

Selection of Reliable Biomarkers from PCR Array Analyses Using Relative Distance Computational Model: Methodology and Proof-of-Concept Study

Chunsheng Liu^{*†}, Hongyan Xu[†], Siew Hong Lam, Zhiyuan Gong^{*}

Department of Biological Sciences, National University of Singapore, Singapore, Singapore

Abstract

It is increasingly evident about the difficulty to monitor chemical exposure through biomarkers as almost all the biomarkers so far proposed are not specific for any individual chemical. In this proof-of-concept study, adult male zebrafish (*Danio rerio*) were exposed to 5 or 25 µg/L 17β-estradiol (E2), 100 µg/L lindane, 5 nM 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD) or 15 mg/L arsenic for 96 h, and the expression profiles of 59 genes involved in 7 pathways plus 2 well characterized biomarker genes, *vtg1* (*vitellogenin1*) and *cyp1a1* (*cytochrome P450 1A1*), were examined. Relative distance (RD) computational model was developed to screen favorable genes and generate appropriate gene sets for the differentiation of chemicals/concentrations selected. Our results demonstrated that the known biomarker genes were not always good candidates for the differentiation of pair of chemicals/concentrations, and other genes had higher potentials in some cases. Furthermore, the differentiation of 5 chemicals/concentrations examined were attainable using expression data of various gene sets, and the best combination was the set consisting of 50 genes; however, as few as two genes (e.g. *vtg1* and *hspa5* [*heat shock protein 5*]) were sufficient to differentiate the five chemical/concentration groups in the present test. These observations suggest that multi-parameter arrays should be more reliable for biomonitoring of chemical exposure than traditional biomarkers, and the RD computational model provides an effective tool for the selection of parameters and generation of parameter sets.

Citation: Liu C, Xu H, Lam SH, Gong Z (2013) Selection of Reliable Biomarkers from PCR Array Analyses Using Relative Distance Computational Model: Methodology and Proof-of-Concept Study. PLoS ONE 8(12): e83954. doi:10.1371/journal.pone.0083954

Editor: Raya Khanin, Memorial Sloan Kettering Cancer Center, United States of America

Received: September 8, 2013; **Accepted:** November 18, 2013; **Published:** December 12, 2013

Copyright: © 2013 Liu et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by the Singapore National Research Foundation under its Environmental & Water Technologies Strategic Research Programme and administered by the Environment & Water Industry Programme Office (EWI) of the PUB, grant number R-154-000-328-272. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: liuchunshengdid@126.com (CL); dbsgzy@nus.edu.sg (ZG)

† These authors contributed equally to this work.

Introduction

Increasing attention has been drawn to the wide occurrence of natural and man-made chemicals in the aquatic environment. Many chemicals can be bioaccumulated in the aquatic organisms and magnified in the food chains, thus threatening human health. The Minamata disease is a typical case, where methylmercury (MeHg) poisoning occurred in human due to the ingestion of fish and shellfish contaminated by MeHg [1]. Such scenarios have promoted researchers to develop early-warning methods for monitoring contaminants in the aquatic system through both chemical monitoring and biomonitoring.

As new pollutants in the environment are emerging rapidly, it becomes increasingly unfeasible to monitor all contaminants in the environment. Since the presence of a foreign chemical in a segment of the environment does not always indicate adverse biological effects [2], it is important to combine chemical monitoring with the biomonitoring for a reliable environmental risk assessment. An ideal approach is to examine biological responses that can reflect the contaminants in the exposed organisms [2]. Under this concept, various biomarkers from fish have been proposed and used for biomonitoring aquatic contaminants. However, most of biomarkers proposed were not specific

for individual chemicals. For example, biomarker for estrogen, *vtg1* mRNA could be induced not only by the native female hormone, 17β-estradiol (E2), but also by many other compounds that can interact with estrogen receptors, including many xenobiotics, such as lindane [3]. The expression of *cyp1a1* was up-regulated by 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD) as well as by other chemicals such as arsenic in mice [4].

