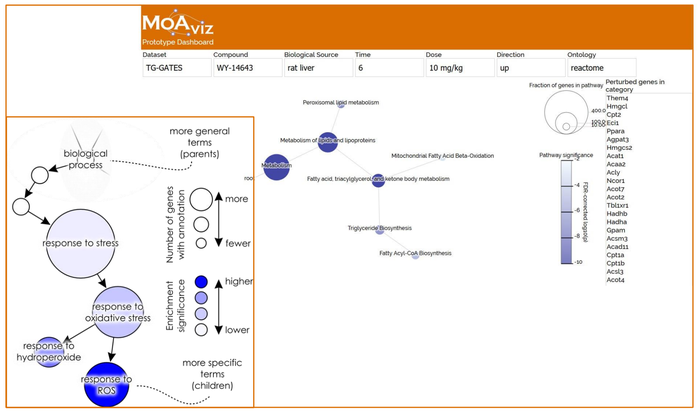
Integration of transcriptomic point of departure metrics into the MoAviz visualization framework

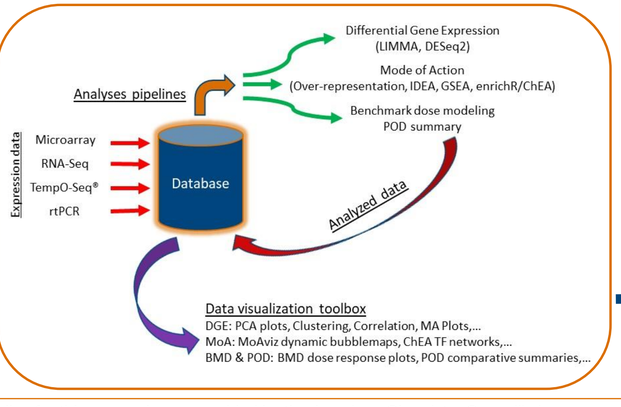
M. B. Black[1], K. Bronson [1], S. N. Pendse [1], J. Fitzpatrick [1], R. A. Clewell[2], M. E. Andersen [1], P. D. McMullen [1]
[1] ScitoVation LLC, Durham, NC 27713 & [2] 21st Century Tox Consulting, Chapel Hill, NC, USA
For more information on this project or other ScitoVation research initiatives, please email science@scitovation.com

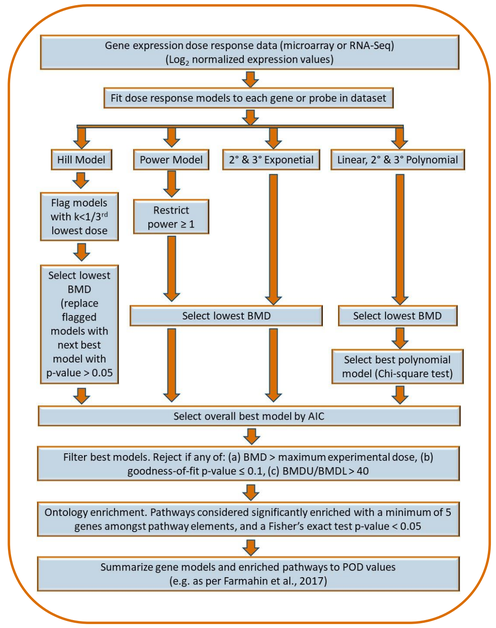
Abstract. Gene expression profiling is emerging as a viable way to evaluate mode of action and points of departure and has the potential to drastically reduce testing costs and product development time. The first steps of using transcriptional responses as a basis of safety assessment are being taken. However, several important considerations remain unresolved, including how to translate expression changes into adverse outcome pathways or other definitions of mode of action, and the best manner with which to summarize gene expression data into a point of departure. Recently we developed an interactive browser application, MoAviz, to facilitate the examination of gene expression data across dose and time for mode of action studies of chemical perturbation for 204 compounds (spanning 290 million gene expression change values). We used MoAviz to quantitatively compare pathway-level transcriptomic signatures across compounds with well-known modes of action, and across different model systems, providing the groundwork for performing "biological read-across" between compounds based on their transcriptomic fingerprints. We evaluated the extent to which gene expression changes from in-life exposures could be associated with mode of action by developing a novel similarity index—the Modified Jaccard Index (MJI)—that provides a quantitative description of genomic pathway similarity. While typical compound-compound similarity is low (MJI = 0.026), clustering of the TG-GATES compounds identifies groups of similar compounds. Some clusters aggregated compounds with known similar modes of action, including PPARα agonists (MJI = 0.330) and NSAIDs (MJI = 0.327). We continue to extend the MoAviz interface and database by incorporating whole transcriptome benchmark dose analyses and point of departure (POD) summary, including the command line modeling features of the BMDExpress2 software. This integration will include statistical pre-filtering of transcriptomic gene expression data, dose response modeling of individual genes, ontology over-representation of genes, and POD summary based on current proposed best practices for gene-based and pathway-based derivation of POD. By combining mode of action and POD tools in an interactive interface, MoAviz will facilitate the use of transcriptomics data over a variety of chemical safety contexts.



