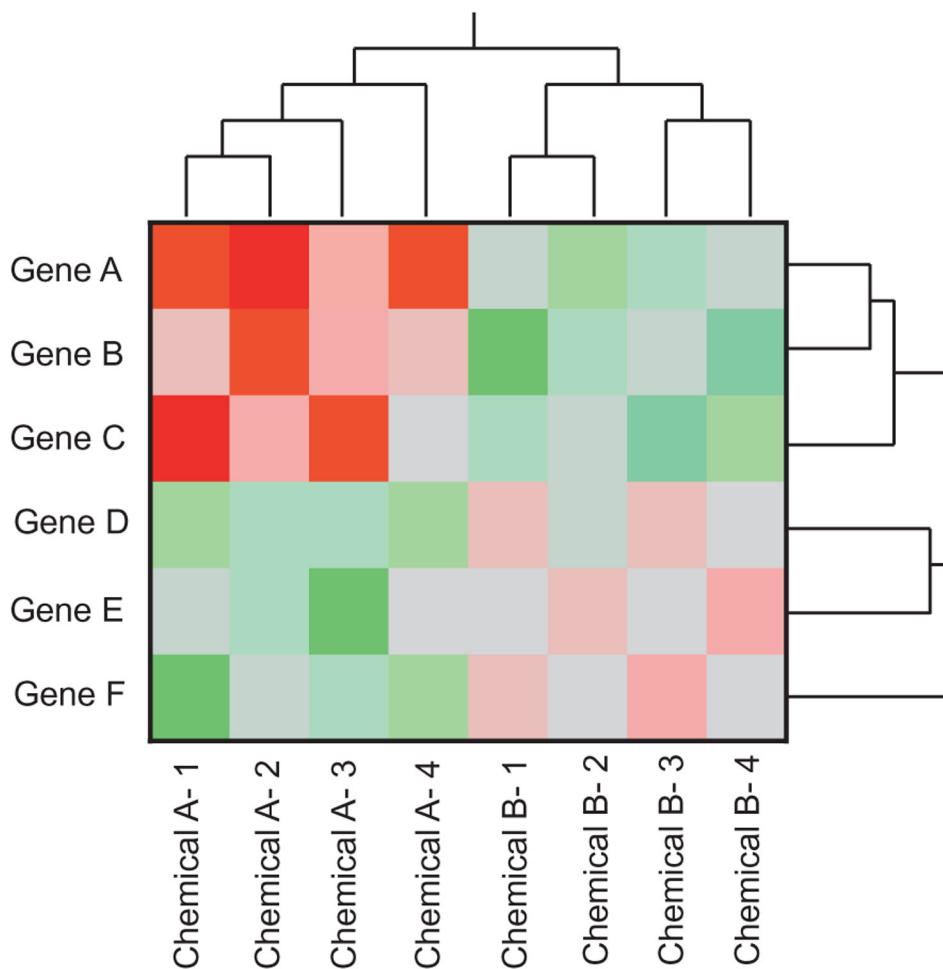


Fig. 2. Schematic representation of a typical heat map with hierarchical clustering. Red, green and gray demonstrates increased regulation, decreased regulation and no-change, respectively. The intensity of the color refers to greater fold-changes. In this example, chemical A increases the expression of genes A, B and C, whereas genes D, E and F are down-regulated. Chemical B down-regulates genes A, B, C and causes modest up-regulation in genes D, E and F. The clustering on the *X*-axis demonstrates how closely the samples correlate with each other. Clustering on the *Y*-axis represents subsets of genes with similar functions (not always depicted).