It has been demonstrated that exposure to single chemicals generated unique gene expression signature in experimental animals [5–9]. Therefore, a multi-parameter quantitative real-time PCR (qRT-PCR) array could be developed as a useful tool to differentiate a complicated set of chemical groups. However, in previous studies, the parameters (genes) were selected only based on responsive difference of gene expression among chemicals after exposure [10–11] and did not represent the best parameter (gene) set for the discrimination of chemicals. Therefore, a proof-of-concept study was designed and conducted in the present study, with the objective of finding the best parameter (gene) set for the discrimination of chemicals tested. Especially, a relative distance (RD) computational model was developed to select gene sets from 61 gene examined for chemical discrimination. Therefore, it is feasible to integrate qRT-PCR arrays and RD computational

model to develop a reliable biomonitoring tool for chemical exposure.

Materials and Methods

Chemicals and reagents

E2, lindane, TCDD and arsenic ($\text{Na}_2\text{HAsO}_4 \cdot 7\text{H}_2\text{O}$) were purchased from Sigma (St. Louis, MO, USA). Arsenic was dissolved in deionized water directly and the other three chemicals were dissolved in dimethyl sulfoxide (DMSO) as stock solutions. The TRizol reagent and LightCycle FastStart DNA Master SYBR Green I were obtained from Invitrogen (New Jersey, NJ, USA) and Roche Applied Science (Mannheim, Germany), respectively.

Fish and chemical exposure

In this study, experimental procedures were carried out following the approved protocol by Institutional Animal Care and Use Committee of National University of Singapore (Protocol 079/07). Adult male zebrafish (*Danio rerio*, 5-month old) were purchased from a local aquarium farm (Mainland Tropical Fish Farm, Singapore), and acclimated for at least two weeks in our aquarium before chemical treatment. After acclimation, fish were exposed to 5 nM TCDD, 5 $\mu\text{g}/\text{L}$ E2, 50 $\mu\text{g}/\text{L}$ E2, 100 $\mu\text{g}/\text{L}$ lindane or 15 mg/L arsenic for 96 h in a static condition. Each tank (5 L size) included 3 L exposure solution and 3 fish, and each concentration included 3 replicated tanks. During the exposure period, fish were fed once a day with commercial frozen bloodworms (Hikari) as described before [12]. The concentrations of these chemicals were chosen based on previous studies of ours and others [12–16], where biological effects of these concentrations have been confirmed by significant changes of some mRNAs examined. For E2, two concentrations were used to test the feasibility to develop a gene expression based model to differentiate exposure concentrations besides different chemicals. Fresh chemical solutions were daily replaced during the exposure experiment. For E2, lindane and TCDD exposure experiments, treatment and control groups received 0.01% DMSO, and for arsenic exposure experiments, treatment and control groups received 0.01% deionized water in this study. After 96-h exposure, the fish were anesthetized with MS-222 (1 mM) and livers were collected and preserved in TRizol reagent at -80°C until RNA isolation.

Selection of target genes for PCR array

A PCR array of sixty-one zebrafish genes was designed as follows. First, seven well characterized pathways commonly affected by chemicals were selected: oxidative and metabolic stress [17–18], apoptosis signaling [19–20], proliferation and carcinogenesis [21–22], DNA damage and repair [23–24], growth arrest and senescence [25–26], heat shock [27–28], and inflammation pathways [29–30]. Representative genes from these pathways were selected by referring Molecular Toxicology PathwayFinder PCR array from SABioscience Gene Network Central (http://www.sabiosciences.com/rt_pcr_product/HTML/PAHS-3401Z.html). Second, annotated zebrafish orthologues of human genes were searched from Ensemble website and confirmed using online synteny tool [31]; unannotated zebrafish orthologues were manually determined first by amino acid sequence comparison with human candidate sequences through UCSC website (<http://genome.ucsc.edu/>) and then confirmed by comparison of genomic organization, chromosomal locations and chromosomal synteny analysis as conducted in a previously study [32]. Finally the zebrafish orthologues of 59 human genes were obtained for designing of PCR primers. In addition, two well-

established biomarker genes, *vtg1* and *cyp1a1*, were also included in order to compare the potentials of biomonitoring between traditional biomarkers and genes/gene sets developed in this study, as inducers of *vtg1* and *cyp1a1* such as E2 and TCDD were also used in the present exposure experiments. The complete list of genes in PCR array and their PCR primer sequences are presented in Table S1. The number of genes in each pathway was 14, 10, 10, 6, 4, 13 and 2 for oxidative and metabolic stress, apoptosis signaling, DNA damage and repair, proliferation and carcinogenesis, growth arrest and senescence, heat shock and inflammation pathways, respectively.