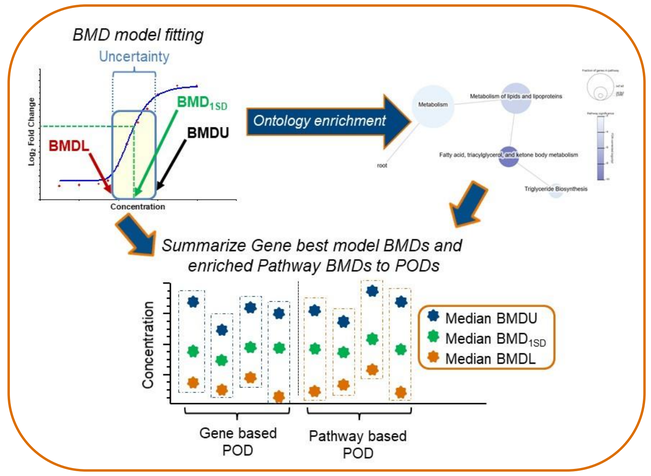
In 2019, we presented a visualization framework to facilitate the interpretation of differential gene expression and ontology enrichment for the determination of cellular modes of action (MoA) [1]. This tool presents ontology over-representation enrichment patterns in the form of interactive bubblemaps, allowing for interactive study of patterns of enrichment across dose and time. Statistical significance of enrichment is displayed by shading, and the number of differentially expressed genes found amongst a pathway's elements is indicated by the size of nodes in the bubblemap. Edges in the bubblemap figure retain the hierarchical structure of the ontology. Individual genes resulting in significant enrichment of an ontology category can be listed so that discrete genes of interest may be identified.

In 2020 we are expanding this tool to include transcriptomic benchmark dose model fitting, pathway enrichment and methods to summarize gene based BMD values and pathway based BMD values to single values for use as points of departure (POD) [7]. Adding this capability to a MoA visualization framework will facilitate interpretation of transcriptomic PODs for use in evaluating compounds in risk assessment and regulatory decision making by providing biological functional relevance to POD derivations.





The BMD pipeline process begins with normalized, Log_2 transformed gene or probe expression values. These data are used in BMDExpress2 command line analyses to fit dose response models to individual microarray probes or genes from RNA-Seq experiments, followed by statistical analyses of fitted dose response models for the selection of a single overall best fit model for each probe or gene [5,6,7].