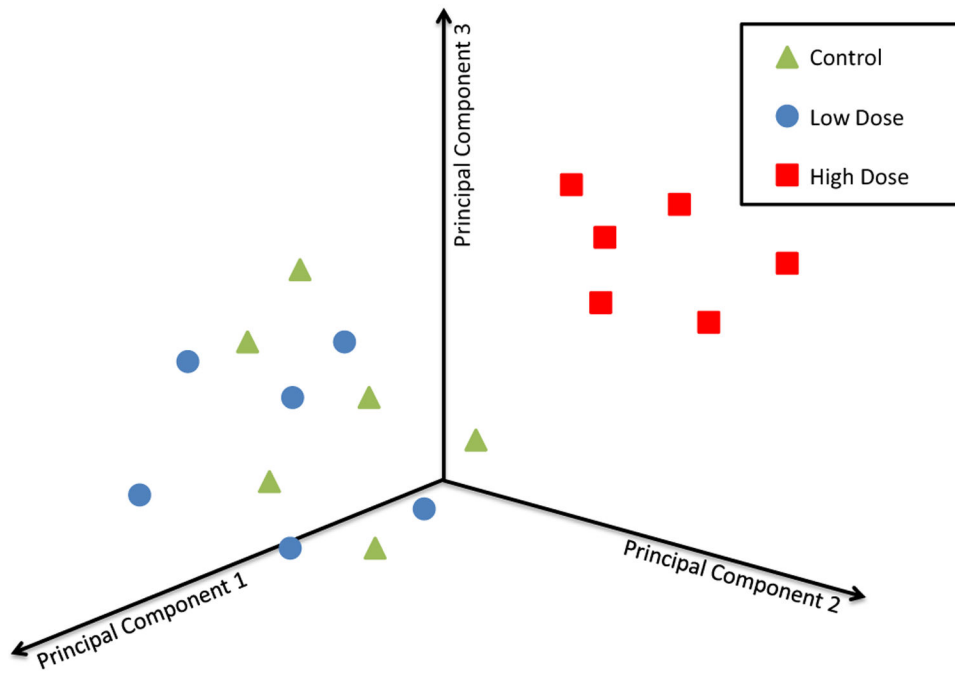


Fig. 3. Schematic representation of a 3D PCA plot. This PCA diagram depicts treatment effects at the high dose, but not at the low dose. PC1 should capture more than PC2 and both components should total to >55% and the top three >70%.

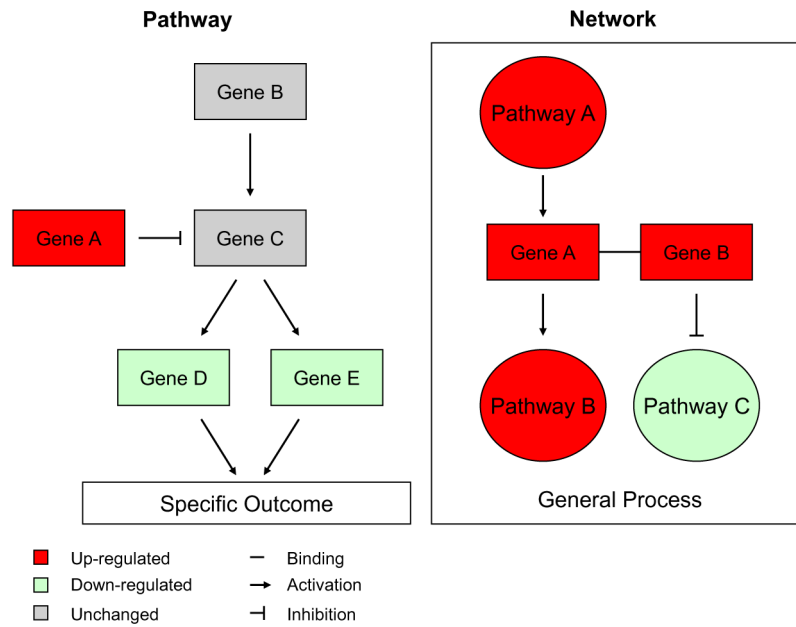


Fig. 4. Schematic representation of biological pathways and networks. The pathway represents molecular interactions leading to a specific biological outcome, whereas the network refers to broader biological interactions representing the overall state of the biological system.