Quantitative real-time PCR (qRT-PCR)

Total RNA was isolated from zebrafish livers with TRizol reagent and used for cDNA synthesis. Real time qPCR was performed using the LightCycler system (Roche Applied Science, Mannheim, Germany) with LightCycler FastStart DNA Master SYBR Green I following manufacturer's instruction. The primer sequences were designed using Primer 3 software (<http://frodo.wi.mit.edu/as>). The amplicon efficiencies of primers were $>90\%$. Three housekeeping genes, β -actin (*beta-actin*), β -2m (*beta-2-microglobulin*) and *rpl13a* (*ribosomal protein L13a*), were used as internal control and the geometric means the expression of the three housekeeping genes were used as the normalized factor by $2^{-\Delta\Delta\text{Ct}}$ method. Each group included three biological replicates and each replicate included a pool of three fish.

Statistical analysis

Gene expression values were logarithmically transformed (\log_2) before statistical analysis. The homogeneity and normality of data were examined using the Kolmogorov-Smirnov and Levene's test, respectively. Statically significant differences between treatment and corresponding control groups were evaluated by ANOVA based on a p -value <0.05 . Average linkage ($p < 0.05$) was used to examine the cluster relationships of different treatment groups based on mRNA expression profiles. The statistical analyses were performed using Kyplot Demo 3.0 software (Tokyo, Japan).

Relative distance (RD) computational model

The differentiation of two chemical/concentration groups not only depends on Euclidean distance between the two groups but also depends on the distance among individual replicates within each group. In this study, the RD computational model was developed to quantitatively describe the potential that three biological replicates from group A can be differentiated from the three replicates in group B based on mRNA expression profiles (fold change), and RD between one replicate from group A treatment and three replicates from group B (rd_{a1b})

$$rd_{a1b} = md_{a1b} - md_{aa} - 1/2 \times SD_{a1b} - 1/2 \times SD_{aa} \quad (1)$$

$$md_{a1b} = (d_{a1b1} + d_{a1b2} + d_{a1b3})/3 \quad (2)$$

$$md_{aa} = (d_{a1a1} + d_{a1a3})/2 \quad (3)$$

$$d_{a_1b_1} = \sqrt{\sum_{j=1}^j (a_{1j} - b_{1j})^2}$$

$$d_{a_1b_2} = \sqrt{\sum_{j=1}^j (a_{1j} - b_{2j})^2} \quad (4)$$

$$d_{a_1b_3} = \sqrt{\sum_{j=1}^j (a_{1j} - b_{3j})^2}$$

$$d_{a_1a_3} = \sqrt{\sum_{j=1}^j (a_{1j} - a_{3j})^2} \quad (5)$$

$$SD_{a_1b} = \sqrt{((d_{a_1b_1} - md_{a_1b})^2 + (d_{a_1b_2} - md_{a_1b})^2 + (d_{a_1b_3} - md_{a_1b})^2) / (3 - 1)} \quad (6)$$

$$SD_{aa} = \sqrt{((d_{a_1a_2} - md_{aa})^2 + (d_{a_1a_3} - md_{aa})^2) / (2 - 1)} \quad (7)$$

where j is the total number of genes examined; a and b are gene expression values in the treatment groups A and B, respectively; md_{a_1b} is the mean Euclidean distance between one biological replicate from treatment group A (a_1) and three replicates from treatment group B (b_1, b_2, b_3); md_{aa} is the mean Euclidean distance between one biological replicate from treatment group A (a_1) and other two biological replicates from the same group (a_2, a_3); SD_{a_1b} is the standard deviation of Euclidean distance between one biological replicate from treatment group A (a_1) and three replicates from treatment group B (b_1, b_2, b_3); SD_{aa} is the standard deviation of Euclidean distance between one biological replicate from treatment group A treatment and other two biological replicates from the same group; $d_{a_1b_1}$, $d_{a_1b_2}$ and $d_{a_1b_3}$ are the Euclidean distance between one biological replicate from treatment group A (a_1) and three replicates from treatment group B (b_1, b_2, b_3); $d_{a_1a_2}$ and $d_{a_1a_3}$ are the Euclidean distance of biological responses between one biological replicate from treatment group A (a_1) and other two biological replicates from the same group (a_2, a_3).