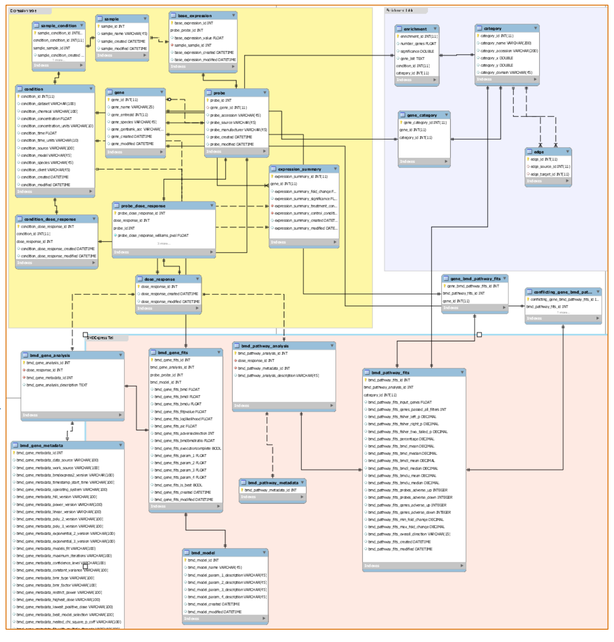


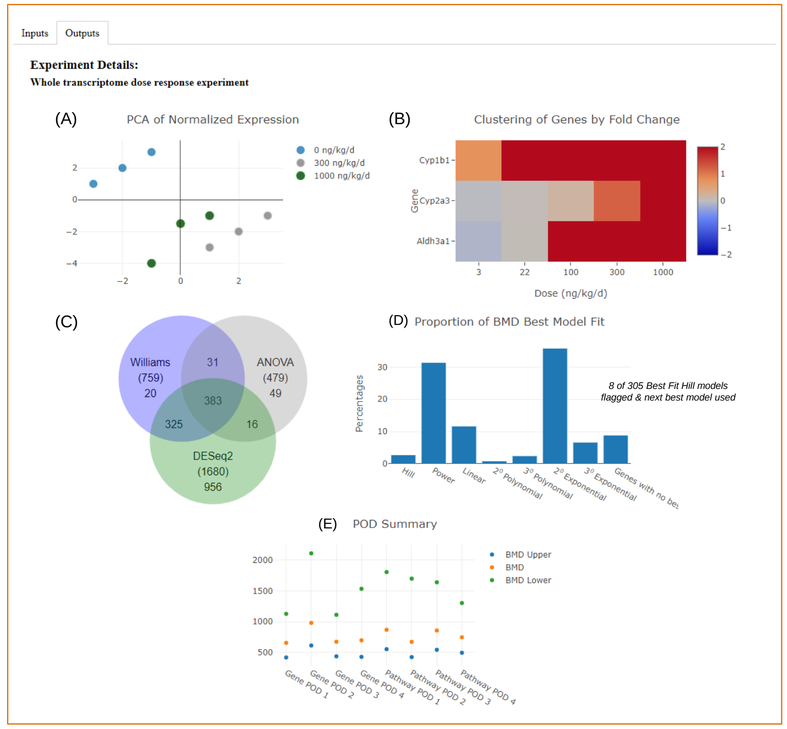
Best fit models meeting QC metrics for acceptable model fit are used in a conventional ontology over representation analyses to define enriched cellular functional pathways. Values for pathway BMDU, BMD and BMDL values are then based on the median of the values for the genes that result in significant enrichment of a pathway. POD values are then derived using both individual gene based summary methods (e.g. 20 genes with lowest BMD), and pathway summaries (e.g. 20 pathways with lowest BMD) [3].

Dynamic visuals and report summaries

Data & analyses captured for comparison across compounds, species or cell type, cellular MoA pathways or POD

MoAviz is built around a database that captures transcriptomic data (e.g., RMA normalized expression, RNA-Seq count tables) providing the inputs for both differential gene expression, and BMD model fitting and POD summary. The BMD model fits, pathway enrichment and POD values are captured in the database to provide inputs for a visualization toolkit to display bubblemaps, BMD summary data and comparative POD plots. This database also allows for re-analyses functions in the event ontology information is updated, or alternative model fitting or POD summary algorithms are to be explored. And it provides a means for comparative analyses of different dose response experiments.





Output Panel Mockup. The figure is a conceptualized mockup of our output panel using in-house data from Affymetrix microarray and RNA-Seq experiments. Our visualization toolbox presents QC data analyses at various steps along the analysis pipeline as well as summary data and results. We use standard statistical tools like Principle component analyses (A) to assess samples and overall experimental results. Clustering tools and heat maps show the pattern of differential expression changes across the dose response (B). Determination of signal in the dataset (C), as a pre-requisite to BMD model fitting, includes the tools built into BMDExpress (ANOVA, Williams trend test) as well as the results of our detailed differential gene expression analyses (using LIMMA for microarray data, and DESeq2 for RNA-Seq count data). Our proposed standard BMD model fitting uses 7 models and the summary of overall best model fits is displayed post analyses as a bar graph (D). Finally, we present POD summaries derived from gene based and pathway based summary algorithms in a comparative plot (E) using proposed best practices.

Funding for this research was provided by the American Chemistry Council's Long-range Research Initiative (ACC-LRI).

References
1 McMullen, P.D., et al., Toxicol In Vitro, 2019. **58**: p. 1-12
2 *NTP Research Report on NTP Approach to Genomic Dose-Response Modeling*, 2018, p. 1-44.
3 Farmahin, R., et al., Arch Toxicol, 2017. **91**(5): p. 2045-2065
4 Thomas, R.S., et al., Toxicol Sci, 2011. 120(1): p. 194-205.
5 Thomas, R.S., et al., Mutat Res, 2012. **746**(2): p. 135-43.
6 Yang, L., B.C. Allen, and R.S. Thomas, BMC Genomics, 2007. **8**: p. 387.
7 Phillips, J.R., et al., Bioinformatics, 2019. **35**(10): p. 1780-1782.1