BIOCHIP JOURNAL, Vol. 3, No. 4, 293-298, December 2009

HazChem Human Array V3: Classification of Environmental Toxicants through Gene Expression Pattern for Risk Assessment

Seung Jun Kim[1,2], Hye-won Park[2,3], So Yeon Yu[2], Jun-Sub Kim[2], Seung Yong Lee[1], Yu Ri An[1], Youn-Jung Kim[4], Moon-Ju Oh[2] , Jae-Chun Ryu[4] & Seung Yong Hwang[1,2]

[1]Department of Biochemistry, Hanyang University, Sa 1-dong, Sangnok-gu, Ansan, Gyeonggi-do 426-791, Korea
[2]GenoCheck Co. Ltd., Sa 1-dong, Sangnok-gu, Ansan, Gyeonggi-do 426-791, Korea
[3]Department of Bio-Nanotechnology, Hanyang University, Sa 1-dong, Sangnok-gu, Ansan, Gyeonggi-do 426-791, Korea
[4]Cellular and Molecular Toxicology Laboratory, Korea Institute of Science and Technology, PO Box 131, Cheongryang, Seoul 130-650, Korea
Correspondence and requests for materials should be addressed to S.Y. Hwang (syhwang@hanyang.ac.kr)

Accepted 24 November 2009

Abstract

Environmental chemicals such as fungicides, dioxin, or cadmium can cause changes in gene expression. Environmental toxicogenomic approaches using gene expression profiles are useful tools that might be exploited in risk assessments of environmental toxicants from natural sources or as the result of human-made pollution. The principal objective of this study was to compare the gene expression profiles of 17 environmental chemicals-16 type-identified chemicals and 1 obscured chemical-and to identify classifications that better characterized toxicity types by exposure. We then utilized 2 human cell lines, and determined the IC20 values of each. In order to classify the gene expression profiles of the 17 chemicals, we used a custom-made HazChem human array V3, based on previous studies. This array included a total of 1136 genes, all of which were specifically differentially expressed by exposure to VOCs, PAHs, POPs, and LTCs (liver-toxicity chemicals). As a result, we detected 286 of these genes that were differentially expressed by drug type, using a statistical method involving type-parametric Welch's t-test and the Benjamini-Hochberg false discovery rate (FDR-adjusted p-value$<$0.01). However, one type-obscured chemical was shown to have endocrine disruption ability, and evidenced liver-toxicity somewhat close to that of POPs. Additionally, we used an SVM (support vector machines) class prediction method, and then selected 150 genes (prediction strength$>$4.148) that could be used to classify the chemical types via Fisher's exact test. We identified 43 common genes via two methods as powerful class-predictive genes, and confirmed their classifications using PCA. These 43 genes may help in advanced screening chemical for similar toxicogenomic effects with 5 chemical-types.

Keywords: Toxicogenomics, HazChem, Class prediction, Environmental toxicants

Introduction

Humans and ecosystems are susceptible to their environments. Chemicals from natural sources or human-made pollution may exert adverse environmental effects. It has been well established that environmental chemicals such as fungicides, parasites, dioxin, and cadmium may induce changes in gene expression[1-4]. The environmental risk is the chance that harmful effects to human health or to ecological systems will result from exposure to environmental mediators. There is a clear need for risk assessment protocols to characterize the nature and magnitude of health risks to humans and ecological receptors from chemical contaminants and other stressors, which may be present in the environment[5,6]. The field of toxicogenomics combines toxicology with genomics, and uses high-throughput molecular profiling technologies. These environmental toxicogenomic approaches using gene expression profiles may prove useful in risk assessments of environmental toxicants from natural sources or human-made pollution[7]. In this study, we employed 5 types of 17 chemicals: six Volatile Organic Compounds (VOCs), three Polycyclic Aromatic Hydrocarbons (PAHs), two Persistent Organic Pollutants (POPs), two Endocrine-Disrupting Chemicals (EDCs), and four Liver-Toxicity Chemicals (LTCs). In order to evaluate the toxicogenomic effects exerted by these 17 environmental toxicants, we utilized a manufactured HazChem Human array V3. It was manufactured with VOCs, PAHs, and POPs-specific or -common expressed genes, and with liver-toxicity related human

