A System-Based Approach to Interpret Dose- and Time-Dependent Microarray Data: Quantitative Integration of Gene Ontology Analysis for Risk Assessment

Xiaozhong Yu,* William C. Griffith,* Kristina Hanspers,† James F. Dillman III,‡ Hansel Ong,*
Melinda A. Vredevoogd,§ and Elaine M. Faustman*,[1]

*Department of Environmental and Occupational Health Sciences, University of Washington, Seattle, Washington 98105; †GenMAPP Development Team, Bioinformatics Research Associate/Conklin Lab Gladstone Institute of Cardiovascular Disease/UCSF, San Francisco, California 94158; ‡Cell and Molecular Biology Branch, U.S. Army Medical Research Institute of Chemical Defense, Aberdeen Proving Ground, Maryland 21010; and §Institute for Risk Analysis and Risk Communication, Environmental and Occupational Health Sciences, Seattle, Wasington

Although microarray technology has emerged as a powerful tool to explore expression levels of thousands of genes or even complete genomes after exposure to toxicants, the functional interpretation of microarray data sets still represents a time-consuming and challenging task. Gene ontology (GO) and pathway mapping have both been shown to be powerful approaches to generate a global view of biological processes and cellular components impacted by toxicants. However, current methods only allow for comparisons across two experimental settings at one particular time point. In addition, the resulting annotations are presented in extensive gene lists with minimal or limited quantitative information, data that are crucial in the application of toxicogenomic data for risk assessment. To facilitate quantitative interpretation of dose- or time-dependent genomic data, we propose to use combined average raw gene expression values (e.g., intensity or ratio) of genes associated with specific functional categories derived from the GO database. We developed an extended program (GO-Quant) to extract quantitative gene expression values and to calculate the average intensity or ratio for those significantly altered by functional gene category based on MAPPFinder results. To demonstrate its application, we applied this approach to a previously published dose- and time-dependent toxicogenomic data set (J. F. Dillman *et al.*, 2005, *Chem. Res. Toxicol.* 18, 28–34). Our results indicate that the above systems approach can describe quantitatively the degree to which functional gene systems change across dose or time. Additionally, this approach provides a robust measurement to illustrate results compared to single-gene assessments and enables the user to calculate the corresponding ED_{50} for each specific functional GO term, important for risk assessment.

Key Words: microarray; GO analysis; dose and time dependent; quantitative.

[1] To whom correspondence should be addressed at Institute for Risk Analysis and Risk Communication, Department of Environmental and Occupational Health Sciences, University of Washington, 4225 Roosevelt Way NE, Suite #100, Seattle, WA 98105. Fax: (206) 616-4875. E-mail: faustman@u.washington.edu.

Microarray technology has become a powerful tool to explore the expression levels of thousands of genes or even complete genomes after exposure to toxicants. A major challenge in the interpretation of toxicogenomic data is the functional interpretation, linking potentially interrelated alterations in gene expression to conventional toxicological end points (Hamadeh *et al.*, 2002; Moggs, 2005). The functional interpretation of microarray data sets represents a laborious and challenging task. Researchers gather annotations from databases and identify characteristics of extracted sets of genes. The Gene Ontology (GO) Consortium initiated the standardization of annotation terms making them applicable for different organisms, facilitating data exchange among laboratories and databases (Ashburner *et al.*, 2000; Camon *et al.*, 2004; Harris *et al.*, 2004). These characteristics render GO annotations a powerful tool for the interpretation of microarray data. The ability to determine which GO terms apply or which biological pathways are associated with each differentially expressed gene from a microarray experiment provides an ideal model to gain an understanding of what molecular processes are affected by the observed changes in gene expression. Recently, many groups have developed methods and tools for pathway analysis that are compatible with GO mapping and reveal statistically significant annotations associated with microarray data (Al-Shahrour *et al.*, 2004; Beissbarth and Speed, 2004; Dennis *et al.*, 2003; Doniger *et al.*, 2003). The combination of using GO and pathway mapping has been proposed as a powerful approach to generate an unbiased view of biological processes and cellular components that are regulated by toxicant exposure at the transcriptional level (Currie *et al.*, 2005). In general, these methods are based on a two-step process. First, a list of significantly altered genes is compiled based on statistical analysis (e.g., ANOVA). Second, annotation terms that are significantly over- or underrepresented within this list based on criteria such as fold change or statistical *p* value are compared to a reference list that usually

consists of all genes on the array. An excellent example of the application of this approach for the analysis of di(2-ethyl-hexyl)phthalate exposure is given (Currie *et al.*, 2005). How-ever, these current methods only allow for a comparison of two experimental settings at a time and/or are restricted to one of the main GO categories or to a specific GO level. Furthermore, the resulting annotations are presented in extensive gene lists lacking quantitative information on each regulated gene. This one-dimensional analysis does not allow for quantitative or qualitative evaluation of possible dose- or time-dependent genomic relationships, which is crucial for the application of toxicogenomic data in the field of risk assessment.

An approach using the data average from many different genes categorized into broad ''omic categories'' has been proposed (Jansen and Gerstein, 2000). However, there is no consensus about the omic categories, and no systemic approach has been proposed to identify and calculate the average data. Since then, there has been a rapid development of the GO project, a collaborative effort to address the need for consistent descriptions of gene products in different databases (Ashburner *et al.*, 2000; Camon *et al.*, 2004; Harris *et al.*, 2004). The GO collaborators are developing structured, controlled vocabularies (ontologies) that describe gene products in terms of their as-sociated biological processes, cellular components, and molec-ular functions in a species-independent manner. The use of GO terms by several collaborating databases facilitates uniform queries across them. The Gene Ontology identifier (GOID) for the term seems to be a perfect candidate for the proposed omic categories. Computer programs such as MAPPFinder are avail-able to link gene expression data to the GO hierarchy and to statistically identify the GO terms altered by toxicant exposure (Doniger *et al.*, 2003). To facilitate quantitative interpretation of dose- or time-dependent genomic data, we propose a system-based approach to integrate the raw gene expression data with the GO Ontological analysis. We developed a computer program called GO-Quant that automatically links functional gene category analysis result from MAPPFinder with the original gene expression data and calculates the average intensity or ratio for those significant genes within each GO term. We applied this approach to demonstrate its application in a published dose- and time-dependent toxicogenomic data set (Dillman *et al.*, 2005). Our results indicate that the above systematic approach can describe quantitatively the degree to which functional gene systems change across dose or time course. Additionally, this approach provides a global measure-ment to illustrate our results compared to single-gene assess-ments and enables the user to calculate the corresponding ED_{50} for each specific functional GO term.

SYSTEMS AND METHODS

Gene expression data set. In this paper, dose and temporal genomic data were obtained from Dillman *et al.* (2005). The complete raw data set was retrieved from NCBI's gene expression omnibus (GEO, http://www.ncbi.nlm.

nih.gov/geo/) through GEO Series accession number GDS 1027. As described in the study of Dillman *et al.* (2005), rats were injected in the femoral vein with liquid bis(2-chloroethyl)sulfide (sulfur mustard, SM), which circulates directly to the pulmonary vein and then to the lung. Rats were exposed to saline, isopropyl alcohol (vehicle control), 1, 3, or 6 mg/kg of SM in vehicle (isopropyl alcohol) and lungs harvested at 0.5, 1, 3, 6, and 24 h postinjection. Three biological replicates were used for each time point and dose tested. Physiological saline administered in a similar manner was also included as an additional control. RNA was extracted from the lungs and used as the starting material for the probing of replicate oligonucleotide microarrays. All experi-ments were performed using Affymetrix Rat RAE230A oligonucleotide arrays (Affymetrix, Santa Clara, CA).

Statistical analysis and comparisons. Raw gene expression data were imported to BRB Array Tools v3.3 for statistical analysis (Wright and Simon, 2003). A log_2 transformation was applied to the data set. Each array was normalized by using the median intensity over the entire set of arrays. Since in this paper we are demonstrating our system approach by applying it to dose or time course microarray data, we conducted one-way ANOVAs across doses or time points. Statistical comparisons were conducted across doses (vehicle control, 1, 3, or 6 mg/kg) for each time point (0.5, 1, 3, 6, and 24 h) or across time points for each dose by using the ''randomized variance'' F test (Long *et al.*, 2001; Wright and Simon, 2003). Significantly changed genes across doses (dose response) at specific time points (6 and 24 h) were selected based on one-way ANOVA test at $p < 0.005$ while significantly changed genes across time points (time course) at a dose of 6 mg/kg were selected based on one-way ANOVA test at $p < 0.001$ since there was a large number of genes changed with time. Hierarchical clustering analysis of functional gene category based on MAPPFinder results was performed across doses at two time points (6 and 24 h) or across time points at a dose of 6 mg/kg for the significantly changed genes. For the selected significant genes, ratios were derived by dividing each normalized gene value by the average value of the controls (vehicle) and then transforming the ratios to their respective log_2 ratios. These log_2-transformed ratios were used as input for the TIGR Multiexperiment Viewer (MEV) (Saeed *et al.*, 2003). Hierarchical clustering analysis was conducted on the output genes using average linkage and Euclidean dissimilarity methods (Eisen *et al.*, 1998). The complete output of the cluster analysis is shown with the fold change indicated colorimetrically.

GO analysis. To establish associations between treatment and the affected GO terms and pathways, we used MAPPFinder (http://www.GenMapp.org) (Doniger *et al.*, 2003), which links gene expression data to the GO hierarchy (Ashburner *et al.*, 2000; Camon *et al.*, 2004; Harris *et al.*, 2004). MAPPFinder calculates a Z score as well as a permutation test p value for the genes that are significantly altered across doses ($p \leq 0.005$) at 6 or 24 h or across time points ($p \leq 0.001$) at a dose of 6 mg/kg doses. Since we proposed to use combined average raw gene expression values (e.g., intensity or ratio) of genes associated with specific functional gene categories derived from the MAPPFinder results and there are two directions of gene expression changes at a certain dose (or time), either upregulation or downregulation within a specific gene category, the average of these two different directions of gene expression alteration will mask the degree of changes. Therefore, the upregulated or downregulated genes will be calculated separately in the MAPPFinder. In this study we found simple patterns of expression of a consistent upregulation or consistent downregulation across dose and time based on K-means cluster analysis (Soukas *et al.*, 2000). K-means cluster analysis was used to group individual genes across doses (or times) into two groups classified as up- or downexpression based on the trend with dose (or time) of the mean expression pattern. The number of clusters used in our analysis was chosen based upon having the mean expression patterns against dose (or time), and this assumption of using only two clusters was tested by examining using more than two clusters. To test the robustness of this method, we compared the K-means cluster analysis with a manual analysis of the trend. In the manual analysis of trend, we assigned an upregulation trend of a gene when the gene expression ratio was significantly elevated in at least three of the four groups or elevated at both SM 3 and 6 mg/kg groups. The trend of downregulation of a gene across groups was assigned when the gene expression

level was significantly downregulated in at least three of the four groups or downregulated at both SM 3.0 and 6 mg/kg groups. The manual analysis of trend tended to create isolated clusters with few genes and larger clusters poorly represented by their mean expression patterns.