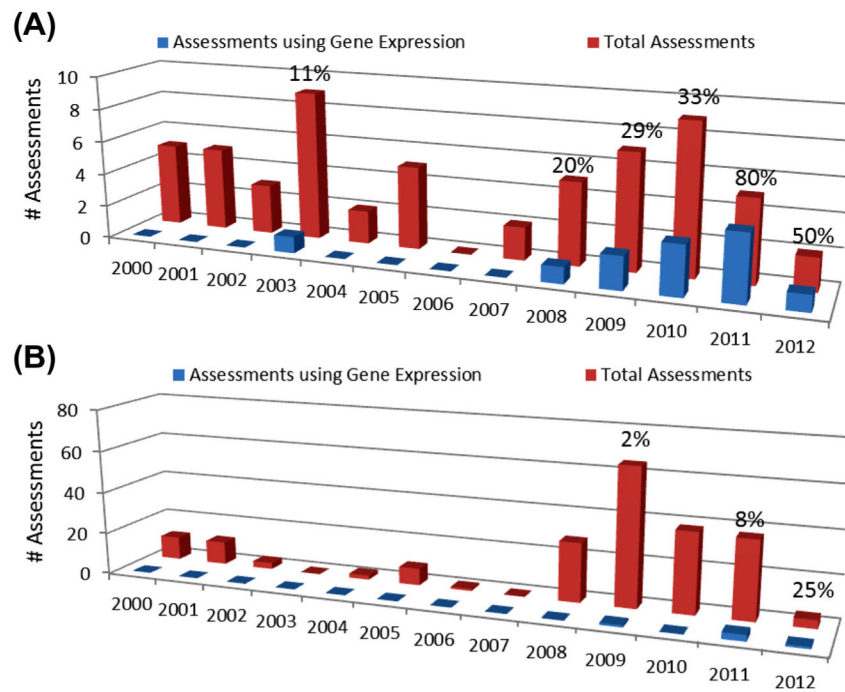


Fig. 5. Number of EPA IRIS (Panel A) and Health Canada ESRAB (Panel B) assessments containing gene expression data (blue) compared to total assessments (red) for 2000–2012. Percentages indicate the proportion of assessments containing gene expression information. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

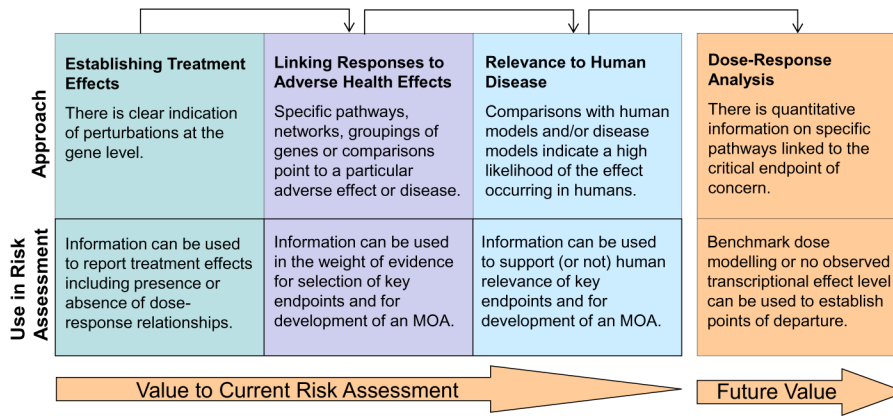


Fig. 6. Utility of toxicogenomics data in risk assessment today and in the future.

Table 1
 Criteria determined to be mandatory (*) or important in evaluating the overall quality of toxicogenomics experiments.