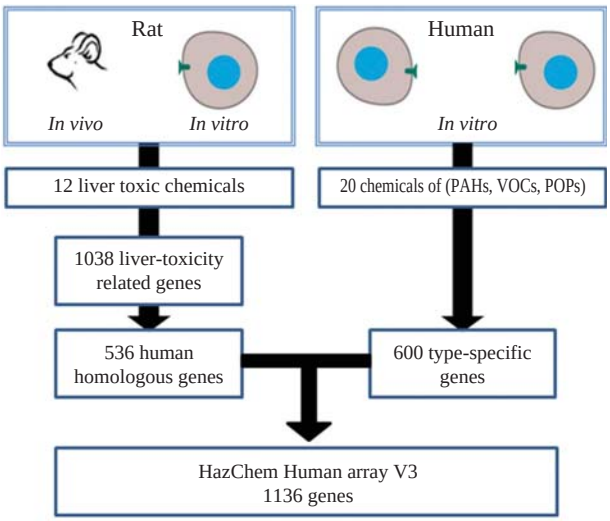


Figure. 1. Contents of HazChem Human array V3. HazChem Human array V3 is composed to 600 genes selected from VOCs and PAHs and POPs specifically expressed on each type of chemicals or commonly expressed on three types, and 536 human homologous genes to 1136 liver-toxicity related genes as shown in previous rat *in vivo* studies.

homologous genes identified in previous studies (Jung *et al*., 2004)[8-10]. We identified type-specific expressed genes and classified the 17 chemicals into each type. In order to select strongly predictive genes, we utilized the statistical method, ANOVA[11], and a class prediction method, SVM[12]. Genes common to the two methods serve as a powerful predictor for the classification of 5 chemical-types.

Results and Discussion

To assess gene expression profiles by these environmental toxicants, we utilized a HazChem Human array V3. The HazChem Human array V3 is composed of 1136 genes selected from VOCs, PAHs, and POPs expressed specifically on each type of chemicals or commonly expressed on the three types as determined in previous studies. It also included liver-toxicity related human homologous genes identified in previous studies (Figure 1). We manufactured our array from an Agilent custom-made array that is 8 blocks per slide and 15 repeat probes per gene.

We conducted microarray experiments of 17 environmental chemicals-16 type-identified chemicals and 1 obscured chemical in human 2 cell lines. Gene expression profiles were analyzed using Agilent Feature Extraction v 10.7 and GeneSpring GX 10 software. In order to eliminate intensity-dependent dye bias from each array, we conducted a LOWESS normalization using GeneSpring. Total gene expression profiles resulting from exposure to 17 chemicals are shown in Figure 2. Hierarchical clustering data demonstrate the similarity between gene expression using Euclidean distance and average linkage algorithm. A total of 1136 gene expression profiles of VOCs samples evidenced very similar patterns. However, the other types differed slightly according to the chemical. The gene expression profile of the vinclozolin sample, which evidenced endocrine disruption and liver-toxicity properties, was shown to cluster in an obscure fashion, neither to the EDCs nor to the LTCs. We noted that the total gene expression profiles could not be successfully classified for the 17 chemicals.

To select genes that could adequately predict 5 chemical-types, we used two known methods. One of these was the statistical method, 1-way ANOVA. We conducted type-parametric Welch's t-test and Benjamini-Hochberg false discovery rate (FDR-adjusted p-value < 0.01) tests. As a result, we identified 286 type-specific differentially expressed genes (Figure 3a). However, the class of vinclozolin remained unclear. As shown in the PCA results, the distance between vinclozolin and other types was similar (Figure 3b). VOCs is shown as a square, PAHs as a diamond, POPs as a star, EDCs as a triangle, and LTCs as a circle. Owing to the weak classification power of the selected genes, we used another method. The other is a SVM (Support Vector Machine) class prediction method using GeneSpring GX 10 software. We then selected 150 genes (prediction strength > 4.148, Fisher's exact test) that could be classify into chemical types. Vinclozolin is closer to LTCs, but PAHs are separate (Data not shown). We hypothesized that the genes common to the two methods can be more definitively classified and class-predicted. We identified those 43 common genes of two methods. As a result, hierarchical clustering was obvious to 5 types of all chemicals. In order to confirm this result, we conducted PCA and demonstrated that 43 common genes of two methods were powerfully class-predictive (Figure 4).

