Mode of action from dose-response microarray data: case study using 10 environmental chemicals

J.R. Pirone[1], D.J. Dix[1], M. DeVito[2], and I. Shah[1]

[1]National Center for Computational Toxicology, Office of Research and Development, USEPA, RTP, NC, USA.
[2]National Health and Environmental Effects Lab, Office of Research and Development, US EPA, RTP, NC, USA.

I. ABSTRACT

Ligand-activated nuclear receptors regulate many biological processes through complex interactions with biological macromolecules. Certain xenobiotics alter nuclear receptor signaling through direct or indirect interactions. Defining the mode of action (MOA) of such xenobiotics is difficult due to the many perturbations in cellular signaling networks resulting from exposure. Microarray data, when collected in a dose-response setting, is a rich source of information for determining the MOA. Analysis presents several challenges: namely, it is difficult to choose a single quantitative model that can be applied to each gene. We have utilized a method for analyzing microarray dose-response data that is flexible, while still capable of capturing the complexity of the responses. Rat primary hepatocytes were incubated with solutions of 0, 10, 30, and 100 µM of 4-nonylphenol, mono-(2-ethylhexyl) phthalate (MEHP), myclobutanil, propiconazole, or triadimefon or 1, 3, and 1 µM of DE-71, PCB-118, or PCB-153. In all cases, the control consisted of the 1% DMSO dosing solution. RNA was extracted 72 hrs after dosing and analyzed using Affymetrix Rat Genome 230 2.0 arrays. Natural cubic splines were fit to the dose-response data for each gene. Statistical significance of each fit was assessed using bootstrap analysis. A set of curve shape and intensity features was extracted from the fit for each gene found to exhibit a significant dose-response. Features included first-derivatives, maximum/minimum expression, dose at maximum/minimum expression, and area under the curve. Using subsets of curve features, genes were first partitioned into groups that responded with a similar shape then into groups with similar expression intensities. Gene groups were subjected to functional and pathway analysis.

II. INTRODUCTION

Dose-response analysis is fundamental to the practice of toxicology. Measuring the response of an organism or bioassay to increasing doses of a xenobiotic is essential for determining a quantitative measure of toxicity. Additionally, the shape of the dose-response curve may be informative of the underlying biology.

In traditional toxicity experiments, the main challenge of designing a dose-response study is choosing doses such that the entirety of the dynamic range of the response is observed. Once the set of doses has been established, a mathematical model (typically nonlinear) is fit to the data and parameters relevant for risk assessment are extracted from the fitted model.

With microarray data it is difficult to find a suitable set of doses that span the dynamic range of each of the thousands of genes on the array. Furthermore, the dose-response to the xenobiotic will differ for each gene, so it is impossible to decide upon a single, or even a small subset of nonlinear models for fitting to the data. The large number of genes also precludes sorting genes on the basis of a similar dose-response shape by visual means. It is necessary to use methods that can:

1. Fit a flexible model to all of the observed dose-response profiles,
2. assess the significance of the fit, and
3. group the genes according to similarities in the shape of their dose-response curves.

III. METHODS

Our fitting method is adapted from Storey *et al* (2005), where natural cubic splines are used to fit the expression data and the significance of the fits is assessed using bootstrap analysis. We have built upon this method by using partitioning around medoids (PAM) to group the genes into groups based on similarities in certain curve shape features extracted from the spline fits to the dose-response curve. Functional analysis was performed using the genes within each shape partition for each chemical. All analyses were carried out using R/Bioconductor (Gentleman *et al*., 2004).

1. Within each chemical, fit a natural cubic spline to the dose-response data for each gene.
2. Compute a "modified F-statistic" for the fit:

$$F_i = \frac{SS_i^0 - SS_i^1}{SS_i^1}$$

3. Compute the residuals for the fit. Bootstrap the residuals and add back into a null model.
4. Fit spline to the new data.

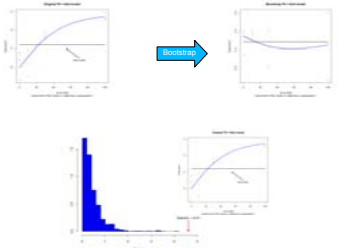

Figure 1: Schematic of method for spline fitting and assessment of statistical significance. Histogram shows the values obtained for the modified F-statistic for 200 bootstrap iterations of the dose-response data for a single gene.

5. Compute the modified F-statistic for the fit to the new data.
6. Repeat a specified number of times.
7. For each gene identified as significant, compute a collection of curve summary measures:
   I. First derivatives at each dose
   II. Dose at expression minimum and maximum
   III. The fitted values at each dose
   IV. The area under the curve (AUC)
8. Partition the genes in to shape groupings on the basis of the values of subsets of the curve summary measures.
9. For the genes in each partition, determine overrepresented GO biological process categories using the hypergeometric test as implemented in the R package GOstats.

IV. RESULTS

The dose-response profiles were partitioned into 4 groups using PAM. Partitioning was based on the values of the derivatives and the doses at the maximum and minimum expressions. A *1-correlation* distance metric was used for the partitioning. Figure 2 shows the average behavior of all (scaled) dose-response curves within each partition. Groups 1 and 2 are made up of genes with monotonically increasing and decreasing dose-response profiles respectively. Groups 3 and 4 are comprised of genes with non-monotonically increasing or decreasing dose-response profiles respectively. Approximately 36.8% of the significant genes were in Group 1, 41.7% in Group 2, 10.7% in Group 3, and 10.8% in Group 4.