Each GO node and the cumulative total of the number of genes either upregulated or downregulated in a parent GO term combined with all its children were calculated. Z scores, a statistical measure of significance for gene expression in a given group, were calculated by subtracting the number of genes expected to be randomly changed in a GO term from the observed number of changed genes in that GO term. This value was then divided by the SD of the observed number of genes under a hypergeometric distribution. The equation can be written out as

$$Z\,\text{score} = (r - n(R/N))/((n(R/N))(1-(R/N))(1-((n-1)/(N-1)))^{1/2}, \quad (1)$$

where N is the total number of genes measured, R is the total number of genes meeting the criterion that the gene be significantly changed based on an F test at significant p value, n is the total number of genes in each specific GO term, and r is the number of genes meeting the criterion in this specific GO term. A positive Z score indicates that there are more genes meeting the criterion in a GO term than would be expected by random chance. If the MAPPFinder data truly obeyed the assumptions of the hypergeometric distribution, then a Z score of 1.96 or -1.96 would correlate with a p value of 0.05. In order to address the multiple testing that occurs in GO database in using Z score, a nonparametric permutation test of the data was calculated (see Doniger *et al.*, 2003, for more detail). This Z score allows us to rank and describe alterations in gene expression by each GO term based on the relative amount of altered gene expression within each category. For a GO term to be included in Tables 2–4, we set that at least three genes changed significantly (nested results) and that the permutation p value ≤ 0.05.

Quantitative integration of GO analysis to evaluate dose- or time-dependent genomic data. MAPPFinder links gene expression data to the GO hierarchy and statistically identifies the GO term altered by the toxicant exposure (Doniger *et al.*, 2003). Selecting genes for each category based on the arbitrary cutoff of fold change or p value change does not take into account the magnitude of the fold change of each gene, which is critical in the evaluation of dose- or time-dependent genomic data. To facilitate quantitative interpretation of dose- or time-dependent genomic data, we propose a system-based approach to integrate the raw gene expression data with the results from GO analysis. We developed a program called GO-Quant which automatically links results from functional GO analysis (MAPPFinder) with the original gene expression data and calculates the average intensity or ratio for those significantly genes within each GO term. For example, the average intensity or ratio of all significantly changed genes under statistically identified GOID was automatically calculated for each dose. Within a GOID, both upregulated and downregulated genes exist, and the calculation of this kind of average can be complicated. Thus, in our MAPPFinder analysis, the upregulated and downregulated genes were analyzed separately, and the average of intensity or ratio within a GOID in the upregulation or downregulation was calculated separately. The output file from the GO-Quant includes GO term, Z score, or average intensity or ratio, as well as gene ID. In the further analysis, the above output file was imported to MEV (Saeed *et al.*, 2003) and a hierarchical clustering analysis on each GO term conducted using average linkage and Euclidean dissimilarity methods (Eisen *et al.*, 1998). This hierarchical clustering analysis allows one to visualize the quantitative GO results. A stand-alone application for the above calculation (GO-Quant) is available now from the authors (http://depts.washington.edu/irarc/Go-Quant/).

RESULTS AND DISCUSSION

Framework for the Quantitative Functional Interpretation of Dose- and Time-Dependent Genomic Data

As shown in Figure 1, typically three main steps are included in the framework for the quantitative functional interpretation
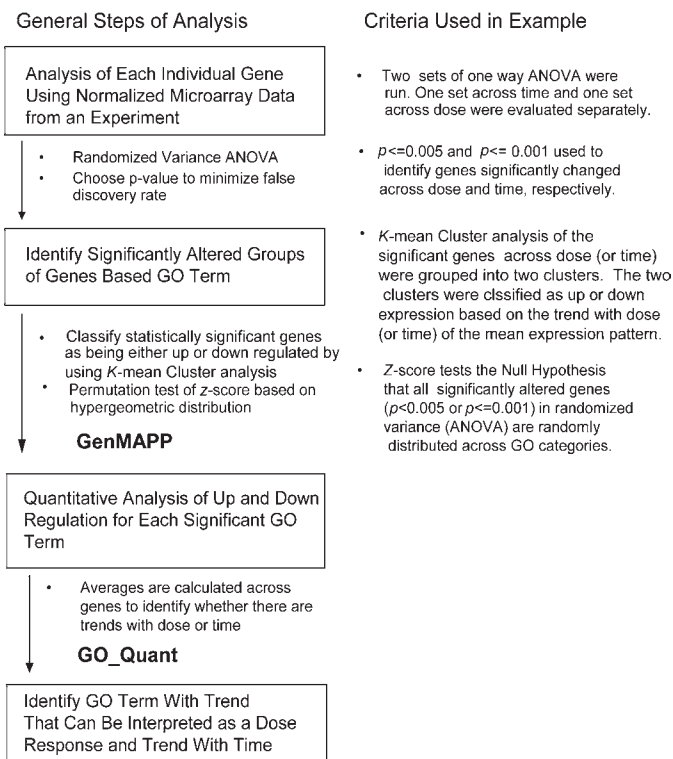
General Steps of Analysis

> Analysis of Each Individual Gene Using Normalized Microarray Data from an Experiment

- Randomized Variance ANOVA
- Choose p-value to minimize false discovery rate

> Identify Significantly Altered Groups of Genes Based GO Term

- Classify statistically significant genes as being either up or down regulated by using *K*-mean Cluster analysis
- Permutation test of z-score based on hypergeometric distribution

> **GenMAPP**

> Quantitative Analysis of Up and Down Regulation for Each Significant GO Term

- Averages are calculated across genes to identify whether there are trends with dose or time

> **GO_Quant**

> Identify GO Term With Trend That Can Be Interpreted as a Dose Response and Trend With Time

Criteria Used in Example

- Two sets of one way ANOVA were run. One set across time and one set across dose were evaluated separately.

- $p <= 0.005$ and $p <= 0.001$ used to identify genes significantly changed across dose and time, respectively.

- *K*-mean Cluster analysis of the significant genes across dose (or time) were grouped into two clusters. The two clusters were clssified as up or down expression based on the trend with dose (or time) of the mean expression pattern.

- *Z*-score tests the Null Hypothesis that all significantly altered genes ($p<0.005$ or $p<=0.001$) in randomized variance (ANOVA) are randomly distributed across GO categories.

FIG. 1. System-based schematic framework for the interpretation of dose- and time-dependent genomic data.

of dose- and time-dependent genomic data. The first step is a traditional microarray analysis, aimed at identifying the statistically significant genes across dose or time course by using the ANOVA method. The next step is to identify significantly altered functional GO gene categories or pathways by using unsupervised GO analysis or a pathway-mapping tool such as MAPPFinder. The last step is to extract the quantitative gene expression data and calculate the average intensity or ratio for those genes significantly altered within each functional gene category based on MAPPFinder by using our GO-Quant program. In order to demonstrate its power in quantitative evaluation of dose- and time-dependent microarray data, we applied this approach to a published dose- and time-dependent toxicogenomic data set in rats treated with SM.

Table 1 shows the outline of gene alteration across dose and time. Time course of the total number of significantly changed genes across doses (0, 1, 3, or 6 mg/kg) for each time point (0.5, 1, 3, 6, and 24 h) by one-way ANOVA comparisons ($p \leq 0.005$) is listed at the right side of the table. The number of significantly changed genes increased in a time-dependent manner, peaking at 6 h. Dose-dependent increase in the number of significantly changed genes ($p \leq 0.005$) across time points (0.5, 1, 3, 6, and 24 h) was listed in the left side of the table. Less than 137 genes changed with isopropyl alcohol across the time period observed, and more than 2600 genes changed with a dose of 6 mg/kg of SM.

TABLE 1
Dose- and Time-Dependent Gene Alterations after Injection of SM

Doses (mg/kg)	Number of genes changed across time ($p \leq 0.005$)	Time (h)	Number of genes changed across dose ($p \leq 0.005$)
0	137	0.5	49
1	957	1	54
3	570	3	133
6	2604	6	737
		24	441

Note. The number of significantly changed genes across doses (isopropyl alcohol, 1, 3, or 6 mg/kg) for each time point (0.5, 1, 3, 6, and 24 h) or the number of significantly changed genes across time points (0.5, 1, 3, 6, and 24 h) for each dose (saline, isopropyl alcohol, 1, 3, or 6 mg/kg) was identified by one-way ANOVA comparisons at p value ≤ 0.005.

Dose-Dependent Alteration of Gene Expression after Injection of SM

Hierarchical clustering analysis. Figure 2 shows the hierarchical cluster analysis of gene expression alterations across dose. Significant gene expression changes were seen in a dose-dependent manner after injections of SM at 6 (A) and 24 h (B). Genes significantly changed across doses ($p \leq 0.005$) at 6 (737 genes) or 24 h (441 genes) after injection were chosen to conduct the hierarchical clustering analysis. Examples of dose-dependent changes included downregulation of protein-tyrosine phosphatase receptor type G and upregulation of cyclin-dependent kinase inhibitor 1A and cyclin G1 transcription after 6 h postinjection (A and B clusters of Fig. 2A). Additionally, dose-dependent downregulation of T-cell receptor beta-chain and granzyme M and upregulation of cyclin-dependent kinase inhibitor 1A and cyclin G1 transcription were observed 24 h postinjection (A and B clusters in Fig. 2B). However, based solely on the above statistical analysis, it is difficult to interpret these example changes in gene expression within a biological context.

GO analysis of SM-induced genes alterations. We used MAPPFinder to find significant changes within each GO functional category. MAPPFinder linked the 11,212 probe sets measured in this experiment to the GO database. A total of 4485 probes were linked to an Ensembl ID and 3144 genes linked to a GO term in the hierarchy. MAPPFinder calculates a Z score as well as a permutation test p value for the genes that are significantly altered across doses ($p \leq 0.005$) at 6 or 24 h, or across time points ($p \leq 0.001$) at a dose of 6 mg/kg. MAPPFinder calculated the percentage of genes meeting the criteria of either a statistically significant increase or decrease in gene expression across dose (or time). In addition, a Z score for each GO term was calculated based on Formula 1 to identify which GO terms had a significant number of altered genes. Since there are two directions of gene expression alteration

after treatment at certain a gene category, the separation of the genes based on their trend (either upregulation or downregulation) of average gene expression pattern across dose (or time) allows to quantitatively evaluate the degree of changes across dose or time. In this study we identified two simple patterns of either consistent upregulation or consistent downregulation for gene expression across doses (or time) within GO terms based on K-means cluster analysis. MAPPFinder results represent a global picture of biological processes, cellular components, and molecular functions that are significantly altered (upregulated or downregulated) at the transcriptional level after treatment (Tables 2 and 3). Significant downregulations of genes involved in protein modification (GOID 6464), such as phosphorylation (GOID 6468) and signaling transduction (GOID 7165, 8277, 7242, 9966, and 9968), were observed based on biological process gene category 6 h after SM injection (Table 2). In contrast, significant upregulation of genes involved in the cell cycle (GOID 7049) was observed based on biological process gene category 6 h after SM injection (Table 2). As shown in Table 3, in the biological process, genes in cell cycle, especially in M phase, such as M phase of mitotic cell cycle (GOID 87), mitosis (GOID 7067), cell division (GOID 51301), M phase (GOID 279), and mitotic cell cycle (GOID 278), were upregulated. Transferase activity (GOID 16746) was also upregulated. Significant downregulation in immune responses (GOID 6955) such as cell activation (GOID 1775) and cell adhesion (GOID 7155) was observed.

