Regulatory Toxicology and Pharmacology 67 (2013) 63–74

Contents lists available at ScienceDirect

Regulatory Toxicology and Pharmacology

journal homepage: www.elsevier.com/locate/yrtph

Gene batteries and synexpression groups applied in a multivariate statistical approach to dose–response analysis of toxicogenomic data ☆

C. Parfett [a,*], A. Williams [b], J.L. Zheng [a], G. Zhou [a]

[a] Genetic Toxicology Laboratory, Mechanistic Studies Division, Environmental Health Science and Research Bureau, Environmental and Radiation Health Sciences Directorate, Healthy Environments and Consumer Safety Branch, Health Canada, Canada

[b] Biostatistics Section, Population Studies Division, Environmental Health Science and Research Bureau, Environmental and Radiation Health Sciences Directorate, Healthy Environments and Consumer Safety Branch, Health Canada, Canada

ARTICLE INFO

Article history:
Received 2 May 2013
Available online 4 July 2013

Keywords:
Risk assessment
Point of departure
Stress-response pathways
Signal response
Adverse outcome

ABSTRACT

Univariate statistical analyses have limited strength when employed in low-dose toxicogenomic studies, due to diminished magnitudes and frequencies of gene expression responses, compounded by high data dimensionality. Analysis using co-regulated gene sets and a multivariate statistical test based upon ranks of expression were explored as means to improve statistical confidence and biological insight at low-doses. Sixteen gene regulatory groups were selected in order to investigate murine hepatic gene expression changes following low-dose oral exposure to the beta-adrenergic agonist, isoproterenol (IPR). Gene sets in this focussed analysis included well-defined gene batteries and synexpression groups with co-regulated responses to toxin exposures and linkage of chronic responses to adverse outcomes. Significant changes of target gene expression within Nfkb, Stat3 and 5′ terminal oligopyrimidine (5′TOP) batteries, as well as the acute phase and angiogenesis synexpression groups, were detected at IPR doses 100-fold lower than doses producing significant individual gene expression values. IPR-induced changes in these target gene groups were confirmed using a similar analysis of rat toxicogenomic data from published IPR-induced cardiotoxicity studies. Cumulative expression differences within gene sets were useful as aggregated metrics for benchmark dose calculations. The results supported the conclusion that toxicologically-relevant, co-regulated genes provide an effective means to reduce microarray dimensionality, thereby providing substantial statistical and interpretive power for quantitative analysis of low-dose, toxin-induced gene expression changes.

Introduction

Toxicogenomic gene expression profiling has been explored as an adjunct methodology for dose–response characterizations of chemical toxicity, with the expectation that it could offer a means to improve this core element of quantitative human health risk assessment. An attractive feature of toxicogenomic profiling resides in its potential for more extensive insights into toxicological modes-of-action (MOA) as compared to traditional toxicological endpoints. Also, gene expression changes are thought to be potentially more sensitive measures of effects at early time points and at lower doses than many apical toxicological measures. Such expectations have generally been met in a range of studies. Gene expression biomarkers have been used to describe dose–response characteristics resulting from the action of estrogen receptor ligands (estrogen, and xenoestrogens bisphenol A and genistein) (Naciff et al., 2005), and in comparing toxicity equivalence factors (TEFs) for Ah-receptor agonists (TCDD, PCB126, TCDF) (Kopec et al., 2010). Other studies have determined threshold doses for MOA-related measures of gene expression in molecular pathways perturbed by the toxin formaldehyde (Andersen et al., 2010), or drugs such as peroxisome proliferator activated receptor agonist fenofibrate, histamine H1 receptor antagonist, methapyrilene (Bercu et al., 2010), beta-adrenergic receptor agonist isoproterenol (Zheng et al., 2011), as well as various tyrosine kinase inhibitors (Ji et al., 2009). Therefore, identification of statistically significant thresholds of toxicological effect, affirmed by toxicogenomic analyses, have offered opportunities to directly apply toxicogenomic information in assigning benchmark doses as points of departure for calculating exposure levels that would impose minimal excess risk for human health (Bercu et al., 2010; Thomas et al., 2007, 2011).

* Corresponding author. Address: Environmental Health Centre 0803A, 50 Columbine Driveway, Tunney's Pasture, Ottawa, K1A 0K9 Ont., Canada. Fax: +1 613 941 8530.

E-mail address: Craig.Parfett@hc-sc.gc.ca (C. Parfett).

0273-2300/$ - see front matter © 2013 The Authors. Published by Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.yrtph.2013.06.005

In dose–response experiments, statistical analyses of toxicogenomic data have primarily employed per gene analysis, using univariate methods with appropriate false discovery corrections. The power to detect individual, differentially expressed genes is diminished by the high dimensionality of the data sets, in which the variables (genes on the microarray) are very large compared to the number observations (experimental samples). Proper experimental design (Wei et al., 2004), pre-processing, normalization and filtering (Bourgon et al., 2010; Hackstadt and Hess, 2009; McClintick and Edenberg, 2006; Yauk et al., 2006) prior to statistical analysis and/or using statistics that borrows strength across genes (Baldi and Long, 2001; Cui et al., 2005; Jain et al., 2003; Smyth, 2004; Tong and Wang, 2007; Wright and Simon, 2003)can increase the ability to detect differentially expressed genes, while still controlling the false discovery rate.

Further gains in statistical sensitivity have been derived from the knowledge that genes never act alone. The fact that genes interact with each other and are normally expressed in functionally relevant patterns suggests that gene-expression data can be stratified and clustered into relatively homogenous groups (Barry et al., 2005; Leung and Cavalieri, 2003; Lu et al., 2005; Qin et al., 2008; Tuglus and Van Der Laan, 2009). Genes within these groups are often correlated and dimensionality reduction into groups of biologically relevant genes may improve screening power while minimizing information loss to detect meaningful genes and or expression profiles (Lu et al., 2005; Qin et al., 2008). Gene sets, either through pathways, or modules (Engreitz et al., 2010; Huang et al., 2007; Rahnenführer et al., 2004; Segal et al., 2003b,c,a) are a natural means to reduce the dimensionality of microarray data while gaining increased interpretation (Alvo et al., 2010; Beltrame et al., 2009; Pavlidis et al., 2002; Rahnenführer et al., 2004). Using multivariate techniques to utilize the information hidden in gene interactions can provide more powerful, biologically meaningful and reproducible results (Beltrame et al., 2009; Maglietta et al., 2010; Stiglic et al., 2010; Zhao et al., 2009) for finding subsets of differentially expressed genes which may not be detectable when using univariate approaches (Engreitz et al., 2010; Lu et al., 2005; Szabo et al., 2003; Zahn et al., 2006; Mar et al., 2011).

Established gene batteries and synexpression groups are known, a priori, as co-expressed gene sets controlled by specific molecular mechanisms or more complex combinations of controls, respectively. Dimensional reduction of toxicogenomic data into these sets should provide a direct, mechanistic approach with high statistical power to detect toxin-induced differential gene expression across exposure levels. Gene sets with direct relevance to toxicological effects include the single transcription-factor controlled gene batteries that respond in major cellular stress response pathways, including responses to oxidative stress, heat shock, DNA damage, hypoxia, unfolded protein accumulation in the endplasmic reticulum, metal stress, osmotic stress, and inducers of inflammation (Jennings et al., 2013; Simmons et al., 2009). The ligand-activated nuclear receptor/transcription factors control specific batteries of drug and xenobiotic metabolic enzymes, catalysing detoxification or toxic activation of xenobiotics as well as conversion of endogenous substrates that may be important for tissue homeostasis (Tirona and Kim, 2005; Woods et al., 2007). Additionally, endocrine disrupting agents may perturb the functioning of receptor/transcription factors for endocrine hormones, thereby altering normal expression patterns of the respective target gene batteries (Tabb and Blumberg, 2006).

Synexpression groups are regulated by more complex, signal-response pathways and are highly coordinately expressed (across tissues, time, or conditions) under the control of common sets of signals and second messengers, usually in order to achieve a biological outcome (Niehrs and Pollet, 1999). It is thought that genes within these groups share compound cis- and trans-acting control elements to achieve coordinated gene expression in response to various signals (Davidson, 2010; Nachman and Regev, 2009). Developmental programs rely on spatio-temporal synexpression of gene groups while certain tissue and cellular activities such as the co-ordination and execution of the cell cycle or production of the acute-phase response of the liver (as two examples) are regulated transcriptionally in response to diverse biological and metabolic inputs, including stress and toxicity. Less defined groupings of genes discovered by clustering algorithms in genome-wide RNA profiling experiments, often labelled as synexpression groups, were not included in this analysis because in many cases there is little evidence to define the underlying control mechanisms or toxicological significance of the expression changes.

Post-transcriptional mechanisms may also define gene groups with co-ordinated regulation in response to cellular stresses. Transcript levels are controlled at multiple post-transcriptional levels by RNA-binding proteins and miRNAs (Keene, 2007) and at the translational level (entry or exit from polysomes), which affects transcript stability. For example, the latter mechanism is known in mRNAs with 5′ terminal oligopyrimidine (5′TOP) sequence motifs, as a response to growth arrest (Balagopal and Parker, 2009; Meyuhas, 2000).

Monitoring gene expression changes coordinated through defined molecular events in stress-response pathways could provide an integrated measure of outcome due to effects of toxins on these complex signalling/response networks. Secondly, such coordinated changes give insights into resultant adaptive, maladaptive, or toxic consequences for cellular and tissue function. This report demonstrates a multivariate gene-set approach for statistical analysis of experimental dose–response data obtained for a selected panel of gene batteries and synexpression groups of toxicological concern. The data are converted to aggregated gene-set responses and applied in a benchmark dose assessment.

Materials and methods

Animal treatment and microarray analysis

Male B6C3F1 mice (age 27–35 days; Charles River Laboratories) were housed in individual cages under a 12 h light/12 h dark lighting schedule and were acclimatized for 2 weeks prior to the start of the study. Mice were randomly assigned to treatment and control groups (five animals per group) and were given a single dose of 250, 50, 5 and 0.5 mg/kg/day IPR via oral gavage and sacrificed 8 h after exposure. The control group received equivalent volumes of 0.9% saline. Liver was removed and total liver RNA was isolated and used for the microarray hybridization on ToxArray™ chip v 1.2 (containing 1624 genes, 4 replicate probes per gene). Arrays were scanned on a ScanArray Express (Perkin Elmer Life Sciences, Woodbridge, ON, Canada). The images were quantified using ImaGene 5.6 (BioDiscovery, Los Angeles, CA), and median signal intensities (not background subtracted) were used for analyses. The complete microarray data are publicly available (Zheng et al., 2011) and via http://www.ebi.ac.uk/arrayexpress/, accession number: E-MEXP-2716.

Composition of gene battteries and synexpression groups for analysis

We employed pre-determined, and experimentally annotated gene sets assembled from several sources: (1) TfactS (Essaghir et al., 2010), for Stat3, Nfkb, Upr, Srebp1, and Creb1; (2) Gene Ontology classifications for the angiogenesis synexpression group (a combination of GO: 0045765, regulation of angiogenesis, and GO: 0001525, angiogenesis), KEGG pathways for the PPAR gene battery; and (4) published literature for the acute phase synexpression