Conclusions

In this study, we analyzed gene expression profiles using the HazChem Human array V3, and then identified powerful predictors for the classification of 17 environmental chemicals into 5 types. Recently, these toxicogenomic approaches have been applied to risk assessments of environmental toxicants in other many studies[13]. If more varied types of environmental chemicals or toxicants are tested, risk assessments for unknown substances may be possible, prior to their ex-

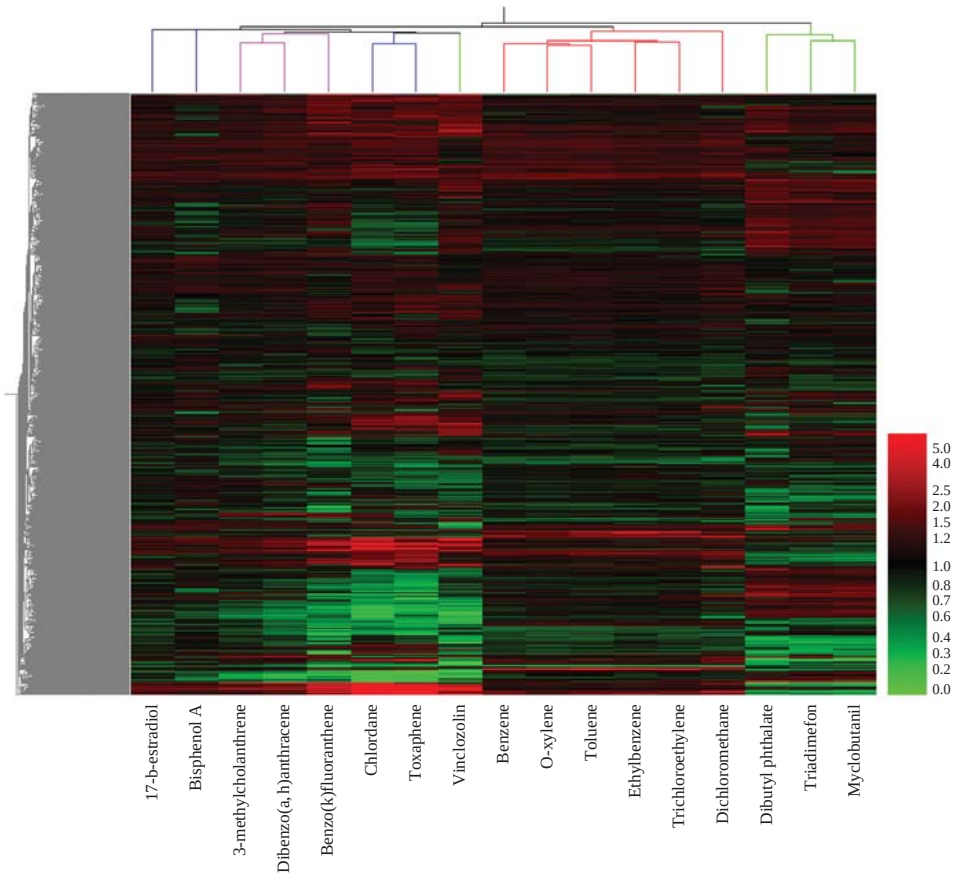


Figure 2. Total gene hierarchical clustering. A total of 1136 genes with similar expression patterns are clustered. Similarity Measure is a Euclidean distance, the clustering algorithm is an average linkage. Red color indicates overexpression, green color indicates downregulation, and a black color is normal.

posure to humans or ecosystems.

Materials and Methods

Chemicals and Reagents

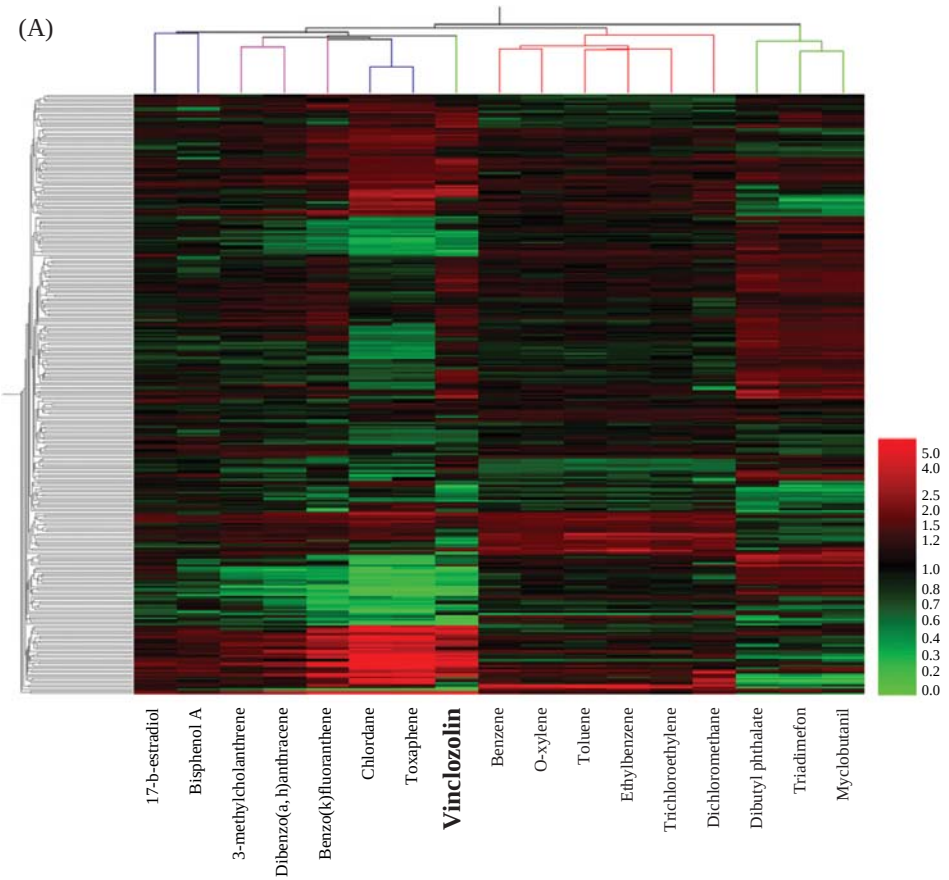
Toxaphene (CAS No. 8001-35-2) and chlordane (CAS No. 57-74-9) were purchased from TCI-EP (Japan), benzene (CAS No. 71-43-2), toluene (CAS No. 108-88-3), o-xylene (CAS No. 95-47-6), ethylbenzene (CAS No. 100-41-4), trichloroethylene (CAS No. 79-01-6), dichloromethane (CAS No. 75-09-2), dibenzo [a,h]antracene (CAS No. 53-70-3), 3-methylcholanthrene (CAS No. 56-49-5), benzo[k]fluoranthene (CAS No. 2007-08-9), Dibutyl phthalate (CAS No, 84-74-2), triadimefon (CAS No. 43121-43-3), vinclozolin (CAS No, 50471-44-8), myclobutanil (CAS No. 88671-89-0), 17β-estradiol (CAS No. 50-28-2), bisphenol A (CAS No. 80-05-7), dimethylsulphoxide (DMSO) and 3-(4,5-dimethylthizol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) were purchased from Sigma-Aldrich Chemical Company (USA). RPMI-1640 Culture Medium, Dulbecco's Phosphate Buffered Saline (PBS) and Fetal Bovine Serum (FBS) were obtained from GIBCO™ (USA). All other chemicals used in this study were of analytical grade or the highest available grade.

Cell Line and Culture

The human hepatocellular carcinoma cell line (HepG2) and the human promyelocytic leukemia HL-60 cell line culture were referred to in a previous paper[14]. Additionally, MTT cell viability assay experiments were conducted as described by Park *et al.*[8]. Respective IC20 values are not shown.

Manufactured of HazChem Array

The HazChem Human array V2 is composed of 600 genes selected from VOCs and PAHs and POPs specifically expressed on each types of chemical, or shown in previous studies to be commonly expressed upon exposure to those three chemical types[9]. In order to assess the toxicogenomic effects of these environmental toxicants, we used a HazChem Human Array V3. In previous *in vivo* studies in the rat, we selected 536 human genes homologous with 1038 liver-toxicity related genes. These previous studies used 12 chemicals that have been shown to be strongly liver-toxic.



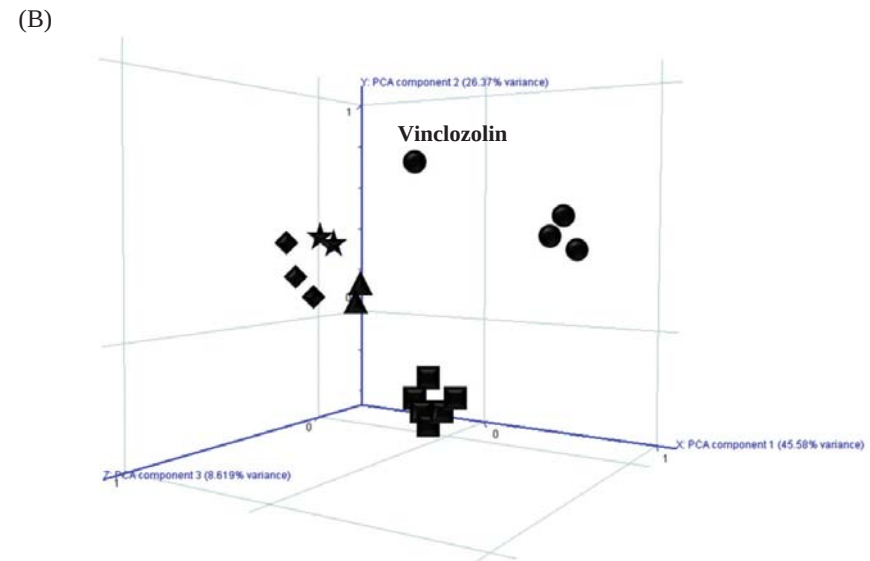


Figure 3. ANOVA selected genes. (A) Hierarchical clustering of ANOVA-selected genes. We selected expressed genes with FDR-adjusted P values of less than 0.01 according to the ANOVA (test type is Welch's t-test). 286 distinct genes with similar expression patterns are clustered. (B) PCA results. VOCs are shown by squares, PAHs are diamonds, POPs are stars, EDCs are triangles, and LTCs are circles.

We manufactured an Agilent custom-made HazChem human array V3, which included a total of 1136 genes.

Microarray Experiment

Total RNA was extracted from the cells using Trizol reagent (Invitrogen Life Technologies) and purified with an RNeasy Mini Kit (Qiagen, USA) in accordance with the manufacturer's instructions. The quantity of each RNA concentration was quantified using Nanodrop, and the RNA quality was assessed with an Agi-

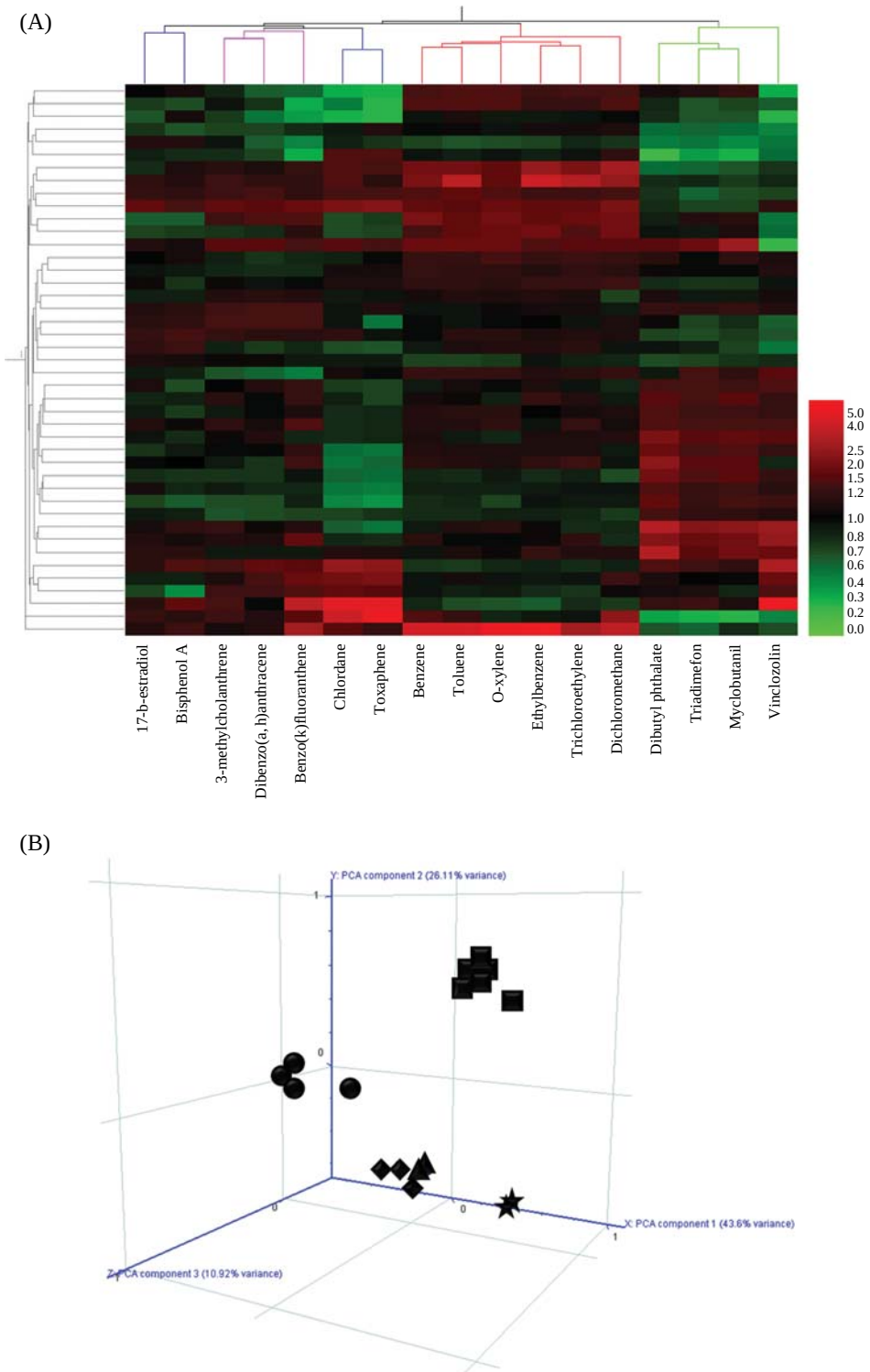


Figure 4. 43 predictors of class by ANOVA and SVM prediction. We identified 43 common genes via two methods as powerful class predictive genes, and confirmed classification using PCA.

lent Bioanalyzer 2100 (Agilent Technologies, CA). Each extracted total RNA sample (30 µg) was labeled with Cyanine (Cy3) or Cyanine (Cy5)-conjugated dCTP (Amersharm, Piscataway, NJ) via a reverse transcription reaction using PrimeScrip Reverse Transcriptase (TaKaRa, Japan). The two labeled cDNAs were then mixed, placed on HazChem Human array V3 (Agilent Technologies, CA) and covered with an Agilent 8-plex

Gasket chamber (Agilent Technologies, CA). The slides were hybridized using an Agilent hybridization system and were washed in 2 X SSC, 0.1% SDS for 2 min, 1 X SSC for 3 min, and then 0.2 X SSC for 2 min at room temperature. The slides were then dried via 20 seconds of centrifugation at 3,000 rpm.

Microarray Data Analysis

The hybridized slides were scanned with an Agilent scanner and the scanned images were analyzed using the Feature Extraction v10.7 software program (Agilent Technologies, CA) and GeneSpring GX 10 (Agilent Technologies, CA). Spots adjudged as substandard via the visual examination of each slide were flagged and excluded from further analysis. Spots that harbored dust artifacts or spatial defects were manually flagged and excluded. In an effort to filter out the unreliable data, spots with signal-to-noise (signal-background-background SD) ratios below 10 were not included in our data. Data were normalized via intensity LOWESS normalization for data reliability. Expression profile data were clustered groups of genes that behaved similarly across the drug treatment experiments using GeneSpring GX 10 (Agilent Technologies, CA). We utilized an algorithm based on Euclidean distance and average linkage to separate genes evidencing similar patterns. Benjamini-Hochberg correction for false discovery rate (FDR) was used for the selection of type-specific genes. We selected expressed genes with FDR-adjusted P values of less than 0.01 according to the ANOVA (test type is Welch's t-test) using GeneSpring GX 10. We conducted a SVM (Support Vector Machine) class prediction method using GeneSpring GX 10 software. Class predictors were selected according to prediction strengths of greater than 4.148 using Fisher's exact test. Additionally we conducted a PCA (principle component of analysis) test for all selected genes[15].

Acknowledgements

This subject was supported by the Korean Ministry of Environment as a component of "The Eco-technopia 21 Project".

References

1. Andreasen, E.A., Mathew, L.K. & Tanguay, R.L. Regenerative growth is impacted by TCDD: gene expression analysis reveals extracellular matrix modulation. *Toxicol. Sci*. **92**, 254-269 (2006).
2. Yoon, C.Y. *et al*. Gene expression profile by 2,3,7,8-tetrachlorodibenzo-p-dioxin in the liver of wild-type (AhR+/+) and aryl hydrocarbon receptor-deficient (AhR−/−) mice. *J. Vet. Med. Sci.* **68**, 663-668 (2006).
3. Koizumi, S. & Yamada, H. DNA microarray analysis of altered gene expression in cadmium-exposed human cells. *J. Occup. Health* **45**, 331-334 (2003).
4. Goetz, A.K. & Dix, D.J. Toxicogenomic effects common to triazole antifungals and conserved between rats and humans. *Toxicol. Appl. Pharmacol.* **238**, 80-89 (2009).
5. U.S. Environmental Protection Agency, Myclobutanil; pesticide tolerances for emergency exemptions. Office of Prevention, Pesticides and Toxic Substances, Washington DC. *Fed. Regist.* **70**, 49499-49507 (2005).
6. Gonzalez, F.J. & Gelboin, H.V. Role of human cytochrome P-450s in risk assessment and susceptibility to environmentally based disease. *J. Toxicol. Environ. Health* **40**, 289-308 (1993).
7. Waters, M.D., Olden, K. & Tennant, R.W. Toxicogenomic approach for assessing toxicant-related disease. *Mutat. Res.* **544**, 415-424 (2003).
8. Park, H.W. *et al*. Gene expression patterns of environmental chemicals in human cell lines using HazChem Human array. *BioChip J.* **3**, 65-70 (2008).
9. An, Y.R. *et al*. Classification of environmental toxicants using HazChem Human array V2. *Mol. Cell. Toxicol.* **5**, 250-256 (2009).
10. Jung, J.W. *et al*. Gene expression analysis of peroxisome proliferators- and phenytoin-induced hepatotoxicity using cDNA microarray. *J. Vet. Med. Sci.* **66**, 1329-1333 (2004).
11. Churchill, G.A. Using ANOVA to analyze microarray data. *Biotechniques* **37**, 173-175, 177 (2004).
12. Pirooznia, M. & Deng, Y. SVM classifier-a comprehensive java interface for support vector machine classification of microarray data. *BMC Bioinformatics* **7(**Suppl 4), S25(2006).
13. Teuschler, L.K., Groten, J.P., Hertzberg, R.C., Mumtaz, M.M. & Rice, G. Environmental chemical mixtures risk assessment: current approaches and emerging issues. *Comments Toxicol.* **7**, 453-493 (2001).
14. Oh, M.J. *et al.* Validation of Human HazChem array using VOC exposure in HL-60 cells. *Mol. Cell. Toxicol.* **4**, 45-51 (2008).
15. Hayden, D. *et al*. Assessing statistical significance in microarray experiments using distance between microarrays. *PLoS One*. **4**, e5838 (2009).