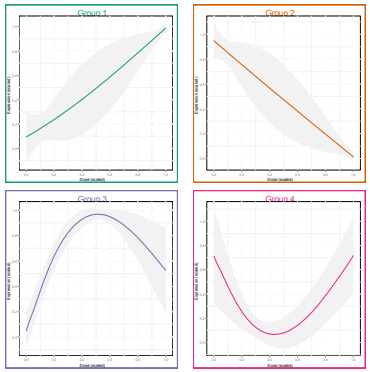

Figure 2: The curves represent the average dose-response behavior for the genes partitioned into that group. The expressions and doses were scaled to lie between 0 and 1. The grey shaded areas cover the mean ± standard deviation.

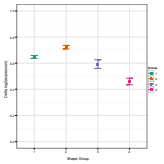

Figure 3 presents the mean dynamic range (maximum fitted expression – minimum fitted expression) for each shape group. The error bars represent the standard errors of the mean dynamic range for each shape group.

Figure 3: The points represent the mean value of the range (max-min) expression values for each gene in each partition as calculated from the fitted curve. The error bars represent ± standard error.

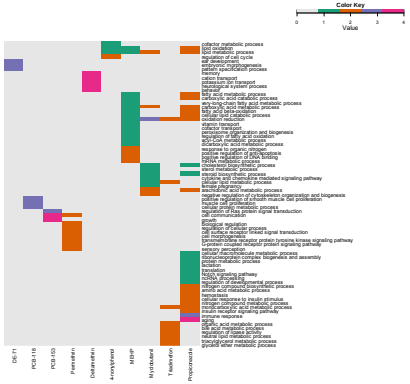

Figure 4: Selection of overrepresented GO biological processes categories by chemical. Color indicates the shape group for genes in the category. Light grey indicates that the GOBP category was not associated with a particular chemical.

For each chemical, the genes in each shape group were tested for GO term association. Figure 4 shows a subset of the significant GO biological processes (GOBP). The shape group of the genes associated with each process is denoted by color. The overrepresented processes are consistent with some of the known modes of action for several of the tested chemicals. For example, many genes that exhibit a significant dose-response to MEHP fall into group 1 (monotonically increasing) and are associated with lipid and fatty acid metabolism. As MEHP is known to act as a PPAR-α agonist, these results are consistent with the established MOA for MEHP (Rusyn, *et al*., 2006). The conazoles triadimefon, myclobutanil, and propiconazole have been shown to impact genes involved in sterol, cholesterol, and arachidonic acid metabolism (Goetz and Dix, 2009). Using our method, GO BP categories consistent with this observation were identified for these chemicals.

Figure 5 shows the shape groupings for the subset of genes exhibiting a significant dose-response that are common to all three conazoles. For the most part, genes that are common to all conazoles are partitioned into the same shape groups. In most cases, the common genes were partitioned into groups 1 and 2. However, some genes were partitioned into group 3 and this partitioning was similar between chemicals. This provides partial evidence that the non-monotonic dose-response profiles identified using our method represent real biological phenomena.

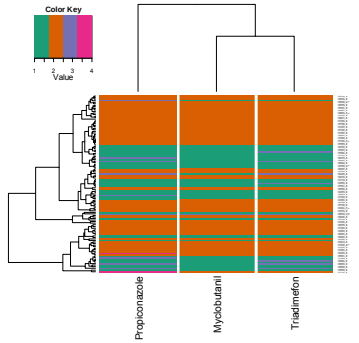

Figure 5: Shape partitions for 114 genes that exhibited a significant dose-response in all three conazoles (propiconazole, myclobutanil, and triadimefon). The shape group for each gene is indicated by color. Dendrograms are based on the distances between genes or chemicals based on the quantitative values of the extracted curve shape features.

V. DISCUSSION

While this approach may be overly simplistic – in most cases genes with multiple dose-response profiles will contribute to a particular MOA – the method shows that genes can be grouped into biologically meaningful categories on the basis of similarities in the shapes of their dose-response curves. Similarity in shape may be strongly related to similarities in promoter architecture. We are currently conducting an analysis of the promoters of the genes within the shape groups to look for overrepresented modules of transcription factor binding sites.

Given the huge number of genes on an array and the complexity of gene regulation, it is not surprising that genes are found that have non-monotonic dose-response profiles. However, this does not suggest that any toxic effects are similarly non-monotonic.

VI. REFERENCES

Gentleman, R., Carey V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., *et al.* (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biology*, **5**, R80.

Goetz, A. & Dix, D.J. (2009) Toxicogeneomic effects common to triazole antifungals and conserved across species. In preparation

Rusyn, I., Peters, J.M., & Cunningham, M.L. (2006) Modes of action and species-specific effects of di-(2-ethylhexyl)phthalate in the liver. *Crit. Rev. Toxicol.* **33**, 655-780.

Storey, J.D., Xiao, W. Leek, J.T., Tompkins, R.G., & Davis, R.W. (2005) Significance analysis of time course microarray experiments. *Proc. Natl. Acad. Sci. USA* **102**, 12837-12842.