	In vitro	In vivo animal	In vivo human
	<ul style="list-style-type: none"> Cytotoxicity was assessed, and at least some of the concentrations are below those inducing overt toxicity* Unexposed control cells were cultured at the same time as the treated cells using identical cell culture procedures* A minimum of three experimental replicates (plates) were used (in order to reach desired power)* The appropriate cell type was used, and there is a rationale for the chemical concentrations, exposure duration and harvest time selected If dose-response data is considered for risk modeling (including estimation of the BMD), a minimum of three doses plus control was used Tests to assess various toxicities (e.g., markers of genotoxicity or oxidative stress) were done using the same biological samples 	<ul style="list-style-type: none"> Control animals were handled alongside treated animals using identical procedures (e.g., controls in oral gavage experiments received vehicle only) and at similar times.* A minimum of three biological replicates (animals) were used per group (in order to reach desired power)* If temporality is considered, time-matched controls were used* The appropriate animal model and tissue was used, and there is a rationale for the doses selected If dose-response is considered for risk modeling (including estimation of the BMD), a minimum of three doses plus control was used Ideally, at least one of these doses should be near the NOAEL Tests to assess various toxicities (e.g., histopathology, biomarkers of disease) were done using the same biological samples 	<ul style="list-style-type: none"> Appropriate control and exposed populations were identified* A sufficiently large sample size was used to ensure adequate power* The exposure scenarios considered do not involve appreciable exposures to additional chemicals* The level of exposure was addressed and duration of the exposure is known The data take into consideration potential confounding effects* There is a rationale for the tissue sample selected for dosimetry (blood, urine, other) If dose-response data is considered for modeling (including estimation of the BMD), the data were divided into ordered categories of exposure, or continuous exposure data were available for individual study participants

Table 2

Criteria that are required (*) or should be considered in DNA microarray methodologies.

-
- RNA A_{260}/A_{280} ratios are reported and are above 1.8 to indicate sample purity, or are consistent across samples
 - The integrity of RNA was assessed (common strategies include an RNA integrity number (RIN), an RNA quality indicator (RQI) or 28s:18s ratio) to ensure minimal RNA degradation or consistency across samples
 - When multiple microarrays are necessary and the experiment was run over different days, the samples were randomized across the slides/days to avoid confounding effects (often referred to as a block design) Note: not always specified in the methods
 - Generally, gene annotation and data quality are more robust when commercially produced microarray platforms are used
 - Species appropriate microarrays were used (i.e., mouse arrays for mouse samples)
 - Labeling and hybridization were done according to manufacturer protocol. Any deviations are reported
 - When co-hybridizations of treated and control samples are done (use of different fluorophores for control and treated samples), dye-swapping experiments were done, or there is an indication that dye bias was assessed statistically
 - Scanner specific quality control software was used to test microarray quality
 - Data quality was assessed (through MA plots, heat maps, boxplots, scatterplots, signal to noise ratio, etc.)
 - In the case that outliers are identified, there is a minimum of three replicates remaining per group and a justification for removal has been provided
 - The data were preprocessed (e.g., background subtracted and log transformed) and normalized (i.e., adjusted to remove technical variations between arrays) prior to statistical analysis*
 - An appropriate statistical analysis of data was conducted to identify differentially expressed genes*
 - Data were adjusted to account for false positives (most commonly referred to as false discovery rate (FDR) adjustment or Benjamini–Hochberg method). Note: this is not always done, though a lower p-value may be set to minimize false positives, with more focus given to pathway/functional effects over individual genes
 - Software versions, parameters, and gene annotation/references versions and builds were recorded
 - The number of genes considered significant was restricted based on fold-change induction (typically not less than 1.5-fold)
 - Gene significance was restricted on the basis of some measure of statistical significance (not more than 10%)
 - Validation of results was assessed using one of the following approaches: (1) the predicted biological effect was verified directly in the test animals or through the literature; (2) changes in important genes were confirmed using an alternative method such as real-time quantitative PCR; (3) other measures were carried out to confirm the predicted response
 - Data files were made available through an open access public database such as Gene Expression Omnibus (GEO), Chemical Effects in Biological Systems (CEBS) or ArrayExpress
-

Table 3

Criteria that are required (*) or should be considered in real-time quantitative PCR methodologies.

<ul style="list-style-type: none"> • RNA A_{260}/A_{280} ratios are reported and are above 1.8 to indicate sample purity, or are consistent across samples • The integrity of RNA was assessed (common strategies include an RNA integrity number (RIN), an RNA quality indicator (RQI) or 28s:18s ratio) to ensure minimal RNA degradation or consistency across samples • Reference genes (also referred to as housekeeping genes) were used to normalize the data* • There is evidence that reference gene expression was not affected by treatment* • It is ensured that the primer used is specific to the gene of interest (common strategies include the use of melting curves, gel electrophoresis or sequencing)* • PCR arrays generally contain pre-validated primers that are gene-specific • No reverse transcription (NRT) controls were included to test for genomic DNA contamination of RNA samples • Technical replicates were performed • The number of cycles required to detect a true signal should be below 35* • PCR efficiencies calculated using the slope of the standard curve are reported

Table 4

Criteria that are required (*) or should be considered in RNA sequencing methodologies.