In this study, first, we calculated all the RD values between two chemical treatment groups using expression data of individual genes. When all six RD values were >0 for each pair of chemicals, it was considered that the gene could be used to differentiate the two chemicals/concentrations. The cluster analyses (average linkage) were performed using commercial software (Kyplot Demo 3.0, Tokyo, Japan) (p -value <0.05) to confirm the feasibility of RD model in screening genes for the differentiation of chemical/concentration treatments. Second, the mean RD values were calculated to quantitatively compare the potentials of individual genes in differentiating two chemicals/concentrations. Finally, a C-language computational program (see Program S1) was edited for selecting genes and generating gene sets that could be used to differentiate all of five chemical/concentration treatments simultaneously using the RD model developed in this study, and maximum mean RD of each gene sets with the same amount of genes and the corresponding components of genes were outputted.

Results

Broad changes of gene expression patterns in the seven selected pathways in response to chemical insults

Adult male zebrafish were treated with 5 nM TCDD, 5 μ g/L E2, 50 μ g/L E2, 100 μ g/L lindane or 15 mg/L arsenic for 96 hours and no mortalities were observed throughout the exposure experiment. As shown in Figure 1 and Table S2, exposure to different chemicals led to different gene expression profiles. TCDD exposure significantly down-regulated the expression of most selected genes involved in the oxidative and metabolic stress, apoptosis signaling, DNA damage and repair, proliferation and carcinogenesis, growth arrest and senescence, heat shock and inflammation pathways, while the expression of *cyp1a1*, *hspa5* and *hsp70* (heat shock protein 70-kDa) was among the highest up-regulated. Treatment with arsenic significantly altered the expression of most selected genes in the seven pathways, such as up-regulation of expression of *ptgs1* (prostaglandin-endoperoxide synthase 1), *cyp1a1* and *hsp90aa1* (heat shock protein 90, alpha, class A member 1, tandem duplicate 1), and down-regulation of *b1p1* (Bcl-XL-like protein 1), *tnfr* (tumor necrosis factor receptor) and *vtg1*. A significant up-regulation in the expression of *vtg1* was observed upon exposure to 5 or 50 μ g/L E2, clearly showing estrogenic activity. Similar to TCDD, exposure to E2 (5 or 50 μ g/L) significantly down-regulated the expression of most selected genes included in the seven pathways investigated. In contrast, exposure to lindane up-regulated the expression of most selected genes in the seven pathways; with exception of only few down-regulated genes, notably *cdkn1a* (cyclin-dependent kinase inhibitor 1A, transcript variant 1) in the growth arrest and senescence pathway and *fmo5* (flavin containing monooxygenase 5) in the oxidative and metabolic stress pathway.

Correlation of RD and potential differentiation of chemical treatment pairs

Using an RD computational model, we calculated all of RD values between two chemical/concentration treatment groups based on expression fold change of individual genes and the results are presented in Figure 2 (see details in Table S3) for all of the 10 possible chemical/concentration pairs. The ability of each of the 61 genes to discriminate the chemical/concentration pairs was tested by the software Kyplot Demo 3.0 program and the findings are presented in Figure 2. There was a good correlation of the RD and the ability to discriminate pair of chemicals/concentrations. All the genes with top and high RD values were found to be able to discriminate pair of chemicals/concentrations. For example, the two best known biomarker genes, *vtg1* and *cyp1a1*, were able to discriminate eight of the ten pairs: TCDD/arsenic, TCDD/E2_high, E2_high/lindane, E2_high/arsenic, TCDD/E2_low, TCDD/lindane, lindane/E2_low, and arsenic/E2_low. However, for the lindane/arsenic pair, *cyp1a1* could not be used to discriminate them, while for the E2_low/E2_high concentration pair, both *vtg1* and *cyp1a1* failed to discriminate them. Interestingly, *vtg1* and *cyp1a1* were not always among the top of the list based on the calculated RD. There were also many other genes (even with better RD) that could be also used to differentiate the corresponding pair of chemicals.

Selection of discriminating gene sets based on RD computational model

While it is relatively easy to discriminate a pair of chemical treatment groups based on expression data from one or few genes, it is more challenging to discriminate multiple treatment groups.

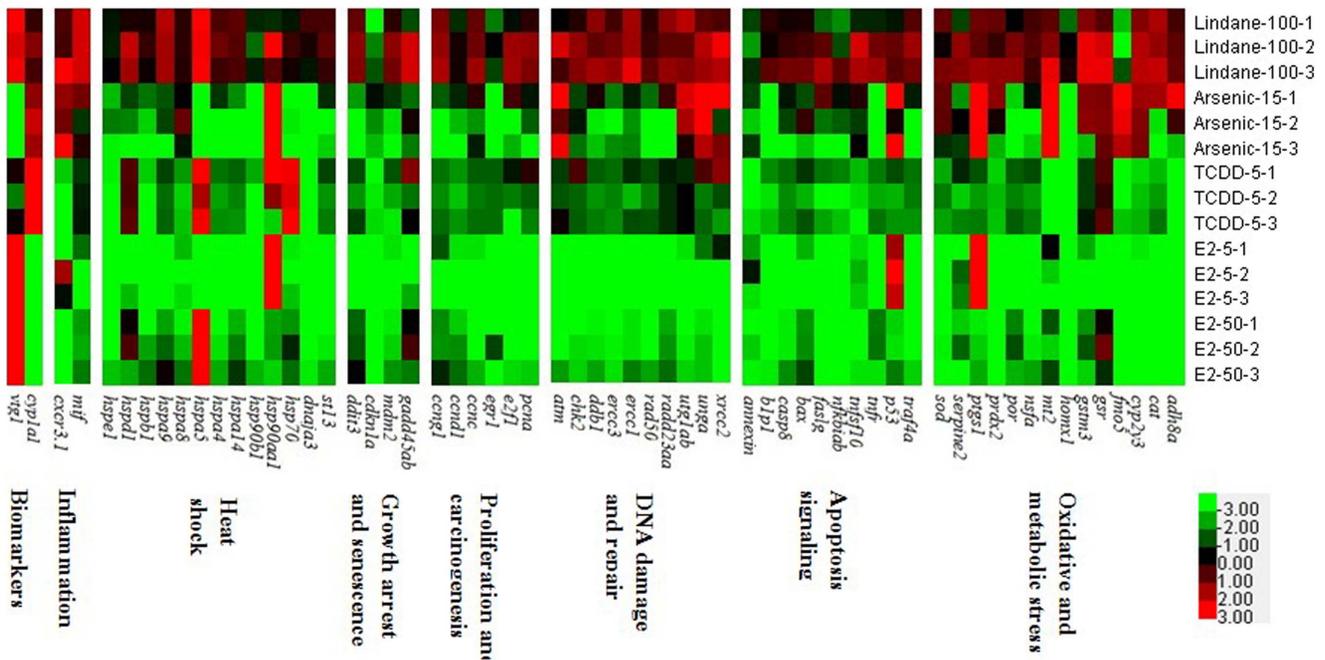


Figure 1. Gene expression profiles included in seven selected pathways in male zebrafish livers after exposure to 100 µg/L lindane, 5 nM 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD), 5 µg/L 17β-estradiol (E2), 25 µg/L E2, or 15 mg/L arsenic for 96 h. There were 3 biological replicates, and each replicate were pooled from 3 fish. Gene expressions were expressed as fold change relative to the corresponding control. The full names of genes can be found in Tables S1 or S2. doi:10.1371/journal.pone.0083954.g001

In the current dataset, no single gene can be used to discriminate all of the five chemical/concentration groups. Thus, it was necessary to select a gene set for discriminating the chemical/concentration groups. Here, we further explored the RD model to select best gene sets for differentiating all of the five chemical/concentration groups. RDs were computed for all possible gene

combinations from one to 61 genes and the highest mean distances for gene sets from 1 to 65 genes are presented in Figure 3. For example, the 2-gene set of the highest mean RD was *vtg1* and *hspa5* with a value of 10.57 (Figure 3 and Table S4) and the two genes can be used to discriminate the five chemical/concentration groups perfectly (Figure 4A). In comparison, the gene pair of best

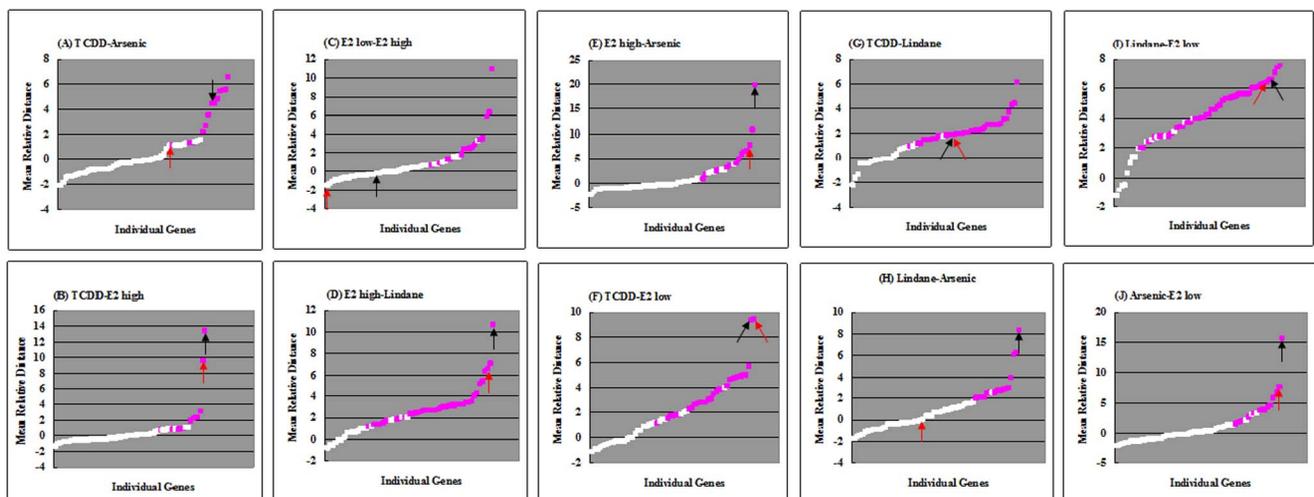


Figure 2. Mean Relative Distances (RDs) between two chemicals/concentration groups. (A) TCDD vs. Arsenic; (B) TCDD vs. E2_high; (C) E2_low vs. E2_high; (D) E2_high vs. Lindane; (E) E2_high vs. Arsenic; (F) TCDD vs. E2_low; (G) TCDD vs. Lindane; (H) Lindane vs. Arsenic; (I) Lindane vs. E2_low; (J) Arsenic vs. E2_low. Black arrows indicate the positions of *vtg1*, and red arrows indicate the positions of *cyp1a1*; White boxes indicate the positions of genes that did not pass the model test and could not be used to discriminate the corresponding two chemicals/concentrations; Pink boxes indicate the positions of genes that passed the model test and could be used to discriminate the corresponding two chemicals/concentrations. TCDD: 5 nM 2,3,7,8-tetrachlorodibenzo-p-dioxin; lindane: 100 µg/L lindane; arsenic: 15 mg/L arsenic; E2_low: 5 µg/L 17β estradiol; E2_high: 50 µg/L 17β-estradiol. The information of RDs and the corresponding genes can be found in Table S3. doi:10.1371/journal.pone.0083954.g002

known biomarkers, *vtg1* and *cyp1a1*, has a value of 10.33 and they could not correctly discriminate all of the five groups, particularly the two concentration groups of E2 treatment (Figure 4B). All other gene sets (3 or more genes) of the highest mean RD were also capable of differentiating all the five chemical/concentration groups correctly (Figure 3). In general, there was an increase of mean RD with the number of genes in gene sets and the maximal mean RD (19.153) was observed in the set with 50 genes, where chemicals were also completely differentiated, including different concentrations (Figure 4C).

Discussion

The environment is continuously loaded with natural and man-made chemicals, and the effects of contaminant exposure to human health have been extensively documented [33–37]. In general, adverse effects of contaminants at population levels in wildlife and human tend to be delayed; when the effects finally become clear, the destructive processes may have been beyond the point where it can be reversed by available remedial actions [2]. Therefore, various biomonitoring methods have been developed in the past few decades for the purpose of early warning. However, most of these methods focused on one or several biological parameters (e.g., biomarkers vitellogenins and cytochrome P450 enzymes 1A1) [38–43]. To search for more biomarker genes to predict chemical contamination, it is common to use high throughput and large scale analyses such as DNA microarray and more recently RNA-seq platform [8,12,44]. However, the methodology for selecting biomarkers from thousands of genes could be a great challenge. Here we performed a proof-of-concept study by selecting a handful of biomarker genes to develop a practical assay with the aid of RD computational model.