Consistent with the previously published analysis (Dillman *et al.*, 2005), we identified alterations in the cell cycle regulation functional gene category as the most striking changes resulting from SM exposure. Furthermore, GO analysis suggests that multiple biological pathways are involved in SM-induced toxic effects. A recent report comparing global gene expression analyses between laboratories and across platforms suggested that GO analysis is a more biologically meaningful and statistically robust approach to data analysis than relying on single-gene analyses (Bammler *et al.*, 2005). Our GO analysis demonstrates that unsupervised GO mapping and subsequent statistical significance analysis constitute a powerful unbiased approach to defining the biological pathways of known and novel SM-induced molecular changes in the rat lung. We noticed in MAPPFinder analysis that among the 11,212 probe sets measured in this experiment, only 4485 probe sets linked to an Ensembl ID and 3144 genes linked to a GO term. Thus, the biological significance of a large number of rat genes still needs to be defined. Once the annotation of these rat genes is completed, GO mapping and subsequent statistical significance analysis of microarray data will prove to be a powerful unbiased approach in defining biological pathways.

Quantitative Integration of GO Analysis to Evaluate Dose Response

Toxicogenomic data can provide profoundly important and novel information to the field of risk assessment. However, we
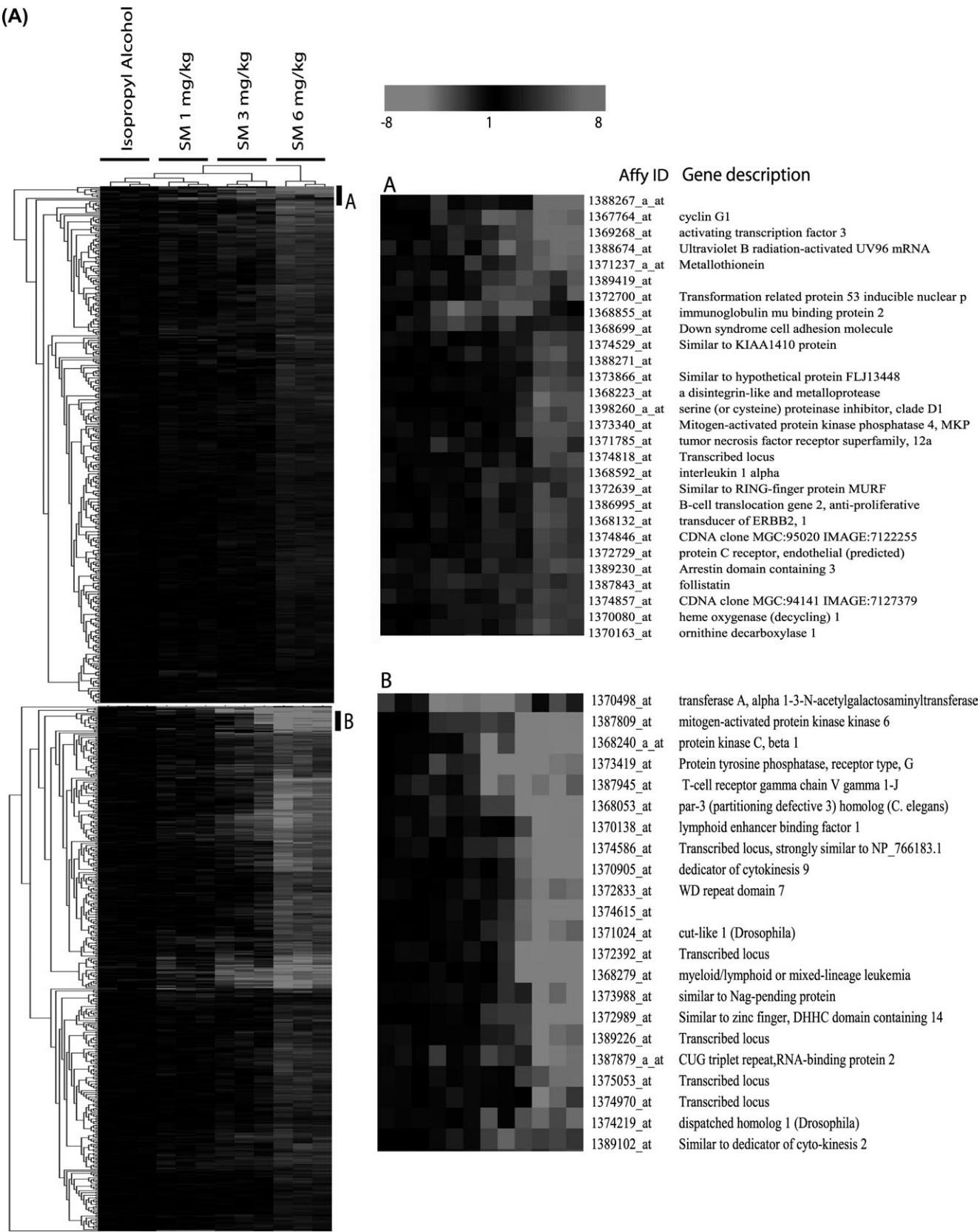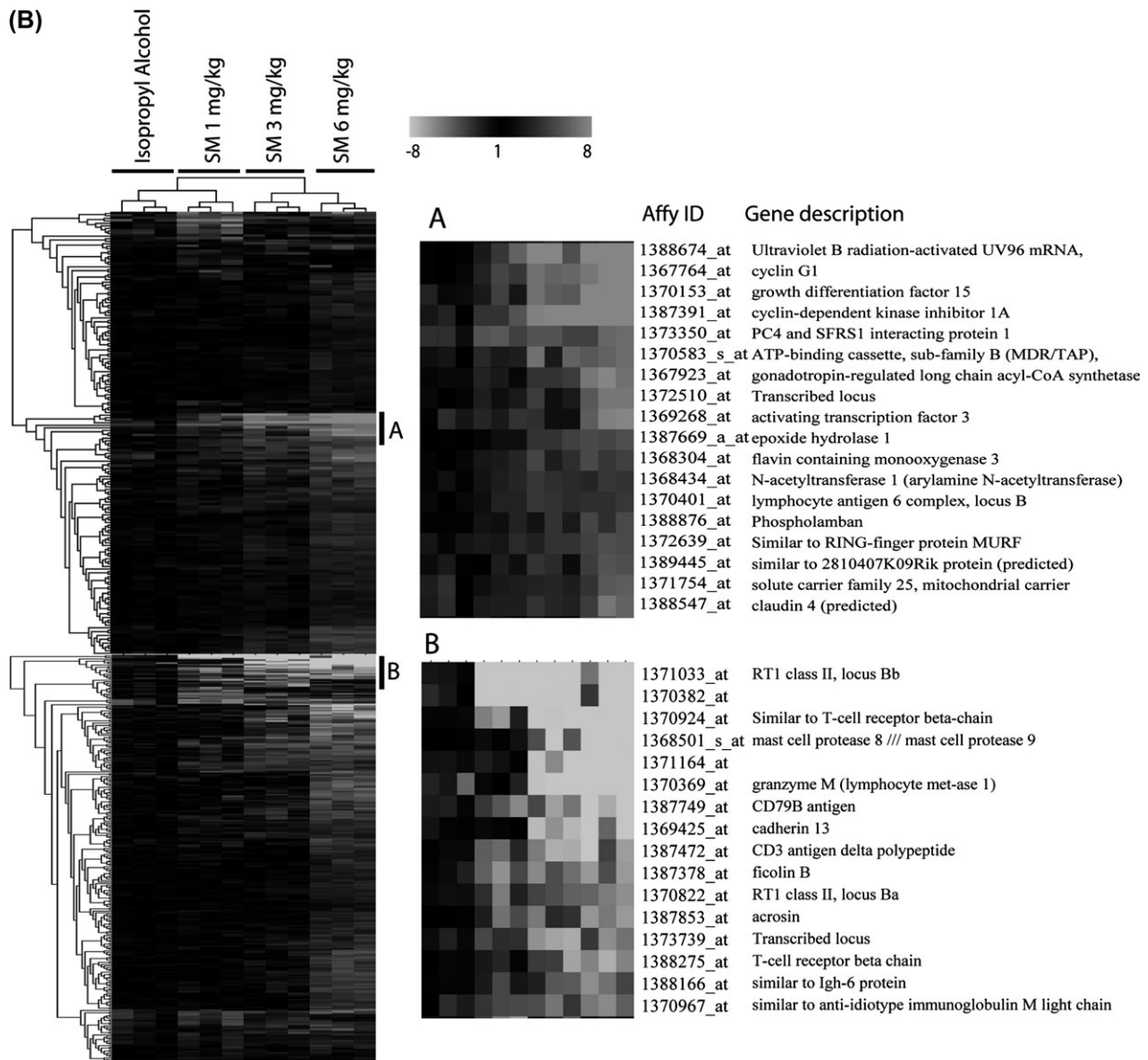
(A)

FIG. 2—*Continued.*

FIG. 2. Hierarchical cluster analysis of dose response gene expression alteration after injection of SM at 6 (A) and 24 h (B). Genes significantly changed across doses at 6 (717 genes) or 24 h (441 genes) after injection were included in hierarchical clustering analysis using average linkage and Euclidean dissimilarity methods. Clusters A and B show the details of genes including Affymetrix probe ID and gene name.

are currently facing the difficult challenge of how to quantitatively interpret dose- or time-dependent relationships resulting from genomic approaches. As presented, the conventional GO analysis does not provide any quantitative information about microarray data in a dose-dependent manner, and no methodology is available to achieve quantitative analysis of pathway-specific information. Thus, we developed a framework to evaluate functional genes from microarray data, as illustrated in Figure 1, and propose the use of such data within a statistically significant functional gene category, such as a specific GO term in the GO database. The above calculation includes a large number of genes within each functional gene category, and there is currently no tool available to conduct such a task. We developed a computer program called GO-Quant, which automatically links functional GO analysis results from MAPPFinder with the original microarray data set, to calculate the average intensity or ratio for those significant genes within each GO term. Figure 3 shows the hierarchical cluster analysis of quantitative GO results across SM treatments after 6 (Fig. 3A) or 24 h (Fig. 3B). Compared to the conventional qualitative GO analysis shown in Tables 2 and 3, Figure 3 shows the dose-dependent alterations of each functional gene category (GOID) following SM treatment. Cell cycle genes (GOID 7049) were dose dependently upregulated 6 h after SM injection (Fig. 3A), and cell cycle alterations, especially in M phase pathway genes (GOID 87,

TABLE 2
Statistically Significantly Changed Gene Ontology Terms across Dose at 6 h

GOID	Path	GO name	Number changed	Number measured	Number in GO	Z score	Permutation p
UP_REGULATION							
Biological process							
7049	0.0.2.2.1	Cell cycle	8	116	204	2.8	0.02
74	0.0.2.2.1.6	Regulation of progression through cell cycle	5	67	114	2.4	0.05
16043	0.0.2.2.7	Cell organization and biogenesis	14	294	689	2.2	0.05
16044	0.0.2.2.7.3	Membrane organization and biogenesis	3	9	10	5.6	0.00
6457	0.0.2.2.9.9.2.8	Protein folding	5	48	138	3.3	0.01
46394	0.0.2.2.9.21.0.4	Carboxylic acid biosynthesis	3	23	31	3.0	0.02
6633	0.0.2.2.9.21.0.4.1	Fatty acid biosynthesis	3	22	30	3.1	0.02
16053	0.0.2.2.9.21.1	Organic acid biosynthesis	3	23	31	3.0	0.02
51244	0.0.2.2.15	Regulation of cellular physiological process	23	518	1186	2.6	0.02
6913	0.0.2.2.17.8.8	Nucleocytoplasmic transport	4	23	44	4.3	0.00
6869	0.0.2.2.17.10	Lipid transport	3	21	42	3.2	0.03
15031	0.0.2.2.17.15	Protein transport	9	163	348	2.2	0.04
6605	0.0.2.2.17.15.0.1	Protein targeting	4	32	70	3.4	0.01
50794	0.0.2.4	Regulation of cellular process	23	554	1247	2.2	0.04
43285	0.0.6.6.1.2.0	Biopolymer catabolism	4	41	97	2.8	0.03
30163	0.0.6.6.1.2.0.1	Protein catabolism	4	39	91	2.9	0.02
50791	0.0.6.8	Regulation of physiological process	24	560	1240	2.5	0.03
42221	0.0.9.1.1	Response to chemical stimulus	6	87	130	2.4	0.03
Cellular component							
5622	0.1.0.8	Intracellular	48	1348	3755	2.5	0.03
5829	0.1.0.8.2.4	Cytosol	6	85	121	2.5	0.03
15630	0.1.0.8.3.4	Microtubule cytoskeleton	4	45	110	2.5	0.05
43229	0.1.0.8.7	Intracellular organelle	44	1167	3089	2.7	0.01
43231	0.1.0.8.7.1	Intracellular membrane-bound organelle	42	1017	2426	3.3	0.00
5634	0.1.0.8.7.1.6	Nucleus	27	565	1641	3.3	0.00
43226	0.1.4	Organelle	44	1167	3089	2.7	0.01
43227	0.1.4.1	Membrane-bound organelle	42	1017	2426	3.3	0.00
Molecular function							
46914	0.2.1.10.0.2	Transition metal ion binding	16	320	1094	2.6	0.02
5506	0.2.1.10.0.2.3	Iron ion binding	7	109	211	2.4	0.04
51082	0.2.1.22.45	Unfolded protein binding	6	34	94	5.3	0.00
16881	0.2.2.9.1.0	Acid–amino acid ligase activity	4	49	228	2.3	0.04
16705	0.2.2.11.12	Oxidoreductase activity, with reduction of molecular oxygen	5	52	100	3.1	0.02
DOWN_REGULATION							
Biological process							
7154	0.0.2.0	Cell communication	48	856	2692	4.3	0.00
7165	0.0.2.0.4	Signal transduction	39	701	2368	3.7	0.00
8277	0.0.2.0.4.0.3.7	Regulation of G-protein–coupled receptor protein signaling pathway	3	14	17	3.8	0.01
7242	0.0.2.0.4.1	Intracellular signaling cascade	20	246	503	4.3	0.00
9966	0.0.2.0.4.2	Regulation of signal transduction	4	35	57	2.7	0.04
9968	0.0.2.0.4.2.0	Negative regulation of signal transduction	3	22	32	2.7	0.04
6464	0.0.2.2.9.9.2.9	Protein modification	20	329	960	2.9	0.01
9187	0.0.2.2.9.19.4.0	Cyclic nucleotide metabolism	3	15	23	3.6	0.01
9190	0.0.2.2.9.19.4.0.2	Cyclic nucleotide biosynthesis	3	15	22	3.6	0.01
6468	0.0.2.2.9.24.0.1.3	Protein amino acid phosphorylation	17	166	407	5.1	0.00
15674	0.0.2.2.17.9.1.0	Di-, trivalent inorganic cation transport	5	54	88	2.4	0.05

TABLE 2—*Continued*

GOID	Path	GO name	Number changed	Number measured	Number in GO	Z score	Permutation p
6897	0.0.2.2.17.18.0	Endocytosis	4	34	44	2.7	0.03
43412	0.0.6.6.3.0.2	Biopolymer modification	20	337	994	2.8	0.02
Molecular function							
8289	0.2.1.12	Lipid binding	7	66	116	3.3	0.01
19992	0.2.1.12.0	Diacylglycerol binding	4	16	21	4.8	0.00
166	0.2.1.16	Nucleotide binding	30	454	1252	4.2	0.00
17076	0.2.1.16.2	Purine nucleotide binding	28	420	1145	4.1	0.00
5524	0.2.1.16.2.0.3	ATP binding	23	313	888	4.2	0.00
5515	0.2.1.22	Protein binding	40	731	1482	3.7	0.00
5516	0.2.1.22.6	Calmodulin binding	8	51	69	4.9	0.00
3779	0.2.1.22.14.0	Actin binding	5	48	114	2.7	0.04
5126	0.2.1.22.37.0.4	Hematopoietin/interferon-class cytokine receptor binding	3	19	32	3.0	0.03
4518	0.2.2.6.4.1	Nuclease activity	4	29	89	3.1	0.02
16740	0.2.2.15	Transferase activity	23	422	1100	2.6	0.02
16772	0.2.2.15.6	Transferase activity, transferring phosphorus-containing groups	19	262	655	3.7	0.01
16301	0.2.2.15.6.1	Kinase activity	19	239	570	4.1	0.00
4672	0.2.2.15.6.1.17	Protein kinase activity	17	170	419	5.0	0.00
4674	0.2.2.15.6.1.17.4	Protein serine/threonine kinase activity	14	126	333	4.9	0.00
4683	0.2.2.15.6.1.17.4.1	Calmodulin-regulated protein kinase activity	3	12	13	4.2	0.01
4702	0.2.2.15.6.1.17.4.9	Receptor-signaling protein serine/threonine kinase activity	3	20	28	2.9	0.04
4713	0.2.2.15.6.1.17.6	Protein-tyrosine kinase activity	7	67	250	3.3	0.01
16773	0.2.2.15.6.3	Phosphotransferase activity, alcohol group as acceptor	18	205	497	4.5	0.00
8047	0.2.4.1	Enzyme activator activity	6	37	61	4.4	0.00
5096	0.2.4.1.1	GTPase activator activity	6	24	44	5.9	0.00
30695	0.2.4.3	GTPase regulator activity	9	45	106	6.3	0.00
5083	0.2.4.3.1	Small GTPase regulator activity	4	26	76	3.4	0.01
5085	0.2.4.3.1.2	Guanyl-nucleotide exchange factor activity	3	20	55	2.9	0.03
16524	0.2.8.1.8.10	Latrotoxin receptor activity	3	5	8	7.0	0.00
5057	0.2.8.3	Receptor-signaling protein activity	4	36	49	2.6	0.05
5262	0.2.12.10.2.2.2	Calcium channel activity	4	29	42	3.1	0.03
5245	0.2.12.10.2.2.2.1	Voltage-gated calcium channel activity	3	12	18	4.2	0.01
Cellular component							
5856	0.1.0.8.3	Cytoskeleton	11	135	337	3.2	0.01
16459	0.1.0.8.3.0.5	Myosin	3	20	39	2.9	0.03
30054	0.1.0.10.8.4	Cell junction	4	29	64	3.1	0.02

Note. MAPPFinder was used to identify the differentially expressed GO terms based on the Z score (≥ 2) and permutation p value (≤ 0.05). The GO identify number (GOID), hierarchical tree (Path), GO term name (GO name), the number of genes changed within each functional gene category, number of genes measured on the array and number of the genes in the GO database, statistical significance value Z score and permutation p value were listed. GO terms with at least 3 genes changed and permutation test p value ≤ 0.05 were listed in table.

279, 278, 7067, and 51301), were dose dependently upregulated 24 h after SM injection (Fig. 3B). Dose-dependent downregulation of protein modification gene pathways, such as phosphorylation, were observed at the 6 h time point, while dose-dependent downregulation of immune response gene pathway (GOID 6952, 1775, and 45321) was observed at 24 h after SM injection (Fig. 3B). Figures 4 and 5 further demonstrate the dose response of quantitative gene expression pathway alterations within each functional gene category in biological process (Fig. 4) and molecular function genes (Fig. 5). Significant upregulation in cell cycle gene pathways becomes obvious at 3 mg/kg both at 6 (Fig. 4A) and at 24 h (Fig. 5A) after injection. Our results confirm that the above system approach is a more robust and clear

TABLE 3
Statistically Significantly Changed GO Terms across Dose at 24 h

GOID	Path	GO name	Number changed	Number measured	Number in GO	Z score	Permutation p
UP_REGULATION							
Biological process							
279	0.0.2.2.1.3	M phase	3	27	46	2.8	0.03
87	0.0.2.2.1.3.1	M phase of mitotic cell cycle	3	21	36	3.4	0.02
278	0.0.2.2.1.5	Mitotic cell cycle	3	26	44	2.9	0.02
51301	0.0.2.2.3	Cell division	3	28	52	2.8	0.03
46164	0.0.2.2.9.0.1	Alcohol catabolism	3	25	72	3.0	0.05
46365	0.0.2.2.9.0.1.0	Monosaccharide catabolism	3	24	68	3.1	0.04
19320	0.0.2.2.9.0.1.0.0	Hexose catabolism	3	24	68	3.1	0.04
6007	0.0.2.2.9.0.1.0.0.0	Glucose catabolism	3	23	67	3.2	0.04
6096	0.0.2.2.9.0.1.0.0.0.0	Glycolysis	3	21	61	3.4	0.02
44275	0.0.2.2.9.7.5.0	Cellular carbohydrate catabolism	3	28	79	2.8	0.05
51186	0.0.2.2.9.10	Cofactor metabolism	4	49	117	2.5	0.04
6351	0.0.2.2.9.19.9.1	Transcription, DNA dependent	14	300	799	2.5	0.04
6790	0.0.2.2.9.27	Sulfur metabolism	3	11	16	5.2	0.01
51244	0.0.2.2.15	Regulation of cellular physiological process	22	517	1186	2.7	0.02
6355	0.0.2.2.15.9.6.5.3	Regulation of transcription, DNA dependent	14	300	790	2.5	0.04
50794	0.0.2.4	Regulation of cellular process	24	553	1247	3.0	0.01
9790	0.0.3.4	Embryonic development	3	10	15	5.5	0.00
16052	0.0.6.6.1.2.1	Carbohydrate catabolism	3	28	80	2.8	0.05
50791	0.0.6.8	Regulation of physiological process	23	558	1240	2.6	0.02
43118	0.0.6.8.0	Negative regulation of physiological process	6	94	147	2.4	0.03
50789	0.0.7	Regulation of biological process	25	613	1355	2.7	0.02
Molecular function							
16747	0.2.2.15.0.2	Transferase activity, other than amino-acyl groups	3	28	74	2.8	0.04
8415	0.2.2.15.0.2.0	Acyltransferase activity	3	28	74	2.8	0.04
4857	0.2.4.2	Enzyme inhibitor activity	5	65	148	2.7	0.04
Cellular component							
19867	0.1.0.10.7	Outer membrane	3	13	22	4.7	0.00
DOWN_REGULATION							
Biological process							
7155	0.0.2.0.0	Cell adhesion	8	116	252	2.5	0.03
9966	0.0.2.0.4.2	Regulation of signal transduction	5	35	57	4.0	0.00
9968	0.0.2.0.4.2.0	Negative regulation of signal transduction	4	22	32	4.2	0.00
6508	0.0.2.2.9.9.2.12	Proteolysis	11	173	458	2.7	0.01
6817	0.0.2.2.17.9.0.0.2	Phosphate transport	3	22	62	3.0	0.04
1775	0.0.6.7.2	Cell activation	3	20	27	3.2	0.02
45321	0.0.6.7.2.0	Immune cell activation	3	20	27	3.2	0.02
6955	0.0.6.7.6	Immune response	9	141	258	2.4	0.02
6952	0.0.9.2.1	Defense response	9	155	289	2.1	0.05
9613	0.0.9.2.3.2	Response to pest, pathogen or parasite	6	87	124	2.2	0.05
Molecular function							
5509	0.2.1.10.0.1	Calcium ion binding	15	256	538	2.8	0.02
8233	0.2.2.6.7	Peptidase activity	12	181	478	3.0	0.01
4175	0.2.2.6.7.2	Endopeptidase activity	11	142	352	3.4	0.00
4252	0.2.2.6.7.2.5	Serine-type endopeptidase activity	8	61	151	4.7	0.00
4263	0.2.2.6.7.2.5.3	Chymotrypsin activity	6	40	119	4.5	0.00
4295	0.2.2.6.7.2.5.20	Trypsin activity	6	43	128	4.3	0.00
8236	0.2.2.6.7.5	Serine-type peptidase activity	8	65	164	4.5	0.00
Cellular component							
12505	0.1.0.10.1	Endomembrane system	5	57	92	2.6	0.04
19898	0.1.0.10.2	Extrinsic to membrane	3	18	26	3.4	0.02

Note. MAPPFinder was used to identify the differentially expressed GO terms based on the Z score (≥ 2) and permutation p value (≤ 0.05). The GOID, hierarchical tree (Path), GO term name (GO name), the number of genes changed within each functional gene category, number of genes measured on the array and number of the genes in the GO database, statistical significance value Z score, and permutation p value were listed. GO terms with at least three genes changed and permutation test p value ≤ 0.05 were listed in the table.

(A)

Columns: 0 mg/Kg, 1 mg/kg, 3 mg/kg, 6 mg/kg

Grayscale bar: 4 — 1 — 4

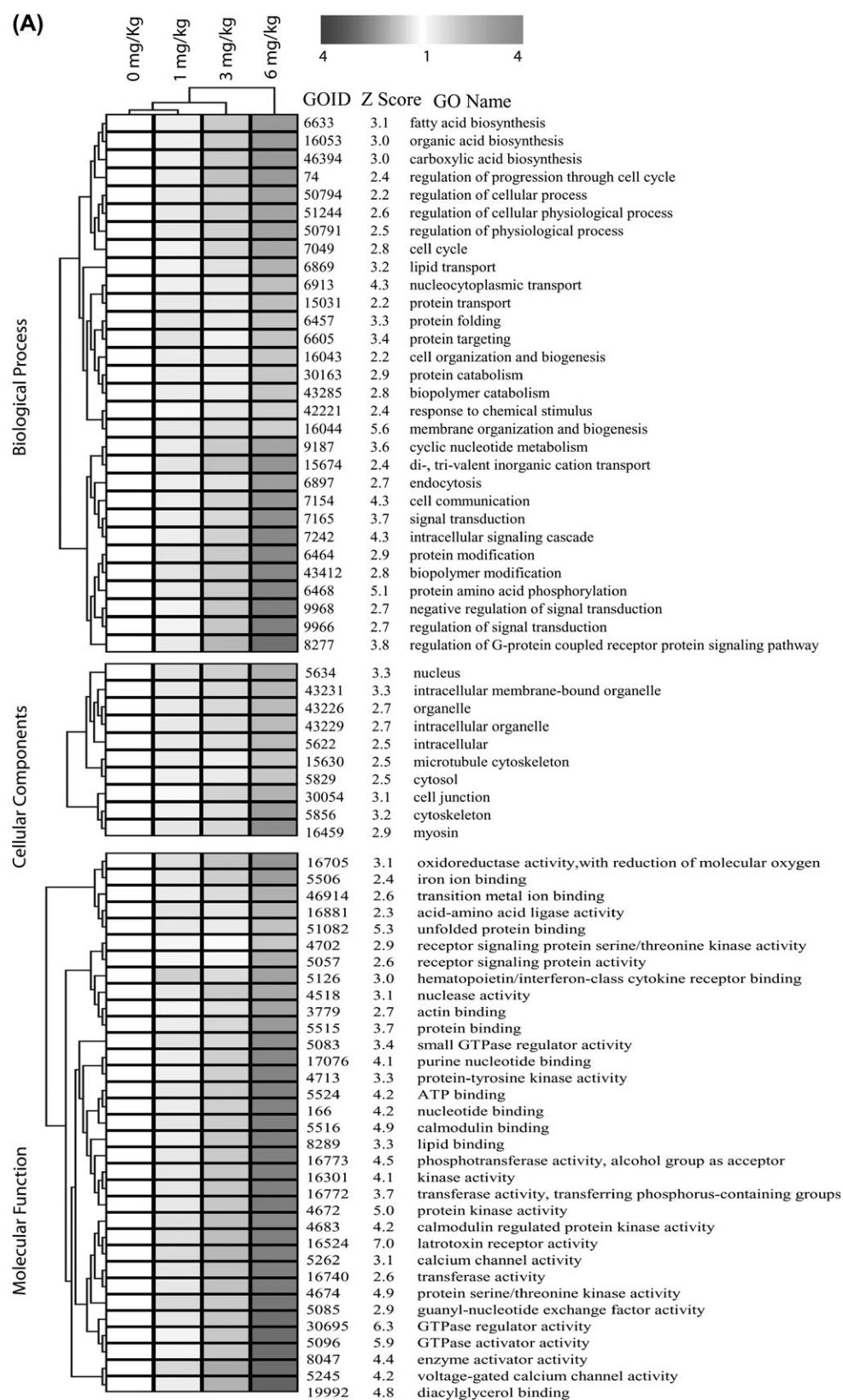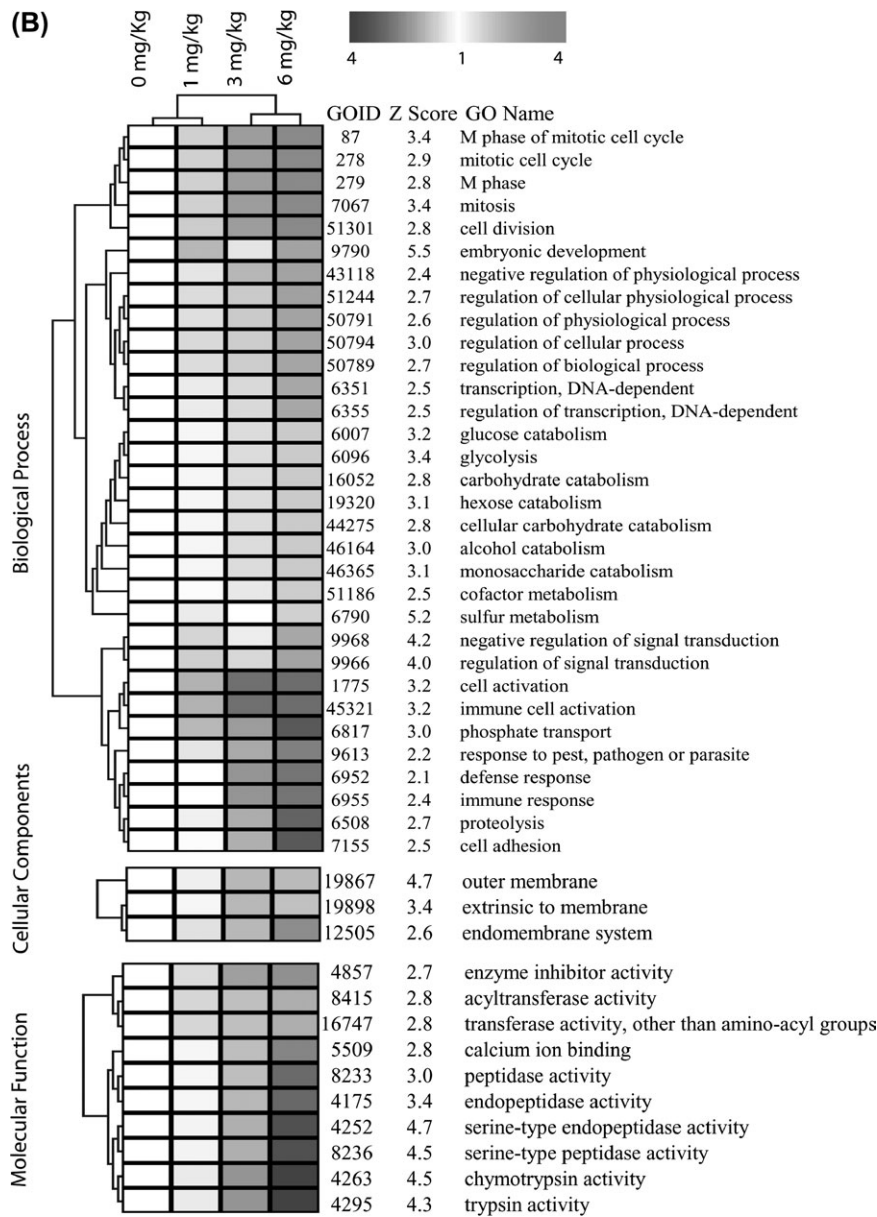| GOID	Z Score	GO Name
		Biological Process
6633	3.1	fatty acid biosynthesis
16053	3.0	organic acid biosynthesis
46394	3.0	carboxylic acid biosynthesis
74	2.4	regulation of progression through cell cycle
50794	2.2	regulation of cellular process
51244	2.6	regulation of cellular physiological process
50791	2.5	regulation of physiological process
7049	2.8	cell cycle
6869	3.2	lipid transport
6913	4.3	nucleocytoplasmic transport
15031	2.2	protein transport
6457	3.3	protein folding
6605	3.4	protein targeting
16043	2.2	cell organization and biogenesis
30163	2.9	protein catabolism
43285	2.8	biopolymer catabolism
42221	2.4	response to chemical stimulus
16044	5.6	membrane organization and biogenesis
9187	3.6	cyclic nucleotide metabolism
15674	2.4	di-, tri-valent inorganic cation transport
6897	2.7	endocytosis
7154	4.3	cell communication
7165	3.7	signal transduction
7242	4.3	intracellular signaling cascade
6464	2.9	protein modification
43412	2.8	biopolymer modification
6468	5.1	protein amino acid phosphorylation
9968	2.7	negative regulation of signal transduction
9966	2.7	regulation of signal transduction
8277	3.8	regulation of G-protein coupled receptor protein signaling pathway
		Cellular Components
5634	3.3	nucleus
43231	3.3	intracellular membrane-bound organelle
43226	2.7	organelle
43229	2.7	intracellular organelle
5622	2.5	intracellular
15630	2.5	microtubule cytoskeleton
5829	2.5	cytosol
30054	3.1	cell junction
5856	3.2	cytoskeleton
16459	2.9	myosin
		Molecular Function
16705	3.1	oxidoreductase activity,with reduction of molecular oxygen
5506	2.4	iron ion binding
46914	2.6	transition metal ion binding
16881	2.3	acid-amino acid ligase activity
51082	5.3	unfolded protein binding
4702	2.9	receptor signaling protein serine/threonine kinase activity
5057	2.6	receptor signaling protein activity
5126	3.0	hematopoietin/interferon-class cytokine receptor binding
4518	3.1	nuclease activity
3779	2.7	actin binding
5515	3.7	protein binding
5083	3.4	small GTPase regulator activity
17076	4.1	purine nucleotide binding
4713	3.3	protein-tyrosine kinase activity
5524	4.2	ATP binding
166	4.2	nucleotide binding
5516	4.9	calmodulin binding
8289	3.3	lipid binding
16773	4.5	phosphotransferase activity, alcohol group as acceptor
16301	4.1	kinase activity
16772	3.7	transferase activity, transferring phosphorus-containing groups
4672	5.0	protein kinase activity
4683	4.2	calmodulin regulated protein kinase activity
16524	7.0	latrotoxin receptor activity
5262	3.1	calcium channel activity
16740	2.6	transferase activity
4674	4.9	protein serine/threonine kinase activity
5085	2.9	guanyl-nucleotide exchange factor activity
30695	6.3	GTPase regulator activity
5096	5.9	GTPase activator activity
8047	4.4	enzyme activator activity
5245	4.2	voltage-gated calcium channel activity
19992	4.8	diacylglycerol binding

FIG. 3—*Continued.*

FIG. 3. GO-based hierarchical cluster analyses of gene expression data across dose after injection of SM at 6 (A) and 24 h (B). GO analysis was performed using MAPPFinder. The average ratio of the genes in a statistically significant functional gene category (GOID in GO database) was calculated with GO-Quant. The output file, including GO term, Z score, average intensity or ratio, and gene ID, was imported to TIGR MEV (Saeed *et al.*, 2003) and a hierarchical cluster analysis conducted using average linkage and Euclidean dissimilarity methods (Eisen *et al.*, 1998).

measurement to determine the degree to which a functional gene system changes across dose. As an alternative to representing annotations by the average of the annotated genes, the sum or the median expression profile could also be calculated using the GO-Quant. In most cases, the resulting plots are virtually identical (data not shown). Such an approach gives one the ability to calculate the ED_{50} for a specific functional GO term, which is important for risk assessment.

Time Course Gene Alteration after a Single Injection of SM at a Dose of 6 mg/kg

In order to demonstrate the above application to the time course within this microarray data set, we further applied the above approach as illustrated in Figure 1 to analyze the time course data after the single injection of SM at a dose of 6 mg/kg. Figure 6 shows the hierarchical clustering analysis of 1111 genes whose transcription was significantly ($p \leq 0.001$) altered across the selected time points (6 mg/kg SM). Using
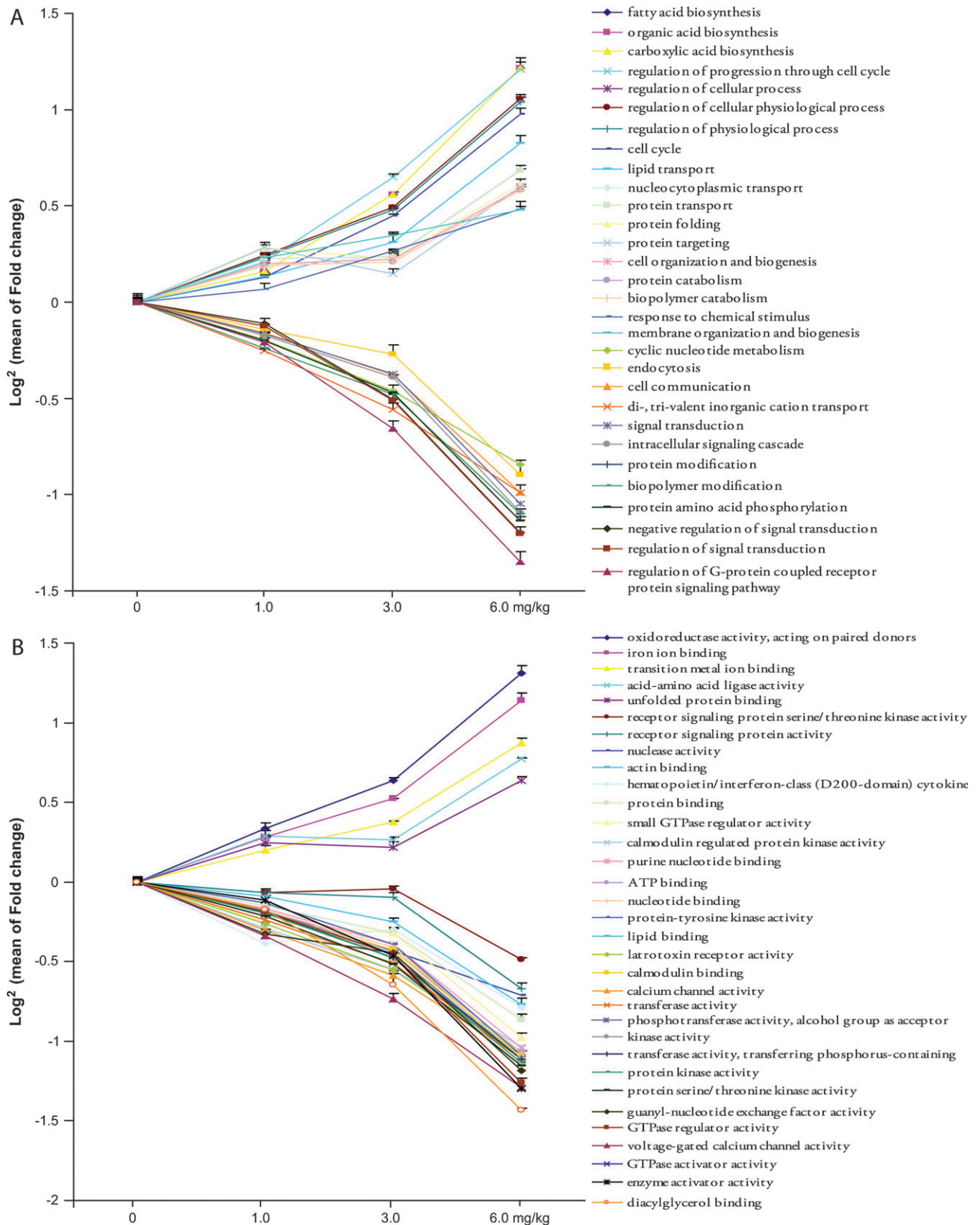
FIG. 4. Dose response of GOID in biological process (A) and molecular function gene categories (B) 6 h after SM injection. GO analysis was performed using MAPPFinder. The average ratio of the genes in a statistically significant functional gene category (GOID in GO database) was calculated with GO-Quant.
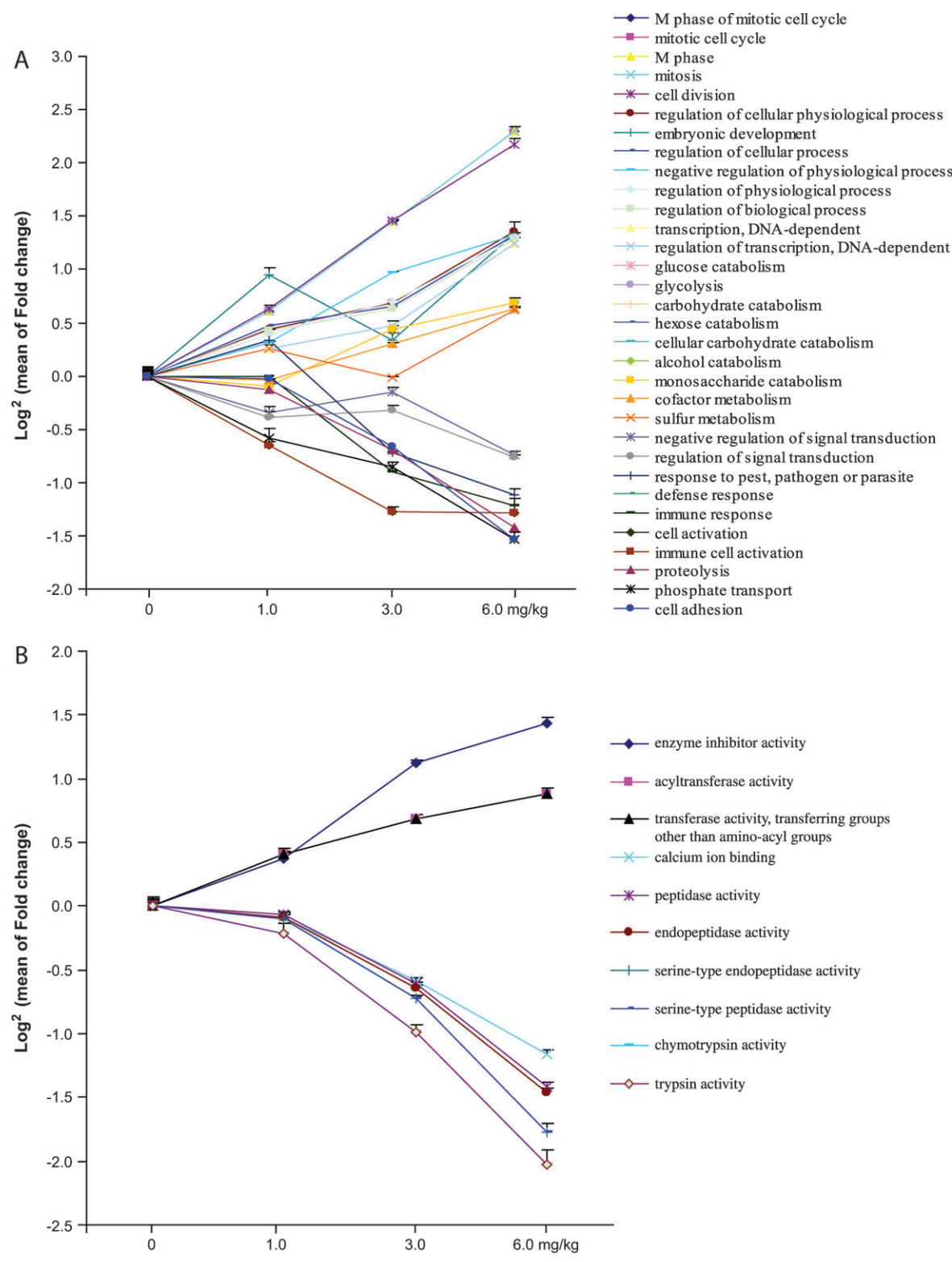
FIG. 5. Dose response of GOID in biological process (A) and molecular function gene categories (B) 24 h after SM injection. GO analysis was performed using MAPPFinder. The average ratio of the genes in a statistically significant functional gene category (GOID in GO database) was calculated with GO-Quant.
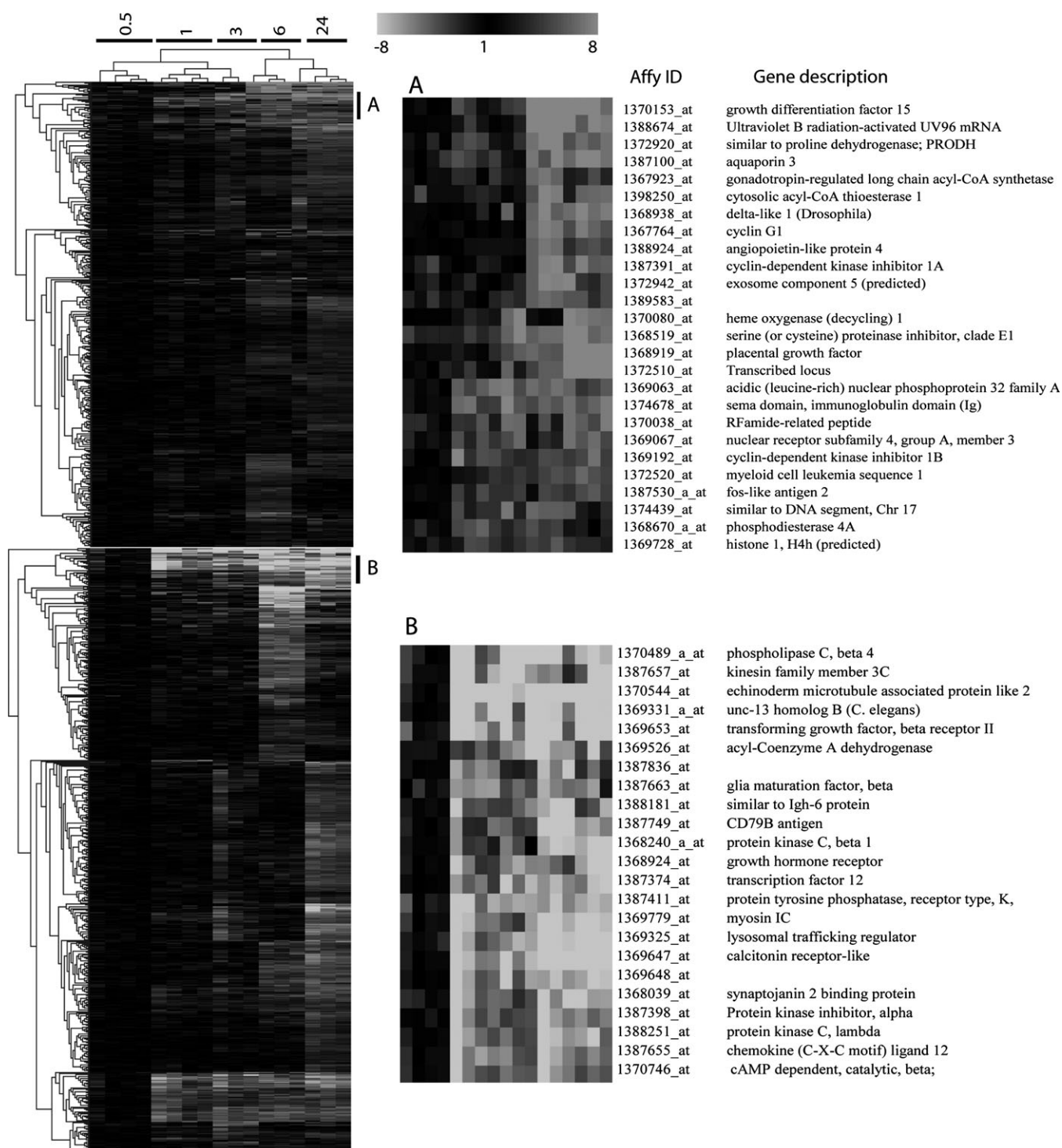
FIG. 6. Hierarchical cluster analysis of altered gene expression in a time-dependent manner after injection of SM at a dose of 6 mg/kg. Genes significantly changed across time points after exposure were included in hierarchical cluster analysis using average linkage and Euclidean dissimilarity methods. Clusters A and B show the details of genes including Affymetrix probe ID and gene name.

cluster analysis, we identified changes in transcription in specific genes, e.g., upregulation of heme oxygenase 1, cyclin G, and cyclin-dependent kinase inhibitor 1A and downregulation of protein kinase C and chemokine ligand 12. As

discussed, it is challenging to interpret these previous results in a biological context. Table 4 shows the list of GOID from the MAPPFinder analysis representing a global picture of biological processes, cellular components and molecular

TABLE 4
Statistically Significantly Changed GO Terms across Time at a Dose of 6 mg/kg of SM

GOID	Path	GO name	Number changed	Number measured	Number in GO	Z score	Permutation p
UP_REGULATION							
Biological process							
7049	0.0.2.2.1	Cell cycle	12	116	204	2.9	0.02
74	0.0.2.2.1.6	Regulation of progression through cell cycle	10	67	114	3.9	0.00
7050	0.0.2.2.1.6.1.0	Cell cycle arrest	3	12	17	3.3	0.01
42157	0.0.2.2.9.9.2.4	Lipoprotein metabolism	3	11	29	3.5	0.01
16567	0.0.2.2.9.9.2.9.17.0	Protein ubiquitination	5	37	189	2.5	0.03
6805	0.0.2.2.9.30	Xenobiotic metabolism	3	12	12	3.3	0.04
51244	0.0.2.2.15	Regulation of cellular physiological process	38	518	1186	3.0	0.01
6869	0.0.2.2.17.10	Lipid transport	4	21	42	3.1	0.02
50794	0.0.2.4	Regulation of cellular process	39	554	1247	2.7	0.02
9888	0.0.3.14	Tissue development	5	35	45	2.6	0.04
40007	0.0.4	Growth	7	62	90	2.4	0.03
50791	0.0.6.8	Regulation of physiological process	38	560	1240	2.4	0.03
9410	0.0.9.1.1.12	Response to xenobiotic stimulus	3	12	13	3.3	0.04
6950	0.0.9.5	Response to stress	17	208	343	2.4	0.03
Molecular function							
8270	0.2.1.10.0.2.8	Zinc ion binding	16	188	815	2.5	0.03
16874	0.2.2.9	Ligase activity	8	75	329	2.4	0.03
16879	0.2.2.9.1	Ligase activity, forming carbon-nitrogen bonds	8	57	250	3.3	0.01
16881	0.2.2.9.1.0	Acid–amino acid ligase activity	8	49	228	3.8	0.01
4842	0.2.2.9.1.0.4	Ubiquitin-protein ligase activity	6	44	216	2.8	0.02
16651	0.2.2.11.10	Oxidoreductase activity, acting on NADH or NADPH	4	11	42	4.9	0.00
16746	0.2.2.15.0	Transferase activity, transferring acyl groups	5	31	80	3.0	0.02
16747	0.2.2.15.0.2	Transferase activity, transferring groups other than amino-acyl groups	5	28	74	3.2	0.01
8415	0.2.2.15.0.2.0	Acyltransferase activity	5	28	74	3.2	0.01
8374	0.2.2.15.0.2.0.5	O-acyltransferase activity	4	9	16	5.6	0.00
4857	0.2.4.2	Enzyme inhibitor activity	8	65	148	2.9	0.02
Cellular component							
151	0.1.0.8.18	Ubiquitin ligase complex	5	35	187	2.6	0.02
DOWN_REGULATION							
Biological process							
7156	0.0.2.0.0.1.3	Homophilic cell adhesion	5	16	76	3.7	0.00
8277	0.0.2.0.4.0.3.7	Regulation of G-protein–coupled receptor protein signaling pathway	4	14	17	3.1	0.02
51301	0.0.2.2.3	Cell division	5	28	52	2.2	0.05
6333	0.0.2.2.7.4.0.1.0.0	Chromatin assembly or disassembly	4	20	93	2.2	0.05
6471	0.0.2.2.9.9.2.9.5	Protein amino acid ADP-ribosylation	3	8	22	3.3	0.02
1775	0.0.6.7.2	Cell activation	5	20	27	3.0	0.01
45321	0.0.6.7.2.0	Immune cell activation	5	20	27	3.0	0.01
46649	0.0.6.7.2.0.1	Lymphocyte activation	4	15	22	2.9	0.02
42493	0.0.9.1.1.4	Response to drug	3	8	18	3.3	0.02
Molecular function							
19992	0.2.1.12.0	Diacylglycerol binding	4	16	21	2.7	0.04
4263	0.2.2.6.7.2.5.3	Chymotrypsin activity	7	40	119	2.5	0.03
4295	0.2.2.6.7.2.5.20	Trypsin activity	7	43	128	2.3	0.04
16763	0.2.2.15.3.2	Transferase activity, transferring pentosyl groups	3	12	32	2.4	0.05
3950	0.2.2.15.3.2.4	NAD+ ADP-ribosyltransferase activity	3	5	12	4.5	0.00
4674	0.2.2.15.6.1.17.4	Protein serine/threonine kinase activity	16	127	333	2.3	0.05
5245	0.2.12.5.0.1.1.2.1	Voltage-gated calcium channel activity	4	12	18	3.5	0.02
Cellular component							
5891	0.1.0.10.4.1.1.7	Voltage-gated calcium channel complex	3	6	11	4.0	0.02
1772	0.1.0.10.8.9	Immunological synapse	3	11	31	2.5	0.04
42101	0.1.0.10.8.9.1	T-cell receptor complex	3	5	5	4.5	0.00

Note. MAPPFinder was used to identify the differentially expressed GO terms based on the Z score (≥ 2) and permutation p value (≤ 0.05). The GOID, hierarchical tree (Path), GO term name (GO name), the number of genes changed within each functional gene category, number of genes measured on the array and number of the genes in the GO database, statistical significance value Z score, and permutation p value were listed. GO terms with at least three genes changed and permutation test p value ≤ 0.05 were listed in the table.
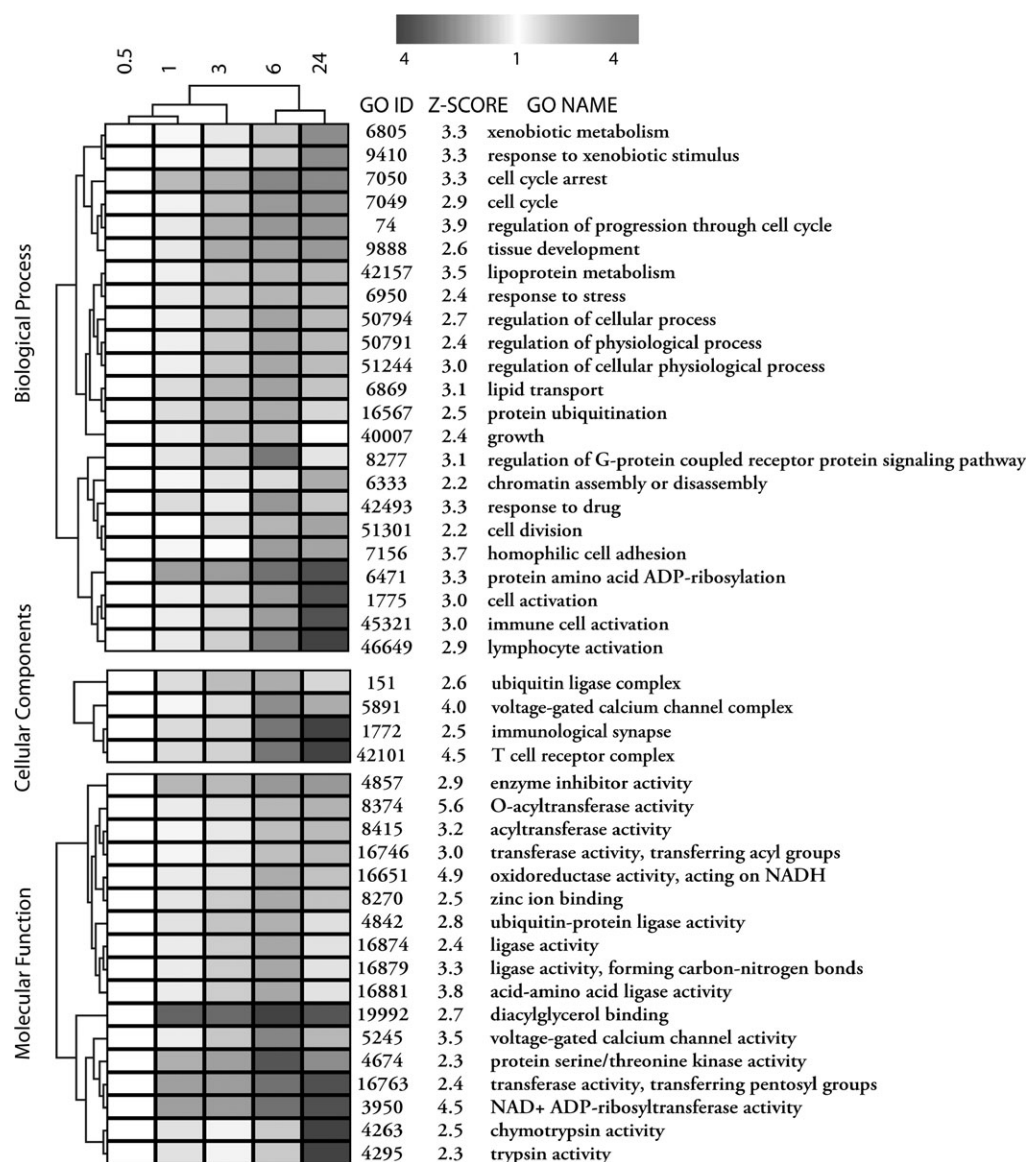
FIG. 7. GO-based hierarchical cluster analyses of gene expression data across time after injection of SM at a dose of 6 mg/kg. GO analysis was performed using MAPPFinder. The average ratio of the genes in a statistically significant functional gene category (GOID in GO database) was calculated with GO-Quant. The output file, including GO term, *Z* score, average intensity or ratio, and gene ID, was imported to TIGR MEV (Saeed *et al.*, 2003) and a hierarchical cluster analysis conducted using average linkage and Euclidean dissimilarity methods (Eisen *et al.*, 1998).

function genes that are significantly altered (upregulated or downregulated) at the transcriptional level following SM exposure at 6 mg/kg (Table 4). As shown in Figures 7 and 8, GO-Quant analysis revealed time-dependent alterations in various functional gene categories within biological process, cellular component, and molecular function genes. In the biological process gene category, time-dependent downregulation in immune cell activation (GOID 45321, 46649, and 1775), cell division (GOID 51310), and response to drug (GOID 42493) were observed, especially at 6 h after SM exposure. In addition, time-dependent upregulations in the regulation of cell cycle (GOID 7050 and 74), protein

ubiquitination (GOID 16567) gene pathways were also observed, particularly at 6 h postexposure. Our results confirm that the above systems approach allows us to demonstrate time-dependent alteration in cellular and molecular gene function based on GO categorization. It is not our focus in this paper to describe why these functional transcriptional alterations after SM injection are observed. Our current analysis is intended to demonstrate that our systematic approach to applying the average data within the statistically significant GOID is a powerful tool to quantitatively describe and interpret dose- or time-dependent relationships discovered using microarray data.
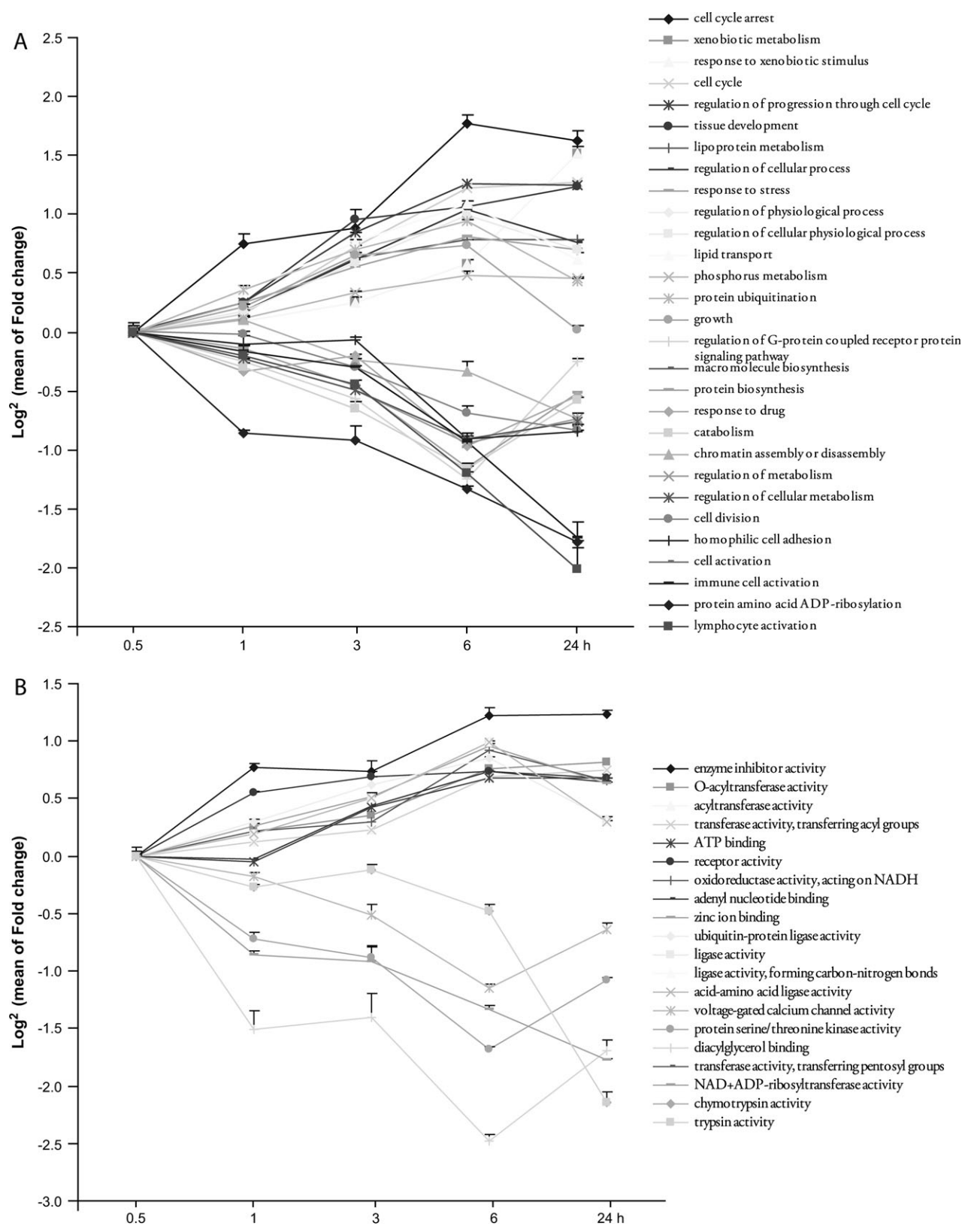
FIG. 8. Time-dependent alteration of GOID in biological process (A) and molecular function genes (B) after SM injection at dose of 6 mg/kg. GO analysis was performed using MAPPFinder. The average ratio of the genes in a statistically significant functional gene category (GOID in GO database) was calculated with GO-Quant.

CONCLUSION

Using gene expression profiling following SM exposure, we have conducted a conventional GO analysis and compared it with our proposed quantitative approach for functional GO analysis. Our results confirmed that GO analysis is a powerful approach to generate an unbiased view of the functional gene alterations by SM. However, the quantitative approach we propose using combined average raw gene expression values (e.g., intensity or ratio) of genes associated with specific functional gene categories derived from the MAPPFinder results in conjunction with the GO-Quant algorithm provides a powerful tool to interpret microarray data in a dose- or time-dependent manner. Additionally, this approach provides a global measurement based on pathway response as compared to single-gene assessments and enables one to calculate the corresponding ED_{50} for each specific functional GO term, which is important for risk assessment. The application of such quantitative interpretation of toxicogenomic data is likely to become increasingly useful for the interpretation of toxicogenomic data generated for evaluating mechanistic similarity of novel chemicals.

ACKNOWLEDGMENTS

REFERENCES

Al-Shahrour, F., Diaz-Uriarte, R., and Dopazo, J. (2004). FatiGO: a web tool for finding significant associations of Gene Ontology terms with groups of genes. *Bioinformatics* **20,** 578–580.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., *et al.* (2000). Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25,** 25–29.

Bammler, T., Beyer, R. P., Bhattacharya, S., Boorman, G. A., Boyles, A., Bradford, B. U., Bumgarner, R. E., Bushel, P. R., Chaturvedi, K., Choi, D., *et al.* (2005). Standardizing global gene expression analysis between laboratories and across platforms. *Nat. Methods* **2,** 351–356.

Beissbarth, T., and Speed, T. P. (2004). GOstat: Find statistically overrepresented gene ontologies within a group of genes. *Bioinformatics* **20,** 1464–1465.

Camon, E., Magrane, M., Barrell, D., Lee, V., Dimmer, E., Maslen, J., Binns, D., Harte, N., Lopez, R., and Apweiler, R. (2004). The gene ontology annotation (GOA) database: Sharing knowledge in Uniprot with gene ontology. *Nucleic Acids Res.* **32,** D262–D266.

Currie, R. A., Bombail, V., Oliver, J. D., Moore, D. J., Lim, F. L., Gwilliam, V., Kimber, I., Chipman, K., Moggs, J. G., and Orphanides, G. (2005). Gene ontology mapping as an unbiased method for identifying molecular pathways and processes affected by toxicant exposure: Application to acute effects caused by the rodent non-genotoxic carcinogen diethylhexylphthalate. *Toxicol. Sci.* **86,** 453–469.

Dennis, G., Jr, Sherman, B. T., Hosack, D. A., Yang, J., Gao, W., Lane, H. C., and Lempicki, R. A. (2003). DAVID: Database for annotation, visualization, and integrated discovery. *Genome Biol.* **4,** 3.

Dillman, J. F., III, Phillips, C. S., Dorsch, L. M., Croxton, M. D., Hege, A. I., Sylvester, A. J., Moran, T. S., and Sciuto, A. M. (2005). Genomic analysis of rodent pulmonary tissue following bis-(2-chloroethyl) sulfide exposure. *Chem. Res. Toxicol.* **18,** 28–34.

Doniger, S. W., Salomonis, N., Dahlquist, K. D., Vranizan, K., Lawlor, S. C., and Conklin, B. R. (2003). MAPPFinder: Using gene ontology and GenMAPP to create a global gene-expression profile from microarray data. *Genome Biol.* **4,** R7.

Eisen, M. B., Spellman, P. T., Brown, P. O., and Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. U.S.A.* **95,** 14863–14868.

Hamadeh, H. K., Bushel, P. R., Jayadev, S., Martin, K., DiSorbo, O., Sieber, S., Bennett, L., Tennant, R., Stoll, R., Barrett, J. C., *et al.* (2002). Gene expression analysis reveals chemical-specific profiles. *Toxicol. Sci.* **67,** 219–231.

Harris, M. A., Clark, J., Ireland, A., Lomax, J., Ashburner, M., Foulger, R., Eilbeck, K., Lewis, S., Marshall, B., Mungall, C., *et al.* (2004). The gene ontology (GO) database and informatics resource. *Nucleic Acids Res.* **32,** D258–D261.

Jansen, R., and Gerstein, M. (2000). Analysis of the yeast transcriptome with structural and functional categories: Characterizing highly expressed proteins. *Nucleic Acids Res.* **28,** 1481–1488.

Long, A. D., Mangalam, H. J., Chan, B. Y., Tolleri, L., Hatfield, G. W., and Baldi, P. (2001). Improved statistical inference from DNA microarray data using analysis of variance and a Bayesian statistical framework. Analysis of global gene expression in Escherichia coli K12. *J. Biol. Chem.* **276,** 19937–19944.

Moggs, J. G. (2005). Molecular responses to xenoestrogens: Mechanistic insights from toxicogenomics. *Toxicology* **213,** 177–193.

Saeed, A. I., Sharov, V., White, J., Li, J., Liang, W., Bhagabati, N., Braisted, J., Klapa, M., Currier, T., Thiagarajan, M., *et al.* (2003). TM4: A free, open-source system for microarray data management and analysis. *Biotechniques* **34,** 374–378.

Soukas, A., Cohen, P., Socci, N. D., and Friedman, J. M. (2000). Leptin-specific patterns of gene expression in white adipose tissue. *Genes Dev.* **14,** 963–980.

Wright, G. W., and Simon, R. M. (2003). A random variance model for detection of differential gene expression in small microarray experiments. *Bioinformatics* **19,** 2448–2455.