-
- DNAase treatment was applied to total RNA samples during extraction (not necessary for polyA selection procedures)
 - RNA A_{260}/A_{280} ratios were reported and were above 1.8 to indicate sample purity, or are consistent across samples
 - The integrity of RNA was assessed (common strategies include an RNA integrity number (RIN), an RNA quality indicator (RQI) or 28s:18s ratio) to ensure minimal RNA degradation or consistency across samples
 - Sample preparation (e.g. RNA enrichment, fragmentation of RNA and library building) was done according to manufacturer protocol. Any deviations were reported
 - When multiple sequencing runs were necessary and the experiments were run over different days, the samples (most likely barcoded and pooled) were randomized across the lanes (Illumina) and/or chips (Ion Torrent) to avoid confounding effects (often referred to as a block design). Note: this is not always specified in the methods
 - There was sufficient sequence depth (i.e., enough sequencing reads per sample were collected to fairly quantify the amount that a gene was expressed)
 - Sequencer-specific quality control software was used to test the sequencing run
 - There is an indication that data quality was assessed
 - In the case that outliers were identified, there was a minimum of three replicates remaining per group and a justification for removal was provided
 - The data were normalized (e.g., adjusted to remove technical variability) prior to statistical analysis*
 - A statistical analysis for count data was done to identify differentially expressed genes.*
 - The number of genes considered significant was restricted on the basis of fold-change induction (typically not less than 1.5-fold increase or decrease in expression) and some measure of statistical significance (not more than 10%)
 - Data were adjusted to account for false positives (most commonly referred to as false discovery rate (FDR) adjustment or Benjamini–Hochberg method). Note: this is not always done and is considered acceptable, though a lower p -value may be set to minimize false positives, with more focus given to pathway/functional effects over individual genes
 - Validation of results was assessed using one of the following approaches: (1) the predicted biological effect was verified directly in the test animals or through the literature; (2) changes in important genes were confirmed using an alternative method such as real-time quantitative PCR; (3) specific proteins corresponding to gene expression changes were quantified; (4) other measures were carried out to confirm the predicted response
 - Data files were made available through an open access public database such as Sequence Read Archive (SRA) or ArrayExpress
-

Table 5

Utility of toxicogenomics approaches within the current risk assessment paradigm.

Approach	Value
Number of differentially expressed genes	General overview of the presence and extent of toxicogenomic response. Comparisons between similar chemicals can be used to draw conclusions on many factors, including potency as well as species-, strain- or tissue-specific susceptibilities
Heat maps and clustering/PCA	Identification of treatment effects, important genes and likely biological outcomes. Comparisons with gene expression profiles for well documented chemicals can be used to elucidate MoA
Gene functionality/pathways and Networks	Identification of major responses and likely biological outcomes that can be used in the weight of evidence. Insight in chemical MoA
Upstream regulators	Identification of plausible molecular initiating events, key events and insight into chemical MoA
Comparisons of work in animals and human cells	Provides evidence that responses associated with adverse effects in experimental animals are relevant to human responses. Can be used to strengthen the rationale for selection of a critical effect for human health risk assessment
Comparisons with publicly available profiles	Provides insight into MoA, potential pathologies and relevance to humans by comparison of gene expression profiles in animals with those of other chemicals, disease models or humans. Can be used to elucidate MoA or to strengthen the rationale for selection of the critical effect
Dose-response analysis	Identification of treatment effects. No-observable-transcriptional-effect level (NOTEL) or BMD can be used to assess modified Bradford-Hill criteria (to verify order of events) if a full MoA analysis framework is considered

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript