Environmental and Molecular Mutagenesis 56:505–519 (2015)

Research Article

Development of a Toxicogenomics Signature for Genotoxicity Using a Dose-Optimization and Informatics Strategy in Human Cells

Heng-Hong Li,[1,2] Daniel R. Hyduke,[1,2,3] Renxiang Chen,[1,2] Pamela Heard,[4] Carole L. Yauk,[5] Jiri Aubrecht,[4] and Albert J. Fornace Jr.[1,2,6]*

[1]*Department of Biochemistry and Molecular and Cellular Biology, Georgetown University Medical Center, Washington, DC*
[2]*Department of Oncology, Georgetown University Medical Center, Washington, DC*
[3]*Biological Engineering Department, Utah State University, Logan, Utah*
[4]*Pfizer Global Research and Development, Drug Safety Research and Development, Groton, Connecticut 06340*
[5]*Environmental Health Science and Research Bureau, Health Canada, Ottawa, Ontario, K1A 0K9, Canada*
[6]*Center of Excellence In Genomic Medicine Research (CEGMR), King Abdulaziz University, Jeddah,Saudi Arabia*

The development of in vitro molecular biomarkers to accurately predict toxicological effects has become a priority to advance testing strategies for human health risk assessment. The application of in vitro transcriptomic biomarkers promises increased throughput as well as a reduction in animal use. However, the existing protocols for predictive transcriptional signatures do not establish appropriate guidelines for dose selection or account for the fact that toxic agents may have pleiotropic effects. Therefore, comparison of transcriptome profiles across agents and studies has been difficult. Here we present a dataset of transcriptional profiles for TK6 cells exposed to a battery of well-characterized genotoxic and nongenotoxic chemicals. The experimental conditions applied a new dose optimization protocol that was based on evaluating expression changes in several well-characterized stress-response genes using quantitative real-time PCR in preliminary dose-finding studies. The subsequent microarray-based transcriptomic analyses at the optimized dose revealed responses to the test chemicals that were typically complex, often exhibiting substantial overlap in the transcriptional responses between a variety of the agents making analysis challenging. Using the nearest shrunken centroids method we identified a panel of 65 genes that could accurately classify toxicants as genotoxic or nongenotoxic. To validate the 65-gene panel as a genomic biomarker of genotoxicity, the gene expression profiles of an additional three well-characterized model agents were analyzed and a case study demonstrating the practical application of this genomic biomarker-based approach in risk assessment was performed to demonstrate its utility in genotoxicity risk assessment. Environ. Mol. Mutagen. 56:505–519, 2015.

Key words: stress genes; DNA-damage inducible; super-paramagnetic clustering; bioinformatics

Additional Supporting Information may be found in the online version of this article.

Grant sponsor: NIH; Grant number: NIEHS 1U01ES016013; NIEHS R01ES020750; NIAAA R01AA018863.
Grant sponsor: HiCi 1432/6-6, Deanship of Scientific Research (DSR), King Abdulaziz University.

*Correspondence to: Albert J. Fornace, Jr., Lombardi Comprehensive Cancer Center, 3970 Reservoir Rd. NW, Room E504, Washington, DC 20057. E-mail: af294@georgetown.edu

Received 18 July 2014; provisionally accepted 19 January 2015; and in final form 00 Month 2015

DOI 10.1002/em.21941

Published online 2 March 2015 in Wiley Online Library (wileyonlinelibrary.com).

INTRODUCTION

There is increasing interest in using omics technologies to guide the development of biomarker panels that can indicate whether a compound will elicit a specific toxic response, such as genotoxicity [Newton et al., 2004; Li et al., 2007; Guyton et al., 2009; Cui and Paules, 2010; Goodsaid et al., 2010; Hartung and McBride, 2011; Wilson et al., 2013]. DNA microarray technologies can be used to measure transcriptomic perturbations mediated by a particular agent, followed by bioinformatics analyses to identify subsets of genes that are characteristic of the induced response caused by the agent. A variety of adverse stimuli, both genotoxic and nongenotoxic in nature, can trigger complex transcriptional responses (i.e., stress response pathways, Supporting Information Fig. S1). Transcriptomic profiles are attractive as biomarkers because gene expression technologies are mature and transcript annotation is well defined as a result of the human genome project. Furthermore, there is considerable information on regulatory events controlling mRNA expression, such that major efforts have been initiated to elucidate pathways of toxicity (PoT) [Hartung and McBride, 2011]. Transcriptome-based approaches have been successful in research applications targeted towards classifying a variety of biologic and pathologic states. After measuring the toxicant-induced transcriptomic responses, bioinformatics approaches are used to identify a subset of features that may be used to classify the potential for toxicity of unknown compounds.

Gene-based biomarker panels have been explored for classifying compounds according to whether they will elicit a specific toxic response. Typically, the biomarker panels consisting of small gene sets are identified via statistical algorithms based on the ability of the panel to discriminate between treated and untreated samples. This approach rarely incorporates biological knowledge during the gene selection process. Nevertheless, toxic agents often display pleiotropic (Supporting Information Fig. S1) effects, inducing multiple stress response pathways [Li et al., 2007; Wilson et al., 2013]. For example, methylmethane sulfonate (MMS) damages DNA and results in the induction of a DNA damage response, in parallel with inducing the unfolded protein response and other endoplasmic reticulum (ER) associated signaling events [Hyduke et al., 2011]. In addition, gene panels that also provide mechanistic insights (i.e., mechanism-based genomic biomarkers) should be better suited to deciphering the complex array of induced stress response pathways, and help delineate those associated with genotoxicity.

Pathway analysis tools are used in conjunction with biomarker panels to develop hypotheses for the underlying regulatory events and response pathways perturbed by a specific chemical. Unfortunately, these methods often generate multiple hypotheses that are not easily testable given the promiscuity of genes. Integration of multiple omics types, including delineation of coordinately regulated transcriptomic responses, may aid in classifying the particular stress response pathways that are invoked following chemical exposure, providing insight into underlying mechanisms of action and PoT [Hartung and McBride, 2011; Jennen et al., 2011; Magkoufopoulou et al., 2011; Wilson et al., 2013].

Previously, we and others have investigated transcriptome perturbations to differentiate among toxicant modes of action and classify the toxicity potential of new compounds [Amundson et al., 2005; van Delft et al., 2005; Lamb et al., 2006]. This has also been the subject of many recent reviews [Li et al., 2007; Cui and Paules, 2010; Hartung and McBride, 2011]. However, in these examples the dosing of cells with tested chemicals was either selected in an ad hoc fashion or was based on the results of cytotoxicity tests. For example, in Amundson et al. (2005), a partial-genome survey using a 7,668 element cDNA array in TK6 cells and its p53-null derivative NH32 was performed to assess 13 known stress agents. In this study, as well as many others, no systematic effort was made to optimize dose selection. Without dose optimization, there is a risk that the transcriptomic response will be either negligible due to under-dosing or obscured by apoptotic and other general responses due to overdosing. In addition, alkylating agents and other protein-damaging agents can disrupt the transcriptional machinery and actually attenuate the transcriptional response at very high doses [Fornace et al., 1989b; Li et al., 2007]. Furthermore, cellular responses to stress detected via mRNA expression changes are also time dependent. The response includes many immediate-early genes and other genes whose transcripts accumulate within a few hours after exposure to genotoxic [Fornace et al., 1989b; Amundson et al., 2005; Ellinger-Ziegelbauer et al., 2009; Hyduke et al., 2011] and nongenotoxic agents (e.g., heat shock) [Fornace et al., 1989a]. The early induced stress response can lead to a cascade of events that cause later cytotoxicity and associated pathway changes [Amundson et al., 2005] that show a similar profile across variety of toxic mechanisms [Ellinger-Ziegelbauer et al., 2009]. In fact, the lack of the appropriate dose-setting metrics and time of exposure cause major problems with interpretation of toxicogenomic studies. Thus, it is critical to establish a standardized dose-setting paradigm in order to ensure that comparisons can be made across chemicals and studies.

In the present study we utilized an in vitro transcriptomics-based approach to develop biomarker gene sets applicable to the assessment of genotoxicity. Our approach measured transcriptomic perturbations in the human lymphoblastoid-derived TK6 cell line because it is p53 proficient, has been well characterized, has been extensively used in toxicologic studies, and provided robust responsiveness in previous stress signaling studies

[Akerman et al., 2004; Islaih et al., 2004; Amundson et al., 2005]. We developed our model using a diverse set of 28 model agents representing DNA-reactive agents that are known to be directly genotoxic, indirect-acting agents causing DNA damage either by inhibition of topoisomerase action or blockage of DNA synthesis (categorized as 'genotoxic'), and non DNA-reactive agents that are negative in genotoxicity tests (categorized as "nongenotoxic"). We note that aneugens that operate via interaction with spindle were classified as nongenotoxic for these purposes.

Complex cellular stress responses following treatment with chemicals are time- and dose-dependent. However, toxicogenomic studies evaluating both time course and dose responses across a large set of compounds are not economically feasible. Therefore, the experimental design necessitated selecting a single dose and time point post-exposure for sample collection. Since early gene expression changes have been shown to be indicative of initial damage and not to be influenced by subsequent molecular processes such as apoptosis [Ellinger-Ziegelbauer et al., 2009], a four hour post-exposure time point was selected for the development of the gene set in the present study [Amundson et al., 2005; Hyduke et al., 2011]. To enable comparison of transcriptome profiles across the whole set of agents at a single dose per chemical, and to establish a strategy for setting doses for new test compounds, we developed a qRT-PCR molecular phenotyping protocol based on expression changes in three known stress response genes in the dose setting experiments; these genes included *ATF3*, *GADD45A* and *CDKN1A* (also referred to as p21, Cip1, or Waf1). Previous studies have demonstrated that these genes exhibit strong responses to a wide variety of agents [Fan et al., 2002; Amundson et al., 2005; Hyduke et al., 2011]. Both *GADD45A* and *CDKN1A* are known to be p53-regulated, and show robust mRNA induction by genotoxic agents, while we have found that *ATF3* is responsive to a wide variety of stress agents in TK6 cells [Amundson et al., 2005].

Using the above strategy, a database of gene expression profiles was derived from a single concentration and time point for each of the chemicals. Within this database there were multiple overlapping responses and a clustering approach was used to identify subsets of genes showing coordinated responses. Many of these subsets were defined by agents sharing known mechanisms of action. This database was also employed to construct an mRNA gene signature that discriminates between genotoxic and nongenotoxic agents. To evaluate the utility of the gene expression signature as a genomic biomarker of genotoxicity, we tested the signature on three model agents: 3-nitropropionic acid (3-NP), isopropyl methanesulfonate (iPMS), and tri-methylxanthine (caffeine). Finally, additional validation exercises were undertaken, and are presented in a companion paper [Buick et al., 2015], to demonstrate that the gene set is also able to accurately classify chemicals as genotoxic or nongenotoxic in TK6 cells in the presence of a metabolic activation system (rat liver S9).

MATERIAL AND METHODS

Cell Culture and Treatments

TK6 cells were grown in RPMI 1640 medium supplemented with 10% fetal bovine serum. The cell culture density was maintained at 1–10 $\times$ 10^5 cells/ml, and was tested periodically for mycoplasma contamination; all mycoplasma results were negative. Exponentially growing cells were treated at a density of 4–5 $\times$ 10^5 cells/ml. All chemicals were purchased from Sigma-Aldrich, except cadmium chloride and potassium chromate, which were purchased from J.T. Baker. The solvents used to prepare the stock solutions are listed in Table I. A stock solution of each chemical was added to exponentially growing cells and incubated at 37°C for 4 hr. An equal volume of the corresponding solvent for each chemical was used to treat cells simultaneously, and labeled as the vehicle control for each experiment. Dose response studies were performed using a range of concentrations for all testing agents except γ-ray exposure, heat shock and ethanol treatments (Table I). The concentration of each chemical that was used to treat cells for microarray analysis was determined based on the dose-response study and is listed in Table I. For γ-ray exposure, cells were irradiated at 0.87 Gy/min to a dose of 4 Gy using a γ-ray irradiator. After the irradiation, cells were incubated at 37°C, in 5% CO_2 for 4 hr. Heat-shock treatment was performed by submerging 35 ml cultures in a T175 flask in a 47°C circulating water bath with gentle agitation for 20 min, while the control was held in the same condition but at 37°C; this severe heat shock treatment has previously been shown to trigger strong responses at the mRNA level [Fornace et al., 1989a]. After 20 min of heat shock, cells were maintained in a 37°C, 5% CO_2 incubator for 4 hr. At the end of each of the 4 h treatments, cells were pelleted by centrifugation and subjected to RNA extraction.

Cell Viability Assays

The trypan blue exclusion [Amundson et al., 2005] and MTT assays (Cayman Chemical, Ann Arbor, MI) were performed after 4 and 24 hr treatments respectively to measure cell viability.

RNA Isolation and Quantitative RT-PCR

RNA was extracted using the Trizol protocol (Invitrogen) followed by RNA cleanup with an RNeasy column (Qiagen). Purified RNA was subjected to spectrometry and Bioanalyzer (Agilent) analysis to determine the quantity and quality, respectively. qRT-PCR was performed with an iCycler (Bio-Rad) to measure the expression of individual genes including GADD45A, CDKN1A, and ATF3. The assay was done in triplicate. Expression levels of genes were normalized with GAPDH for antimitotic agents, histone deacetylase (HDAC) inhibitors, and heavy metals, or 18sRNA for the other agents. Both *GAPDH* and 18sRNA are standard internal control genes and the results showed that they can be interchangeable (Supporting Information Fig. S2). The relative mRNA induction fold change was calculated. The expression alteration with fold change more than 1.7 and p-value less than 0.05 was considered significant. iScript one-step RT-PCR kit for probes (Bio-Rad, Hercules, CA) was used and primers and probe mix of Taqman gene expression assay were purchased from ABI (Foster city, CA). Dose-response studies were performed for most agents to determine the dose for the microarray experiments.

TABLE I. Standard Agents Used in This Study

Categories	Compound names	Solvent	Dose range	Conc. for array
Alkylating agents	cisplatin	0.9% NaCl	10−80 μM	80 μM
	methyl methane sulfonate (MMS)	H_2O	20−200 μg/ml	100 μg/ml
Topoisomerase I inhibitors	camptothecin	DMSO	62.5−500 nM	125 nM
Topoisomerase II inhibitors	etoposide	DMSO	50−400 nM	200 nM
RNA/DNA antimetabolites	5-fluorouracil (5-FU)	DMSO	6.25−50 μg/ml	25 μg/ml
	methotrexate	DMSO	0.05−1 mM	100 μM
DNA antimetabolites	arabinofuranosyl cytidine (AraC)	H_2O	12.5−50 μM	50 μM
	hydroxyurea	H_2O	0.25−mM	0.5 mM
causing DNA strand break by other mechanisms	γ-rays	N/A*	4 Gy	4 Gy
	bleomycin	H_2O	5−40 μg/ml	10 μg/ml
	hydrogen peroxide	N/A*	20−80 μM	80 μM
antimitotic agents	colchicine	Ethanol	62.5−1000	250 ng/ml
	docetaxel	DMSO	25−100 nM	50 nM
	paclitaxel	DMSO	12.5−200 nM	50 nM
	vinblastin	DMSO	50−800 ng/ml	200 ng/ml
histone modification inhibitors	trichostatin A (TSA)	DMSO	5−80 ng/ml	20 ng/ml
	apicidin	DMSO	0.25−4 μg/ml	1 μg/ml
	HC toxin	Methanol	5−80 ng/ml	20 ng/ml
	oxamflatin	DMSO	0.25−4 μM	1 μM
endoplasmic reticulum modulator	tunicamysin	Methanol	1.25−10 μg/ml	2.5 μg/ml
	thapsigargin	Ethanol	62.5−500 nM	250 nM
glycolysis inhibitor	2-deoxy-D-glucose (2-DG)	H_2O	0.16−20 μM	20 μM
energy metabolism inhibitor (uncoupling agent)	antimycin A	Ethanol	25−200 μM	100 μM
heavy metals	cadmium chloride	H_2O	50−800 μM	50 μM
	potassium chromate (VI)	H_2O	25−400 μM	100 μM
	sodium arsenite	H_2O	10−90 μM	30 μM
other stresses	heat shock	N/A*	47°C	47°C
	ethanol	N/A*	2%, 4%	2%, 4%

N/A, not applicable.

Gene Expression Microarrays

Control and exposed RNA for each of the 28 model agents were labeled with SuperScript indirect cDNA labeling system (Invitrogen), which does not involve an amplification step, and hybridized on Agilent human whole genome 44K oligo microarrays that contain probes for more than 41,000 unique human genes and transcripts. Hybridization and washing was performed according to the manufacturer's protocol. Arrays were scanned with an Agilent DNA microarray scanner. Feature Extraction (Version 9.1; Agilent) was used to filter, normalize, and calculate the signal intensity and ratios. Processed data were loaded onto the Rosetta Resolver for data warehousing. All the microarray data were deposited in NCBI's Gene Expression Omnibus and are accessible through GEO Series accession number GSE58431 (Available at: http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc= GSE58431).

Bioinformatic Analyses

Gene expression data were exported from Rosetta Resolver based on Entrez Gene Identifiers. To identify an mRNA signature indicative of whether an agent is genotoxic or not (either directly or indirectly), we applied the nearest shrunken centroids method (Tibshirani et al., 2002) to our transcriptome data, as implemented in the pamr package in the R statistical environment (http://cran.r-project.org). Briefly, the standardized centroid computed for each class within a training set, where the standardized centroid is the mean expression level for each gene in a class divided by its within-class standard deviation. The standard centroid for each class is shrunken toward the overall centroid to produce the nearest shrunken centroid. The method employs a shrinkage parameter that is used to control the number of features used to construct the classifier.

To identify an appropriate shrinkage parameter, we employed 10-fold cross validation [Hastie et al., 2001]. The set of samples was divided into 10 approximately equal-sized parts that were each roughly balanced for the two classes. We then assessed classification accuracy for each part using the other 9 parts to construct a classifier. With a shrinkage threshold of 2.2, we were able to identify a 65-gene panel, hereafter referred to as TGx-28.65 (28 refers to the use of 28 chemicals in the training set), with 100% accuracy based on cross-validation. With smaller gene panels, methotrexate was misclassified as nongenotoxic.

Stable gene subclusters were identified using the coupled two-way clustering (CTWC) biclustering approach [Getz et al., 2000] with the superparamagnetic clustering (SPC) algorithm [Domany, 1999]. Briefly, the expression levels for the set of 1628 genes that were significantly ($P < 0.01$) perturbed (1.7-fold or more) by at least one agent were subjected to CTWC/SPC. The default parameters for the CTWC software (http://ctwc.weizmann.ac.il/) were used for a gene depth of 5 and sample depth of 1.

Pathway Analysis

Gene signatures indicative of classic genotoxicity, heavy metals, heat shock, and nongenotoxicity were identified using two-dimensional clustering. Unique gene identifiers and the associated $\log_{10}$ expression ratios were uploaded to Ingenuity Pathway Analysis version 5.5 to identify gene networks that were significantly over-represented in the different data sets.

RESULTS

Description of Toxicants Used in the Study

We examined transcriptional changes in TK6 cells following exposure to a diverse set of model agents representing DNA alkylating agents (cisplatin, MMS), DNA strand breaking agents (γ-rays, bleomycin, hydrogen peroxide), topoisomerase inhibitors (camptothecin, etoposide), nucleotide antimetabolites [cytosine arabinoside (AraC), fluorouracil (5-FU), hydroxyurea, and methotrexate], ER stress agents (tunicamysin, thapsigargin), energy metabolism inhibitors [2-deoxy-D-glucose (2-DG), antimycin A], HDAC inhibitors [trichostatin A (TSA), apicidin, HC toxin, oxamflatin], microtubule inhibitors (colchicine, docetaxel, paclitaxel, vinblastin), and heavy metals (cadmium chloride, potassium chromate (VI), sodium arsenite) (Table I). Heat shock and ethanol were also included as additional representatives of nongenotoxic stresses. The mechanisms of action for these agents are well-established and were derived from the literature (Table I). Genotoxic agents were considered to be those causing direct DNA damage or those leading to DNA perturbations indirectly. Direct acting genotoxic agents in this group include alkylating agents and other DNA strand breaking agents. The indirectly-acting agents are topoisomerase I and II inhibitors, and antimetabolites interfering with DNA synthesis: some of the latter agents also affect RNA synthesis. These indirect-acting genotoxicants typically induce strand breaks indirectly by blocking resealing of transient topoisomerase strand breaks or by accumulation of gaps and single strand DNA during DNA synthesis. The heavy metal group has diverse mechanisms of action, including DNA damage (e.g. DNA-protein crosslinks) for chromate [Fornace et al., 1981]. Finally, microtubule inhibitors were considered nongenotoxic because they cause aneugenicity through spindle interference, rather than DNA damage and mutations.

Dose and Treatment Time Parameters

To identify an appropriate dose for each agent, we assessed mRNA levels of three well-characterized stress response genes, *ATF3*, *GADD45A*, and *CDKN1A*, which serve as indicators for an effective transcriptional response after toxin exposure. While we recognize that RT-PCR can measure significant increases that are less than 2-fold, we arbitrarily used a threshold of 2-fold to indicate robust induction. As a representative example of our results, the qRT-PCR results for bleomycin showed an expected dose dependent increase in *ATF3*, *GADD45A*, and *CDKN1A* transcript levels (Fig. 1A). The full molecular phenotyping results for the agents are summarized in Figure 1B. In general, treatment with genotoxic agents led to the induction of all three genes, whereas the nongenotoxic treatments primarily induced *ATF3* and *GADD45A* (Fig. 1B). The doses that showed the highest induction of one or more of these genes for each agent were selected for full transcriptomic analysis, after considering measures of cytotoxicity. The doses selected showed no appreciable cytotoxicity at 4 h (Supporting Information Fig. S3) and only moderate effects on viability at 24 h (Supporting Information Fig. S3). When more than one dose met the criteria and behaved similarly, the lower concentration was selected. For example, all doses of bleomycin showed increases in relative mRNA levels for all three genes (several fold increases or more), but the magnitude of *GADD45A* induction decreased slightly at doses greater than 10 μg/ml. Since all doses showed strong responses for the indicator genes, the 10 μg/ml dose was chosen for microarray analysis. All model agents, except for paclitaxel, showed at least a 2-fold increase in one or more of the three-gene battery. In the case of paclitaxel, induction was less than1.5-fold so a biologically effective dose from previous in vitro studies using cultured cells was selected [Fallo et al., 1996].

Global Transcriptional Analysis

After selecting doses that elicited robust expression of genes in our qRT-PCR battery, we measured global transcriptomic response using Agilent human whole genome oligonucleotide arrays. To reduce the effects associated with different labeling efficiencies, we used a two-channel dye-swapping configuration [Patterson et al., 2006] where the vehicle control was used as a reference sample. Visualization of the expression profiles highlighted the pleiotropic nature of these agents [i.e., gene expression profiles consists of more than one defined gene cluster (Fig. 2)]. In the case of genotoxic agents, hierarchical clustering grouped all 13 toxicants together on the right side of this heatmap. Distinct subsets of genes showed coordinate expression, such as those in the region designated by numerous red sidebars. The nongenotoxic agents represent diverse toxicants but those with similar mechanisms of action showed other subsets of responsive genes. For example, the HDAC inhibitors (apicidin, oxamflatin, HC toxin, and TSA) are known to have a similar mechanism of action and elicited highly similar transcriptome perturbations (error-weighted Pearson correlation coefficient range 0.86–0.92) with prominent groupings of induced or repressed genes on the left side of the heatmap. The transcriptome profiles of the more general category of nongenotoxic agents were not as strongly correlated (error-weighted Pearson correlation coefficient range 0.01-0.85), but other clusters can be appreciated by visual inspection. For example, the ER stress agents thapsigargin and tunicamycin grouped together by hierarchical clustering. Interestingly, a subset of induced genes (designated by the arrow in Fig. 2) for these ER stress agents were also strongly induced by a few other agents including MMS, which directly damages proteins. Many

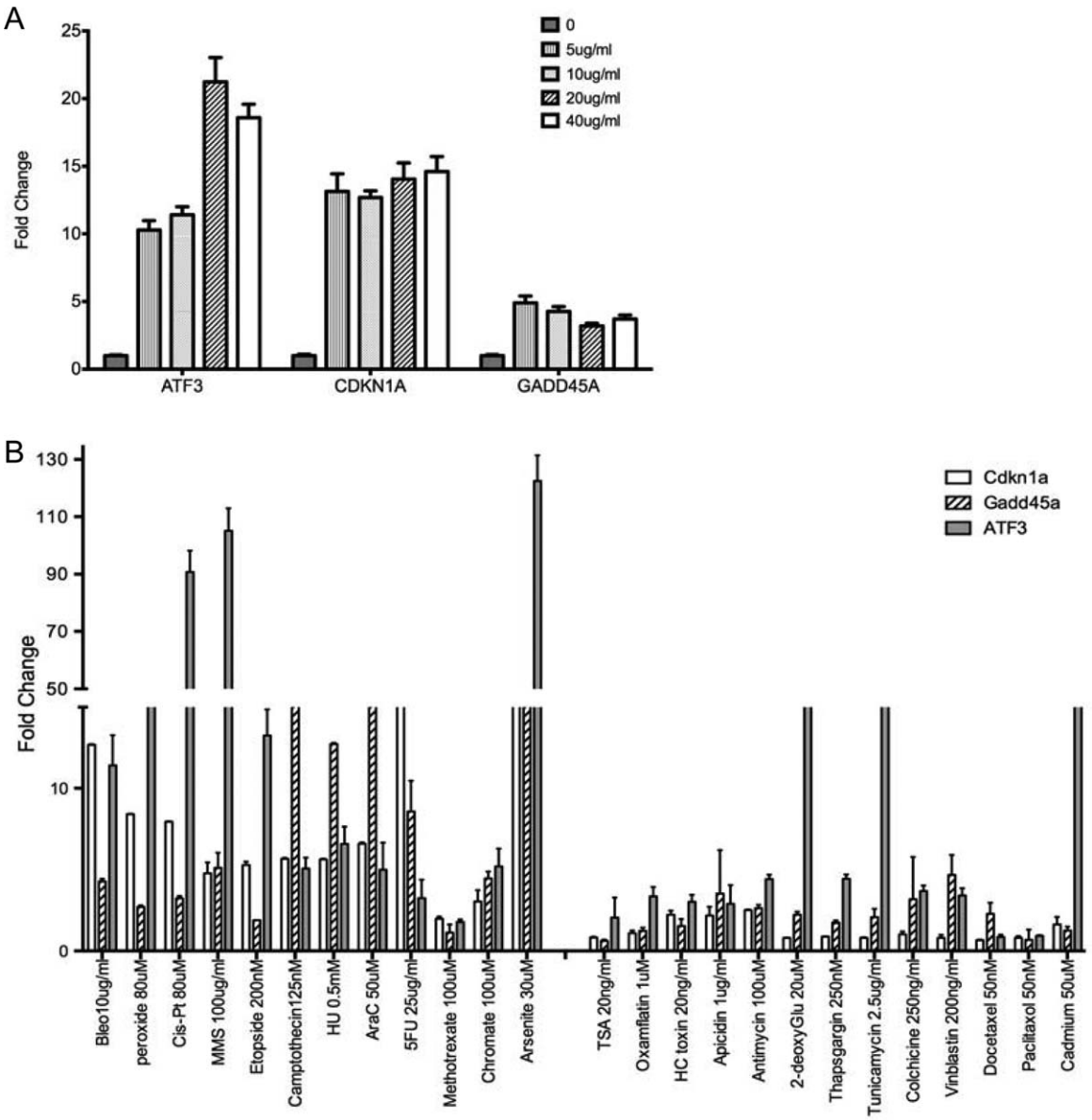

Fig. 1. Quantitative RT-PCR measurements for a panel of known stress response genes (*ATF3*, *CDKN1A*, and *GADD45A*) were used to guide stress agent dose selection. Doses were selected by measuring perturbations of this panel as a function of agent dose; fold change is relative to vehicle control cells done at the same time. The assay was done in triplicate cultures within one experiment. **A**: The dose response profile for bleomycin illustrates increased expression of the stress response genes for 5, 10, 20, and 40 μg/ml with an observable reduction in *GADD45A* upregulation at the two highest doses. **B**: Stress gene panel expression profiles for the doses of chemical agents eliciting a robust response.

toxicants can trigger multiple stress response pathways (Supporting Information Fig. S1), and one of the most striking examples is the case of chromate, which is known to damage DNA upon reduction to chromic anion within the cell [Whiting et al., 1979; Fornace et al., 1981; Salnikow and Zhitkovich 2008]. As expected, chromate induced a set of genes in common with the other genotoxic agents (Fig. 2, red bars). However, chromate also strongly induced expression changes in a set of genes that was only strongly induced by cadmium and heat shock (which did not elicit genotoxic stresses), while these genes were strongly repressed by some of the other genotoxic stress agents (bleomycin, cisplatin, MMS, and arsenite) (Fig. 2, blue bars).

Delineation of Gene Subclusters Using a Biclustering Approach

Although two-dimensional clustering can provide biological insight by organizing large data sets for visual inspection, it does not indicate the extent to which genes participate in multiple pathways (Supporting Information

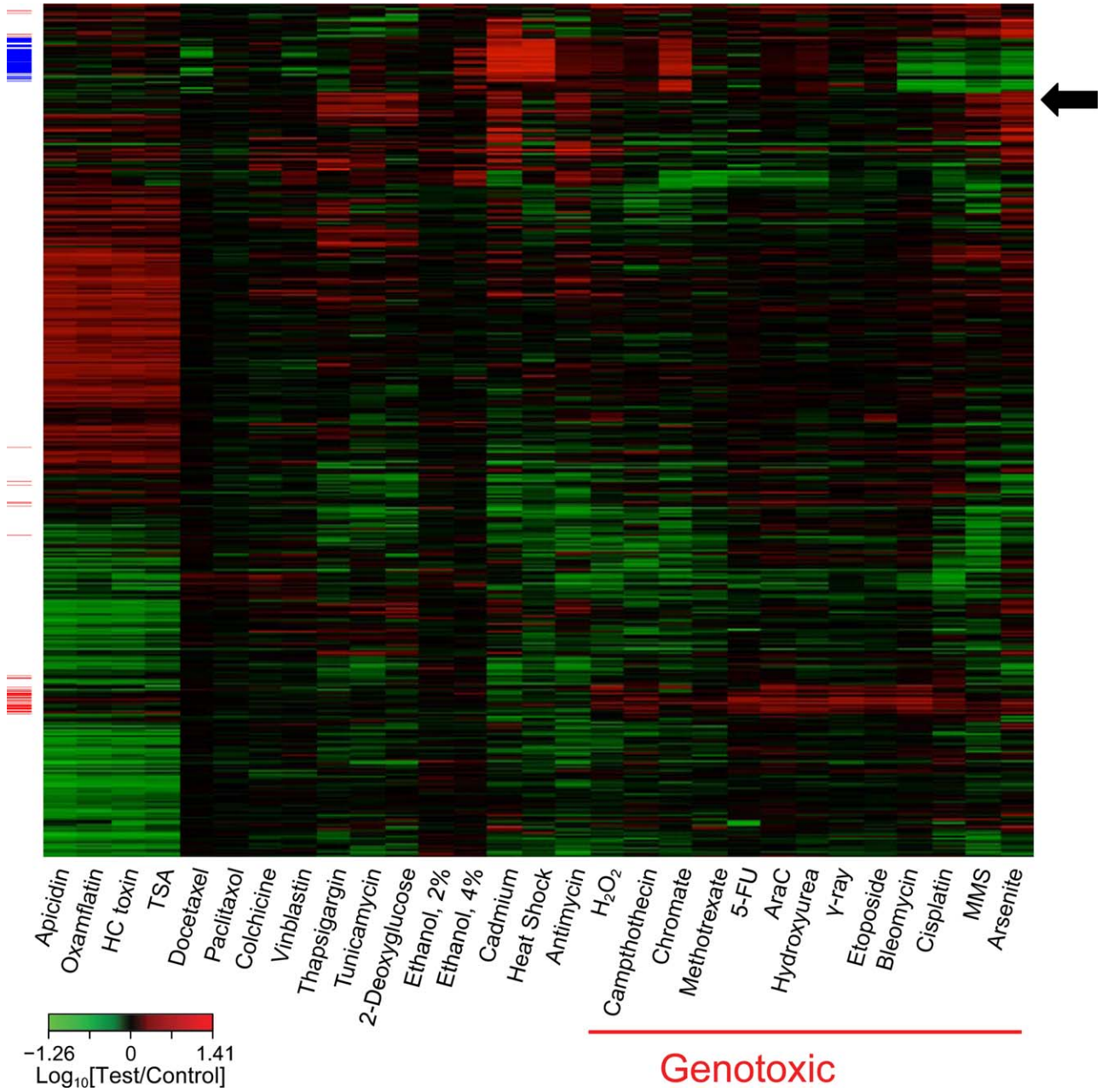

Fig. 2. Heatmap of transcriptome profiling data illustrating coexpressed sets of genes associated with various toxicants, such as genotoxic agents (red side bars). Some agents also exhibit obvious pleiotropy; e.g., chromate induced expression of genes associated with genotoxic agents (red side bars) and genes associated with cadmium and heat shock (blue side bars). This heatmap contains 1,628 genes that were significantly ($P < 0.01$; *t*-test) perturbed at least 1.7-fold, relative to the control, by at least one stress agent. The genes in the heatmap were organized by hierarchical clustering with complete linkage based on their error-weighted Pearson distances. The genotoxic (red side bars) and cadmium and heat shock (blue side bars) were identified using coupled two-way clustering. The arrow highlights a cluster of genes responding to ER stress agents thapsigargin and tunicamycin as well as certain other agents.

Fig. S1). To account for the fact that the response to toxicity can involve overlapping subsets of genes, a variety of biclustering methods have been developed. One of the most promising methods is coupled two-way clustering (CTWC) using superparamagnetic clustering (SPC) [Domany, 1999; Getz et al., 2000]. Not only does this method account for gene products mediating multiple stress response pathways, it also provides a metric for assessing the statistical significance of each cluster of genes.

When we applied the CTWC/SPC biclustering algorithm to our transcriptomic data, we identified 67 gene groups with stability scores ≥ 40 (Supporting Information Fig. S4); these groups will subsequently be referred to as

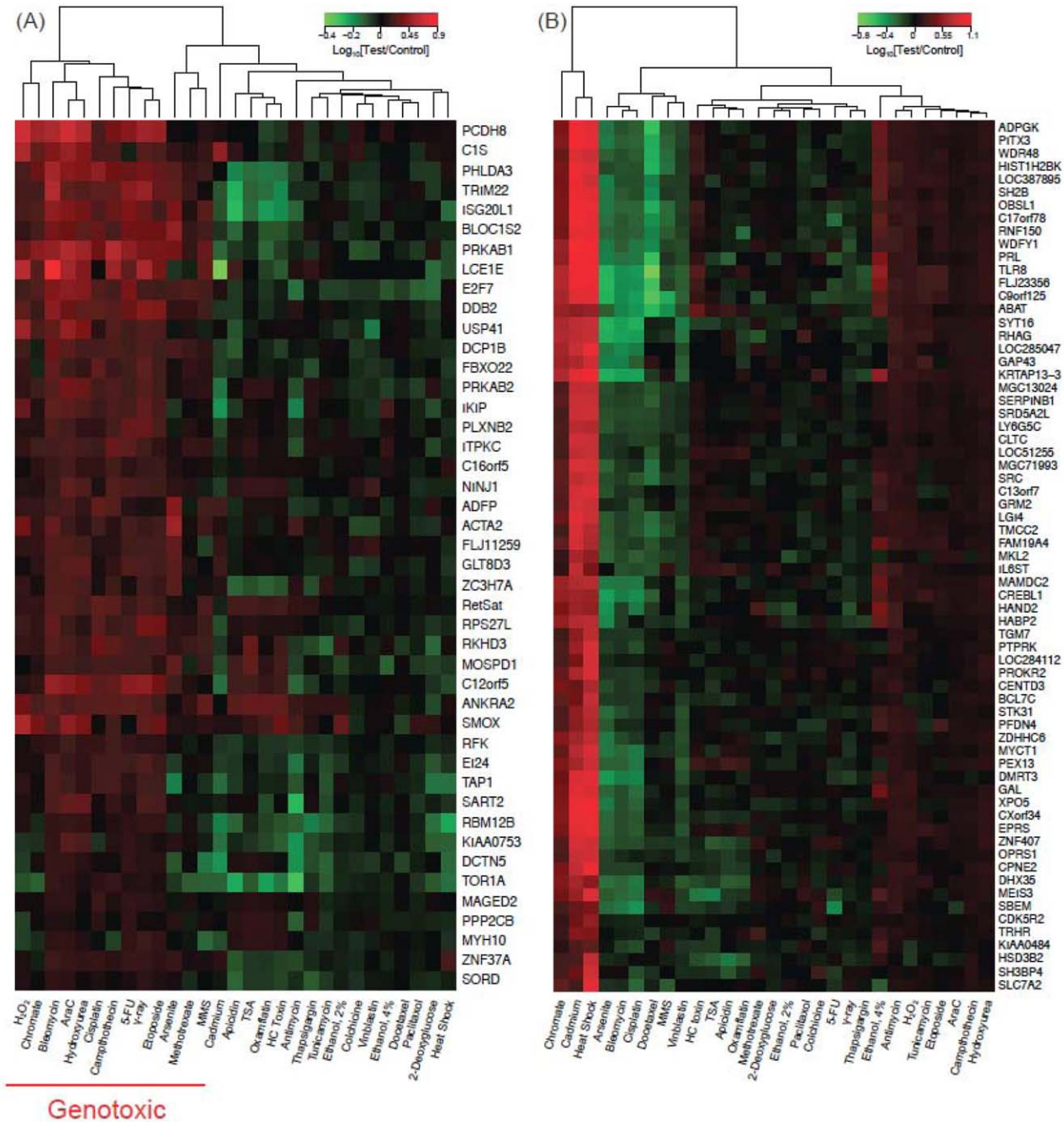

Fig. 3. Chromate induces pleiotropic stress responses that involve genotoxic and nongenotoxic modes. **A**: Cluster of chromate-responsive genes that tend to be upregulated in response to genotoxic agents. **B**: Genes that are upregulated in response to chromate, cadmium, and heat shock but are either repressed or not perturbed by other genotoxic agents. The genes and toxicants in the heatmaps were organized by hierarchically clustering with complete linkage based on their error-weighted Pearson distances after selection using the CTWC/SPC biclustering algorithm.

subclusters to avoid confusion with results from hierarchical clustering. These gene subclusters were each comprised of 20−193 genes, with over 50% of the genes participating in multiple clusters and a single gene participating in as many as eight clusters. This method revealed a subcluster of genes induced by chromate and the genotoxic agents that was further analyzed using a hierarchical cluster approach (Fig. 3A), as well as a set of genes induced by chromate, cadmium, and heat shock (Fig. 3B). Ingenuity pathway analysis indicated that the chromate/genotoxic agent subcluster was enriched for DNA-damage responsive genes, some of which are known to be regulated by p53, such as *AEN*, *C12orf5*, *EI24*, and *PHLDA3* (Supporting Information Fig. S5A); whereas, the chromate/cadmium/heat shock cluster included genes involved in amino acid metabolism, amino acid

transportation, and the unfolded protein response (Supporting Information Fig. S5B). Interestingly, the latter subcluster does not include classic heat shock protein genes, such as *HSPA1A*.

Development of a Genomic Biomarker for Genotoxicity

To identify an mRNA signature indicative of genotoxicity, we applied the nearest shrunken centroids method [Tibshirani et al., 2002] to our transcriptome data. We identified a panel of 65 transcripts whose expression resulted in 100% accuracy as assessed by 10-fold cross-validation (see Material and Methods). This gene panel, designated TGx-28.65, displayed a clear-cut separation between genotoxic and nongenotoxic agents (Fig. 4A). A number of these genes, such as *CDKN1A, GADD45A*, and *TRIM22*, are regulated by p53 [Riley et al., 2008]. For a complete list of the 65 genes in the TGx-28.65 biomarker refer to Table II.

Validation and a Case Study on Utility of the TGx28.65 Biomarker in Human Health Risk Assessment

To assess the practical utility of this genomic biomarker-based approach in the assessment of genotoxic hazard, we evaluated the gene expression profiles of three well-characterized agents: 3-nitropropionic acid (3-NP), isopropyl methanesulfonate (iPMS), and trimethylxanthine (caffeine). 3-NP is an irreversible inhibitor of succinate dehydrogenase (also known as Complex II of the mitochondrial respiratory chain) that participates in the TCA cycle and electron transport chain. It forms a covalent adduct with Arg297 in the active site of succinate dehydrogenase resulting in severe energy (ATP) impairment; therefore, 3-NP is considered to be a nongenotoxic agent based on our definition of genotoxicity. iPMS belongs to a class of DNA alkylating chemicals that are positive in standard in vitro gene mutation and chromosome aberration assays and thus should be readily detectable with our genomic biomarker approach as genotoxic. In contrast, the genotoxicity profile of caffeine is typical of compounds that present challenges for assessment of potential genotoxic hazard. Specifically, caffeine gives negative results in bacterial mutation assays, positive findings in in vitro chromosome aberration assays, and is not genotoxic in vivo (discussed in [Goodsaid et al., 2010]).

To find the appropriate doses for transcriptomic profiling of these agents, we first assessed mRNA levels of *ATF3, CDKN1A*, and *GADD45A* in cells treated in increasing concentrations of the agents. The concentrations that showed the most robust induction of one or more of these genes were selected and it was ensured that these were not overtly cytotoxic (as described above). As shown in Supporting Information Figure S7, the optimal dose for transcriptomic profiling was 3 mM for 3-NP, 250 μg/ml for iPMS and 2 mM for caffeine, respectively. Comparison of the gene expression profiles of these test agents with the TGx-28.65 gene panel classified 3-NP as nongenotoxic and iPMS as genotoxic. Analysis of the gene expression profile of caffeine resulted in it being classified as a nongenotoxic agent using the TGx28.65 gene classifier (Fig. 4B).

DISCUSSION

In this study, we developed a robust transcriptomic signature for the assessment of genotoxicity using a refined in vitro approach. Importantly, we established a protocol for optimal dose selection that can be used to ensure that cells are not over- or under-dosed and enables more effective comparison across toxicants. We identified a gene signature, designated TGx-28.65, generated using 28 model agents and comprised of 65 transcripts whose expression resulted in 100% accuracy in classifying genotoxic and nongenotoxic agents as assessed by 10-fold cross-validation. This signature was derived from both DNA reactive- and DNA nonreactive genotoxic agents (Table I) and clearly distinguished the 13 genotoxic agents from the 15 nongenotoxic agents with widely varying mechanisms of action.

To evaluate the practical utility of this 65-gene panel, we tested three additional agents with known mechanisms of action, including caffeine which is known to be positive in the in vitro chromosome aberration assay [Weinstein et al., 1975] but is negative in in vivo genotoxicity tests [Goodsaid et al., 2010]. iPMS was chosen as a prototypical representative of DNA alkylating agents that is positive for all assays in the standard genotoxicity testing battery in vitro and in vivo [Segal et al., 1986]. While 3-NP is considered nongenotoxic [Oshiro et al., 1991], a contaminant in previous studies may have led to a positive bacterial mutation data reported in the earlier literature [Hansen 1984; Zeiger et al., 1988]. Therefore, we carried out a Salmonella mutation assay to test the genotoxicity of this agent and the results were negative (data not shown). As we expected, both agents with well-characterized mechanisms were correctly classified, supporting that this 65-gene panel can be used to distinguish genotoxic and nongenotoxic agents (Fig. 4B). Caffeine is well-known to show positive findings in the in vitro chromosome aberration assay [Weinstein et al., 1975]; however, there was no genotoxic response in cells treated with caffeine suggesting that the clastogenicity observed in the in vitro mammalian chromosome damage assay is not likely to be caused by DNA reactive mechanisms and is probably the consequence of general toxicity or other molecular processes. Finally, we note that the TGx28.65 biomarker also classifies genotoxicity correctly for chemicals requiring metabolic activation (benzo[a]pyrene,

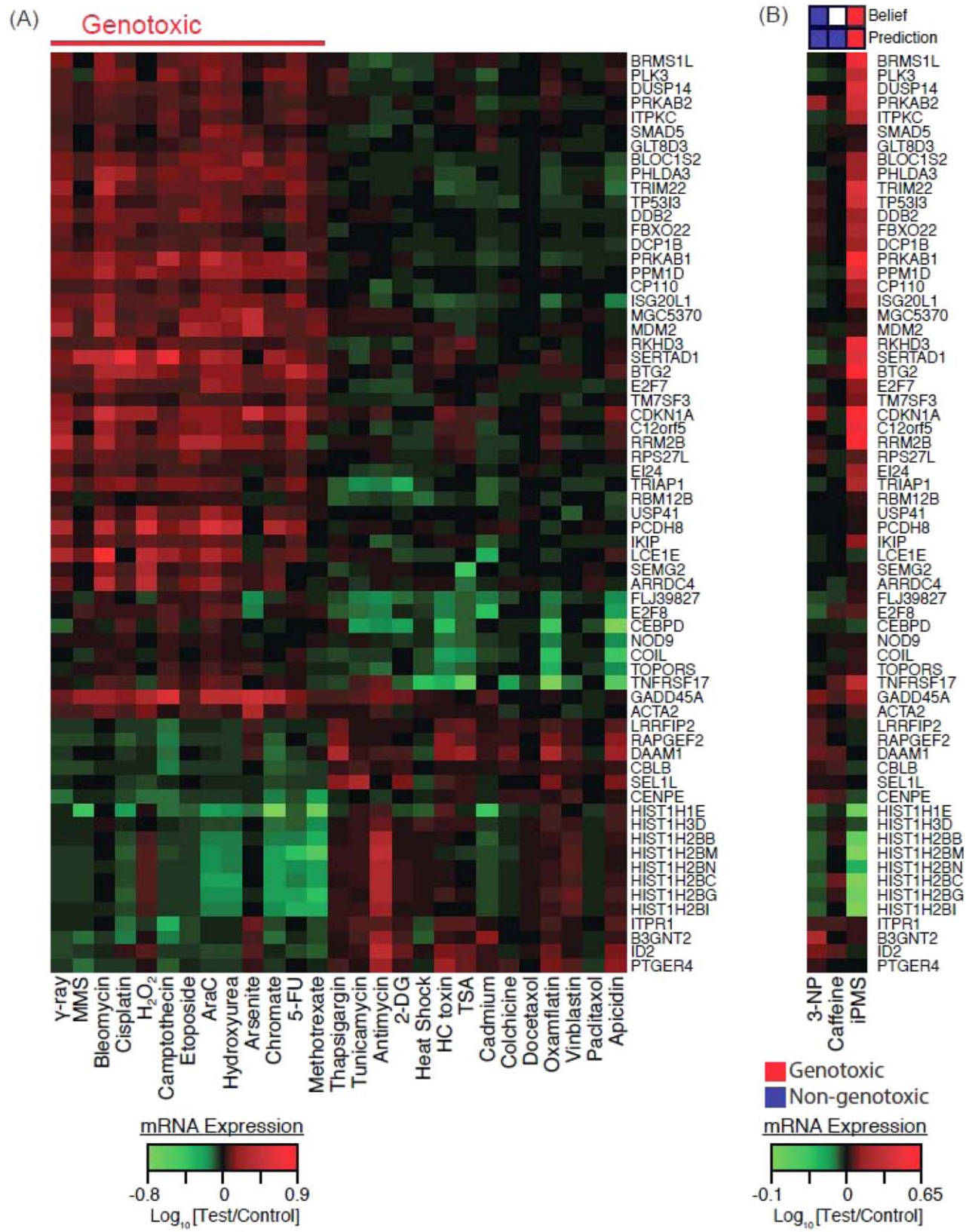
(A)
Genotoxic
BRMS1L
PLK3
DUSP14
PRKAB2
ITPKC
SMAD5
GLT8D3
BLOC1S2
PHLDA3
TRIM22
TP53I3
DDB2
FBXO22
DCP1B
PRKAB1
PPM1D
CP110
ISG20L1
MGC5370
MDM2
RKHD3
SERTAD1
BTG2
E2F7
TM7SF3
CDKN1A
C12orf5
RRM2B
RPS27L
EI24
TRIAP1
RBM12B
USP41
PCDH8
IKIP
LCE1E
SEMG2
ARRDC4
FLJ39827
E2F8
CEBPD
NOD9
COIL
TOPORS
TNFRSF17
GADD45A
ACTA2
LRRFIP2
RAPGEF2
DAAM1
CBLB
SEL1L
CENPE
HIST1H1E
HIST1H3D
HIST1H2BB
HIST1H2BM
HIST1H2BN
HIST1H2BC
HIST1H2BG
HIST1H2BI
ITPR1
B3GNT2
ID2
PTGER4
γ-ray
MMS
Bleomycin
Cisplatin
H2O2
Camptothecin
Etoposide
AraC
Hydroxyurea
Arsenite
Chromate
5-FU
Methotrexate
Thapsigargin
Tunicamycin
Antimycin
2-DG
Heat Shock
HC toxin
TSA
Cadmium
Colchicine
Docetaxol
Oxamflatin
Vinblastin
Paclitaxol
Apicidin
mRNA Expression
-0.8
0
0.9
Log10[Test/Control]
(B)
Belief
Prediction
3-NP
Caffeine
iPMS
Genotoxic
Non-genotoxic
mRNA Expression
-0.1
0
0.65
Log10[Test/Control]

Fig. 4.

aflatoxin B1, dexamethasone, and phenobarbital) in TK6 cells coexposed to the aforementioned chemicals and rat liver S9 in a companion paper [Buick et al., 2015].

One important consideration is that our genotoxic classifier does not include some prominent aneugens in the antimitotic agent group. While the taxels and vinblastine are effective cancer chemotherapy agents, their properties are distinct from typical cytotoxic chemotherapy agents that damage DNA. For example, the cellular responses in the NCI60 cell line panel substantially differed as exemplified by lack of correlation with p53 status with cytoxicity for antimitotic agents [O'Connor et al., 1997; Weinstein et al., 1997].

The strategy used here builds on our earlier report in TK6 cells and its p53-null derivative NH32 cells where seven genotoxic agents were compared to three nongenotoxic agents and three oxidizing agents. In this earlier study using a smaller custom cDNA array with 7.7k features [Amundson et al., 2005], some separation between agent classes was observed, but the contribution by p53 was minimized by inclusion of the p53-null cell line and a robust classifier for genotoxicity was not further developed. When a 16-gene signature was defined to discriminate p53 status, only four genes overlapped with our current 65-gene classifier. This earlier limited study did not employ dose optimization or the current bioinformatics approach, but did assess early responses to stress at 4 hr. As seen previously, there was also no appreciable cytotoxicity detected by trypan blue staining at 4 hr (Supporting Information Fig. S3) and assessment of viability at 24 hr showed only moderate effects on viability for most agents (Supporting Information Fig. S3) in this study. A collaborative effort [Buick et al., 2015] demonstrated the utility of our TGx-28.65 biomarker by accurately classifying benzo[a]pyrene and aflatoxin B1 as genotoxic and dexamethasone and phenobarbital as nongenotoxic in the presence of rat liver S9 at 4 and 8 hr in TK6 cells. These results support the use of early time points (up to 8 hr) in the presence of S9, but suggest that later time points (24 hr) may reveal false classification for nongenotoxic agents under conditions of high cytotoxicity. However, the companion paper also reveals that TGx-28.65 is 100% accurate in classifying genotoxicity in human liver HepaRG cells at much later time points.

Compared to previous studies (discussed in [Li et al., 2007; Cui and Paules, 2010; Godderis et al., 2012]), our approach benefited by combining dose optimization, an informatics approach that takes into account overlapping signaling events (Supporting Information Fig. S1), duration of response time (4 h), and choice of cell line. As discussed previously [Amundson et al., 2005; Li et al., 2007], many toxicants, including DNA-damaging agents, trigger transcriptional and post-transcriptional responses resulting in high and frequently maximal mRNA accumulation within several hours These toxicants usually show a dose-dependence until very high concentrations at which point transcriptional machinery or cell integrity is compromised. The aim here is to generate clear responses and accurate stress-response signatures rather than assess cellular parameters such as lethality and cell cycle redistribution *per se*. Many studies have relied on toxicological parameters, such as micronucleus formation, mutagenesis, or long-term ($>$ 24 h) cytotoxicity for dose determination, so subtle changes in gene expression can impact these parameters but give an unreliable signal for transcriptomics. Thus, a few marginal or sub-threshold responses can compromise classifier development, and even with very large surveys [Lamb et al., 2006], lead to some false negatives. Our qRT-PCR dose optimization approach was effective for both genotoxic and nongenotoxic agents using only three well-characterized stress genes, and, as shown in Figure 1B, the pattern of responses tended to differ for these two classes particularly with more robust *CDKN1A* induction for genotoxic agents. While *GADD45A* is a p53-regulated gene [Kastan et al., 1992], it can be induced by many stresses and was initially isolated in p53-deficient rodent cells [Fornace et al., 1988]. *GADD45A* mRNA has been found to increase following exposure to many nongenotoxic agents, but elements of this gene's promoter may show more specificity [Lynch et al., 2011]. Taken together, our dose optimization procedure provides a standard condition for every agent's effect at the selected dose, and decreases the likelihood of false negatives.

Several mammalian cell lines are commonly used in toxicology screening assays, but TK6 cells have characteristics that make them attractive for transcriptomic genotoxicity testing. For example, the commonly used mouse lymphoma and HepG2 lines are tumor-derived with compromised tumor suppressor pathways while TK6 is a spontaneously transformed lymphoid line [discussed in [Amundson et al., 2005; Li et al., 2007]]. Moreover, the

Fig. 4. Discrimination of genotoxic and nongenotoxic agents based on gene expression. **A**: Heatmap of a 65-gene panel that was able to discriminate genotoxic from nongenotoxic agents with 100% accuracy after 10-fold cross validation. The genes were identified applying the nearest shrunken centroids (NSC) algorithm to the 28 treatments and 1628 genes that were significantly ($P < 0.01$; *t*-test) perturbed at least 1.7-fold, relative to the control, by at least one stress agent. The NSC threshold was set at 2.2 to ensure 100% accuracy. The genes in the heatmap were organized by hierarchically clustering with complete linkage based on their Pearson distances. **B**: Classification of three test agents using the TGx-28.65 gene panel. The agents 3-NP and iPMS showed correct classification compared to their known mechanisms of action. Caffeine was classified as a nongenotoxic agent. Bars above the heatmap: belief indicates the genotoxicity of agents based on their known mechanisms; prediction designates the probability of genotoxicity or nongenotoxicity using Tgx28.65. Red: genotoxic, blue: nongenotoxic, white: unclassified.

TABLE II. Genes in TFx-28.65 Classifier

Entrez ID	GeneSymbol	Response	p53 Requlated
59	ACTA2	↑	Yes
64782	AEN	↑	Yes
7832	BTG2	↑	Yes
57103	C12orf5	↑	Yes
1026	CDKN1A	↑	Yes
1643	DDB2	↑	Yes
11072	DUSP14	↑	Yes
144455	E2F7	↑	Yes
9538	EI24	V	Yes
26263	FBX022	↑	Yes
1647	GADD45A	↑	Yes
121457	IKBIP	↑	Yes
4193	MDM2	↑	Yes
23612	PHLDA3	↑	Yes
8493	PPM1D	↑	Yes
51065	RPS27L	↑	Yes
50484	RRM2B	↑	Yes
9540	TP53I3	↑	Yes
51499	TRIAP1	↑	Yes
10346	TRIM22	↑	Yes
91947	ARRDC4	↑	
10678	B3GNT2	↓	
282991	BLOC1S2	↑	
84312	BRMS1L	↑	
868	CBLB	V	
9738	CCP110	↑	
1052	CEBPD	V	
1062	CENPE	↓	
8161	COIL	V	
23002	DAAM1	V	
196513	DCP1B	↑	
79733	E2F8	↑	
139285	FAM123B	V	
283464	GXYLT1	V	
3008	HIST1H1E	↓	
3018	HIST1H2BB	↓	
8347	HIST1H2BC	↓	
8339	HIST1H2BG	↓	
8346	HIST1H2BI	↓	
8342	HIST1H2BM	↓	
8341	HIST1H2BN	↓	
8351	HIST1H3D	↓	
3398	ID2	V	
80271	ITPKC	↑	
3708	ITPR1	V	
353135	LCE1E	↑	
9209	LRRFIP2	V	
84206	MEX3B	↑	
79671	NLRX1	V	
5100	PCDH8	↑	
1263	PLK3	↑	
5564	PRKAB1	↑	
5565	PRKAB2	↑	
5734	PTGER4	V	
9693	RAPGEF2	↑	
389677	RBM12B	V	
6400	SEL1L	V	
6407	SEMG2	↑	
29950	SERTAD1	↑	
4090	SMAD5	↑	
51768	TM7SF3	↑	
608	TNFRSF17	↑	
10210	TOPORS	V	
373856	USP41	↑	

Observed transcriptomics responses to genotoxic agents by members of the TGx28.65 dataset. Upregulated genes, "↑", represent those showing more than 1.5-fold induction for more than one third of tested genotoxic agents; downregulated genes, "↓", designate those showing more than 1.5-fold repression in more than one third of tested genotoxic agents; remaining genes showed more variable responses and are designated with "V". Genes known to be p53-regulated are indicated in the last column.

mouse lymphoma line [Storer et al., 1997] is p53-deficient and HepG2 cells exhibit less robust responses of stress genes (in particular those where p53 contributes to their regulation, data not shown). While TK6 cells have limited capacity for metabolic activation, the approach by [Buick et al., 2015] clearly shows its broad utility for genotoxicity assessment in the presence of rat liver S9. Our approach can be applied in the future to screen for additional effective lines from other tissue types.

In order to associate cellular function related to particular gene clusters, sets of genes from some of the major clusters in Figure 2 were subjected to Ingenuity Pathway Analysis. The pathways that were affected in the various agent clusters differed significantly, and suggest that pathway analysis can provide mechanistic insight into the action of the agents based on the cellular responses. The pathways of 'p53 signaling' and 'cell cycle: G2/M DNA damage checkpoint regulation' are highly overrepresented in our genotoxic signatures, with P-values of $1.6x10^{-7}$ and $7.9x10^{-4}$, respectively. The same pathways were also the most enriched in our TGx-28.65 classifier gene set (Supporting Information Table SII). It is possible that two or more categories affect the same function, but from a different aspect. For example, functional analysis revealed that both genotoxic agents and microtubule inhibitors affect genes involved in DNA replication, recombination, and repair. However in the genotoxic group, the processes involved in this function are DNA damage recognition and processing, while in the microtubule group the processes are involved in chromosomal structure and organization.

Transcriptome data have been used to classify pharmacologic agents with different mechanisms of action such as in Lamb et al. (2006), where a collection of drug-response transcriptome profiles was termed the Connectivity Map (CMAP). While no dose optimization was carried out, it was possible to use CMAP to identify compounds that may have similar effects on a set of genes selected using the gene set enrichment algorithm (GSEA) [Subramanian et al., 2005]. GSEA identifies

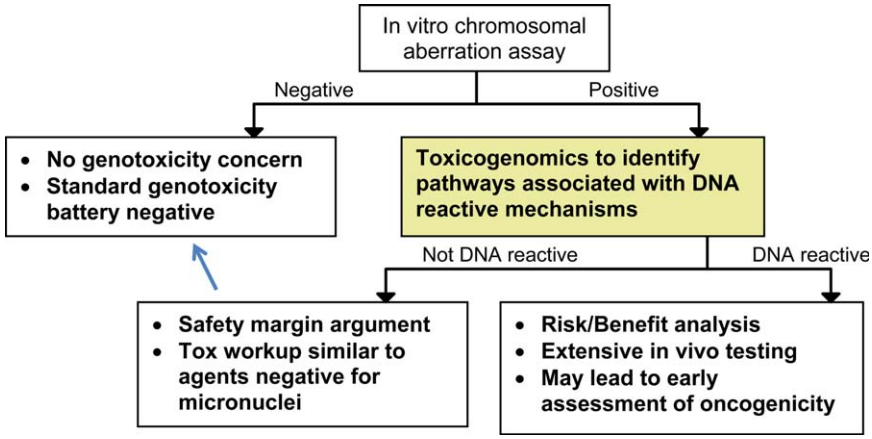

Fig. 5. Proposed scheme for incorporation of transcriptomics biomarker component into genotoxicity risk assessment. As discussed in the text, up to 30% of nongenotoxic agents give a positive result for chromosome aberrations as determined by micronucleus formation, and thus require further extensive evaluation. With the proposed toxicogenomics approach, nonreactive compounds would then be treated similarly to those that are negative in the chromosome aberration assay (designated by blue arrow).

genes that are strongly expressed, or repressed, across a set of transcriptome profiles. While GSEA will indicate whether a set of genes is strongly perturbed in a transcriptome database, it is not designed to account for genes participating in multiple gene sets. In this study we used a deconvolution method to address the challenge encountered with multiple mechanisms activated by one compound, by using a CTWC/SPC algorithm to define the stable subclusters. Biclustering approaches, such as CTWC/SPC [Getz et al., 2000], allow genes to participate in multiple subclusters in line with pleotropic activities. Pathway analysis of gene subclusters may serve to more effectively associate them with existing biological knowledge (Supporting Information Fig. S6).

Even though nongenotoxic agents represent a wide diversity of stresses and probably a myriad of PoT, our approach should have utility in the development of robust classifiers for distinct stress responses. For example, since HDAC inhibitors are known to trigger broad transcriptional responses [Lamb et al., 2006], they showed a pronounced signature in Figure 2, as expected. Other signatures or subclusters can also be discerned visually in Figure 2 or by CTWC/SPC, such as for ER stress agents or heat shock. Unanticipated subclusters were also delineated including a prominent one triggered by Cd, Cr, and a rather severe HS treatment [Fornace et al., 1989a], and probably represent a response to some perturbation of protein structure and/or function (Supporting Information Fig. S5B). Since ethanol has major public health consequences, it was added to the panel of agents even though it typically is studied in longer exposure studies in hepatic cells, typically at 0.1% or less or in vivo. While TK6 cells lack many of metabolic pathways present in hepatic cells, alcohol did elicit an appreciable response and a clear nongenotoxic signature. An ER stress response signature is unsurprising considering the relatively high concentrations of alcohol, but pathway analysis in Supporting Information Table SIII also showed TNF-related signaling, which has also been described in vivo.

In summary, we developed a toxicogenomics approach for differentiation of genotoxic agents from other toxicants based on the analysis of cellular stress responses. Our approach demonstrates the importance of the qRT-PCR molecular dose-selection protocol that established the appropriate dose by monitoring the expression of several known stress response genes to exposure with tested agents in a dose setting experiment. The approach yielded a panel of 65 biologically relevant genes that differentiate whether a particular agent is genotoxic or nongenotoxic. Our method was further independently confirmed by [Buick et al., 2015] in the presence of metabolic activation. While limited in scope, our approach should have the ability to classify such toxicants considering the diversity of nongenotoxic agents assessed in this study, especially as additional agents are assessed in the future.

We propose that our TGx-28.65 genomic biomarker is useful for risk assessment of genotoxic effects, in particular to provide mechanistic insight into positive findings in the in vitro chromosome aberration assays for compounds with otherwise clean genotoxicity profiles. Since the majority of such compounds are also negative in carcinogenicity tests, the risk assessment of these positive in vitro findings is a challenge to industry and regulatory agencies [Goodsaid et al., 2010]. Considering that thirty percent of drug candidates with otherwise clean genotoxicity profiles in the Salmonella assay and in the in vivo

micronucleus assay show positive test results in the in vitro chromosome aberration assays, the application of genomic biomarker-derived information would add significant value to the current genotoxicity testing. In our case study, caffeine as a prototypical agent with positive findings in the chromosomal aberration assay and absence of documented human cancer risk was evaluated by applying our toxicogenomics approach as shown schematically in Figure 5. In this approach the absence of the genotoxic signature is considered as a lack of genotoxicity relevant to human health [Goodsaid et al., 2010]. Although additional studies are necessary, the genomic biomarker approach as demonstrated in the case of caffeine has the potential to complement and/or replace a number of assays currently used for developing a 'weight of evidence' assessment by offering essential mechanistic information that is difficult or impossible to obtain using current methods. Since our biomarker approach is also simple, inexpensive and rapid it could be easily integrated into the evaluation of lead compounds and chemical series in early research stages of drug development. Furthermore, broad incorporation of a genomic biomarker-based genotoxic risk assessment into the new product development process would significantly reduce animal testing.

AUTHOR CONTRIBUTIONS

HHL, JA, and AJF designed the study, participated in analysis and contributed to writing the manuscript. HHL and RXC carried out the cellular and transcriptomic studies and contributed to data analyses. DRH analyzed microarray data and contributed to writing the manuscript. PH contributed to cell culture studies. CLY participated in the analysis, and contributed to writing and editing the manuscript. All authors approved the final manuscript. HHL, JA, AJF, RXC, CLY, and DRH had complete access to the study data.

REFERENCES

Akerman GS, Rosenzweig BA, Domon OE, McGarrity LJ, Blankenship LR, Tsai CA, Culp SJ, MacGregor JT, Sistare FD, Chen JJ, et al. 2004. Gene expression profiles and genetic damage in benzo(a)-pyrene diol epoxide-exposed TK6 cells. Mutat Res 549:43–64.

Amundson SA, Do KT, Vinikoor L, Koch-Paiz CA, Bittner ML, Trent JM, Meltzer P, Fornace AJ Jr. 2005. Stress-specific signatures: Expression profiling of p53 wild-type and -null human cells. Oncogene 24:4572–4579.

Buick JK, Moffat I, Williams A, Swartz CD, Recio L, Hyduke DR, Li HH, Fornace AJJ, Aubrecht J, Yauk CL. 2015. Integration of metabolic activation with a predictive toxicogenomics signature to classify genotoxic versus non-genotoxic chemicals in human TK6 cells. Environ Mol Mutagen.

Cui Y, Paules RS. 2010. Use of transcriptomics in understanding mechanisms of drug-induced toxicity. Pharmacogenomics 11:573–585.

Domany E. 1999. Superparamagnetic clustering of data—The definitive solution of an ill-posed problem. Physica A 263:158–169.

Ellinger-Ziegelbauer H, Fostel JM, Aruga C, Bauer D, Boitier E, Deng S, Dickinson D, Le Fevre AC, Fornace AJJ, Grenet O, et al. 2009. Characterization and interlaboratory comparison of a gene expression signature for differentiating genotoxic mechanisms. Toxicol Sci 110:341–352.

Fallo F, Pilon C, Barzon L, Pistorello M, Pagotto U, Altavilla G, Boscaro M, Sonino N. 1996. Effects of taxol on the human NCI-H295 adrenocortical carcinoma cell line. Endocr Res 22:709–715.

Fan F, Jin S, Amundson SA, Tong T, Fan W, Zhao H, Zhu X, Mazzacurati L, Li X, Petrik KL, et al. 2002. ATF3 induction following DNA damage is regulated by distinct signaling pathways and over-expression of ATF3 protein suppresses cells growth. Oncogene 21:7488–7496.

Fornace AJ, Jr, Seres DS, Lechner JF, Harris CC. 1981. DNA-protein cross-linking by chromium salts. Chem Biol Interact 36:345–354.

Fornace AJ Jr, Alamo IJ, Hollander MC. 1988. DNA damage-inducible transcripts in mammalian cells. Proc Natl Acad Sci USA 85: 8800–8804.

Fornace AJ, Jr, Alamo IJ, Hollander MC, Lamoreaux E. 1989a. Induction of heat shock protein transcripts and B2 transcripts by various stresses in Chinese hamster cells. Exp Cell Res 182:61–74.

Fornace AJ, Jr, Nebert DW, Hollander MC, Luethy JD, Papathanasiou M, Fargnoli J, Holbrook NJ. 1989b. Mammalian genes coordinately regulated by growth arrest signals and DNA-damaging agents. Mol Cell Biol 9:4196–4203.

Getz G, Levine E, Domany E. 2000. Coupled two-way clustering analysis of gene microarray data. Proc Natl Acad Sci USA 97:12079–12084.

Godderis L, Thomas R, Hubbard AE, Tabish AM, Hoet P, Zhang L, Smith MT, Veulemans H, McHale CM. 2012. Effect of chemical mutagens and carcinogens on gene expression profiles in human TK6 cells. PLoS ONE 7:e39205.

Goodsaid FM, Amur S, Aubrecht J, et al. 2010. Voluntary exploratory data submissions to the US FDA and the EMA: Experience and impact. Nat Rev Drug Discov 9:435–445.

Guyton KZ, Kyle AD, Aubrecht J, Burczynski ME, Carl K, Catalano J, Charlab R, Close S, Cornu-Artis C, Essioux L, et al. 2009. Improving prediction of chemical carcinogenicity by considering multiple mechanisms and applying toxicogenomic approaches. Mutat Res 681:230–240.

Hansen TJ. 1984. Ames mutagenicity tests on purified 3-nitropropionic acid. Food Chem Toxicol 22:399–401.

Hartung T, McBride M. 2011. Food for Thought. on mapping the human toxome. ALTEX 28:83–93.

Hastie T, Tibshirani R, Friedman JH. The Elements of Statistical Learning: Data Mining, Inference, and Prediction: With 200 Full-Color Illustrations. New York: Springer; 2001:xvi, 533.

Hyduke DR, Amundson SA, Fornace AJ Jr. Complexity of stress signaling and responses. In: Bradshaw RA, Dennis EA, editors. Regulation of Organelle and Cell Compartment Signaling. San Diego: Academic Press; 2011. pp 187–206.

Islaih M, Li B, Kadura IA, Reid-Hubbard JL, Deahl JT, Altizer JL, Watson DE, Newton RK. 2004. Comparison of gene expression changes induced in mouse and human cells treated with direct-acting mutagens. Environ Mol Mutagen 44:401–419.

Jennen D, Ruiz-Aracama A, Magkoufopoulou C, Peijnenburg A, Lommen A, van Delft J, Kleinjans J. 2011. Integrating transcriptomics and metabonomics to unravel modes-of-action of 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) in HepG2 cells. BMC Syst Biol 5:139.

Kastan MB, Zhan Q, el-Deiry WS, Carrier F, Jacks T, Walsh WV, Plunkett BS, Vogelstein B, Fornace AJ, JR. 1992. A mammalian cell cycle checkpoint pathway utilizing p53 and GADD45 is defective in ataxia-telangiectasia. Cell 71:587–597.

Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, Wrobel MJ, Lerner J, Brunet JP, Subramanian A, Ross KN, et al. 2006. The connectivity map: Using gene-expression signatures to connect small molecules, genes, and disease. Science 313:1929–1935.

Li HH, Aubrecht J, Fornace AJ Jr. 2007. Toxicogenomics: Overview and potential applications for the study of non-covalent DNA interacting chemicals. Mutat Res 623:98–108.

Lynch AM, Sasaki JC, Elespuru R, Jacobson-Kram D, Thybaud V, De Boeck M, Aardema MJ, Aubrecht J, Benz RD, Dertinger SD, et al. 2011. New and emerging technologies for genetic toxicity testing. Environ Mol Mutagen 52:205–223.

Magkoufopoulou C, Claessen SM, Jennen DG, Kleinjans JC, van Delft JH. 2011. Comparison of phenotypic and transcriptomic effects of false-positive genotoxins, true genotoxins and non-genotoxins using HepG2 cells. Mutagenesis 26:593–604.

Newton RK, Aardema M, Aubrecht J. 2004. The utility of DNA microarrays for characterizing genotoxicity. Environ Health Perspect 112:420–422.

O'Connor PM, Jackman J, Bae I, Myers TG, Fan S, Mutoh M, Scudiero DA, Monks A, Sausville EA, Weinstein JN, et al. 1997. Characterization of the p53 tumor suppressor pathway in cell lines of the National Cancer Institute anticancer drug screen and correlations with the growth-inhibitory potency of 123 anticancer agents. Cancer Res 57:4285–4300.

Oshiro Y, Piper CE, Balwierz PS, Soelter SG. 1991. Chinese hamster ovary cell assays for mutation and chromosome damage: Data from non-carcinogens. J Appl Toxicol 11:167–177.

Patterson TA, Lobenhofer EK, Fulmer-Smentek SB, Collins PJ, Chu TM, Bao W, Fang H, Kawasaki ES, Hager J, Tikhonova IR, et al. 2006. Performance comparison of one-color and two-color platforms within the MicroArray Quality Control (MAQC) project. Nat Biotechnol 24:1140–1150.

Riley T, Sontag E, Chen P, Levine A. 2008. Transcriptional control of human p53-regulated genes. Nat Rev Mol Cell Biol 9:402–412.

Salnikow K, Zhitkovich A. 2008. Genetic and epigenetic mechanisms in metal carcinogenesis and cocarcinogenesis: nickel, arsenic, and chromium. Chem Res Toxicol 21:28–44.

Segal A, Seidman I, Melchionne S, Albert RE, Upton AC. 1986. Rapid induction of thymic lymphomas by isopropyl methanesulfonate: A preliminary report. Proc Soc Exp Biol Med 183:132–135.

Storer RD, Kraynak AR, McKelvey TW, Elia MC, Goodrow TL, DeLuca JG. 1997. The mouse lymphoma L5178Y Tk+/- cell line is heterozygous for a codon 170 mutation in the p53 tumor suppressor gene. Mutat Res 373:157–165.

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, et al. 2005. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci USA 102:15545–15550.

Tibshirani R, Hastie T, Narasimhan B, Chu G. 2002. Diagnosis of multiple cancer types by shrunken centroids of gene expression. Proc Natl Acad Sci USA 99:6567–6572.

van Delft JH, van Agen E, van Breda SG, Herwijnen MH, Staal YC, Kleinjans JC. 2005. Comparison of supervised clustering methods to discriminate genotoxic from non-genotoxic carcinogens by gene expression profiling. Mutat Res 575:17–33.

Weinstein D, Mauer I, Katz ML, Kazmer S. 1975. The effect of methylxanthines on chromosomes of human lyphocytes in culture. Mutat Res 31:57–61.

Weinstein JN, Myers TG, O'Connor PM, Friend SH, Fornace AJJ, Kohn KW, Fojo T, Bates SE, Rubinstein LV, Anderson NL, et al. 1997. An information-intensive approach to the molecular pharmacology of cancer. Science 275:343–349.

Whiting RF, Stich HF, Koropatnick DJ. 1979. DNA damage and DNA repair in cultured human cells exposed to chromate. Chem Biol Interact 26:267–280.

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.

Zeiger E, Anderson B, Haworth S, Lawlor T, Mortelmans K. 1988. Salmonella mutagenicity tests. IV. Results from the testing of 300 chemicals. Environ Mol Mutagen 11 Suppl 12:1–157.

Zhan Q, Fan S, Bae I, Guillouf C, Liebermann DA, O'Connor PM, Fornace AJ Jr. 1994. Induction of BAX by genotoxic stress in human cells correlates with normal p53 status and apoptosis. Oncogene 9:3743–3751.

Zhan Q, Eldeiry W, Bae I, Alamo I, Kastan M, Vogelstein B, Fornace A. 1995. Similarity of the DNA-damage responsiveness and growth-suppressive properties of waf1/cip1 and gadd45. Int J Oncol 6:937–946.