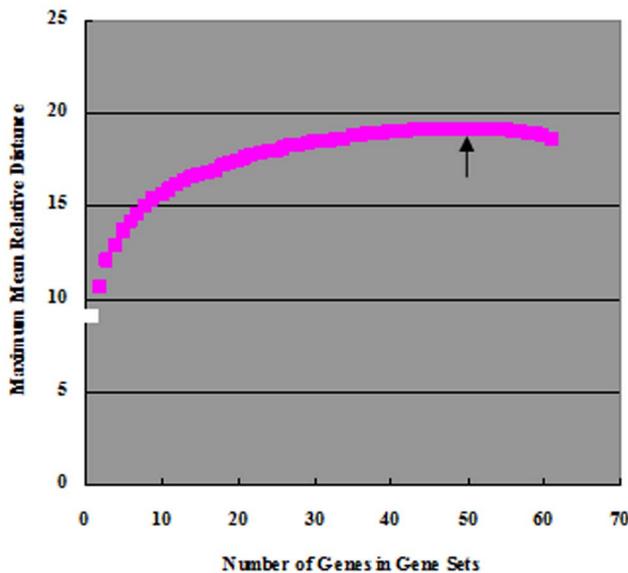


Figure 3. Maximum mean RD of gene sets with different numbers of genes among 5 chemicals/concentrations. Black arrow indicates the position of gene set (50 genes), where maximum RD was achieved. White box indicates the position of gene set (1 gene) that did not pass the model test and could not be used to differentiate the corresponding five chemicals/concentrations; Pink boxes indicate the positions of gene sets that passed the model test and could be used to differentiate the corresponding five chemicals/concentrations. The information about maximum mean RDs and the corresponding components of genes can be found in Table S4. doi:10.1371/journal.pone.0083954.g003

Here four chemicals including E2, lindane, TCDD and arsenic were tested. Both E2 and lindane exposures caused up-regulation of hepatic *vtg1* expression; similarly, treatment with TCDD or arsenic showed up-regulation of *cyp1a1* expression. These observations are consistent with previous studies [3–4,45], suggesting the effectiveness of these chemical exposure experiments. In general, exposure to different chemicals resulted in different gene expression profiles in the seven biological pathways examined. For example, both of E2 and lindane induced *vtg1* expression, but E2 down-regulated the expression of essentially all of the selected genes in the seven pathways while lindane up-regulated the expression of most of these genes. Similarly, TCDD down-regulated the expression of most of genes and arsenic up-regulated many of the genes, especially in two pathways, oxidative_and_metabolic_stress and DNA_damage_and_repair, suggesting a molecular basis for their discrimination.

In the current data set, we found that none of the 61 genes could be used to correctly discriminate all of the five chemical/concentration groups; thus, it has to rely on multiple gene sets for successful discrimination, which should be the direction for future development of multiple gene signatures for discrimination of a multiple chemical groups, as previously proposed [8,12]. To systematically select the best discriminator genes, here we developed a computational model using RD to determine the prediction power of each gene or in combination with others. First, we demonstrated that there was a positive correlation between the RD values and the discrimination of different treatments groups (Fig. 2). In our data set, a minimum of two genes (e.g. *vtg1* and *hspa5*) could be used to successfully discriminate all of the five chemical/concentration groups. There is a general increase of mean RD values with the number of genes added to the gene set, which indicate the power of using more genes for discriminating more complicated data set. In our dataset, we also found that the 50-gene set had the highest mean RD values, indicating that there is an optimal gene number used for the discrimination. From a practical viewpoint, the used of minimal number of genes will minimize workload and ease downstream data analysis. However, using more genes, especially those representing different molecular pathways, provides additional important biological information in molecular-marker based biomonitoring.

In summary, the data of this study demonstrated chemicals that induced similar responses in biomarker (e.g., TCDD and arsenic, E2 and lindane) could cause different biological responses depending on the parameters examined, and the use of parameter sets consisting of different biological responses for biomonitoring should be more appropriate. Furthermore, the computational model based on RD may be useful to select appropriate gene sets to develop efficient biomarker-based biomonitoring. Considering the rapid, sensitive, convenient and high-throughput properties of PCR, a PCR array including multiple gene parameters should be a feasible tool to develop for biomonitoring of chemical exposure.

Supporting Information

Table S1 Sequences of primers for selected genes.
(DOC)

Table S2 mRNA expression profiles in the livers of zebrafish after chemical exposure.
(DOC)

Table S3 Mean relative distances (MRDs) of individual genes between chemicals.
(DOC)

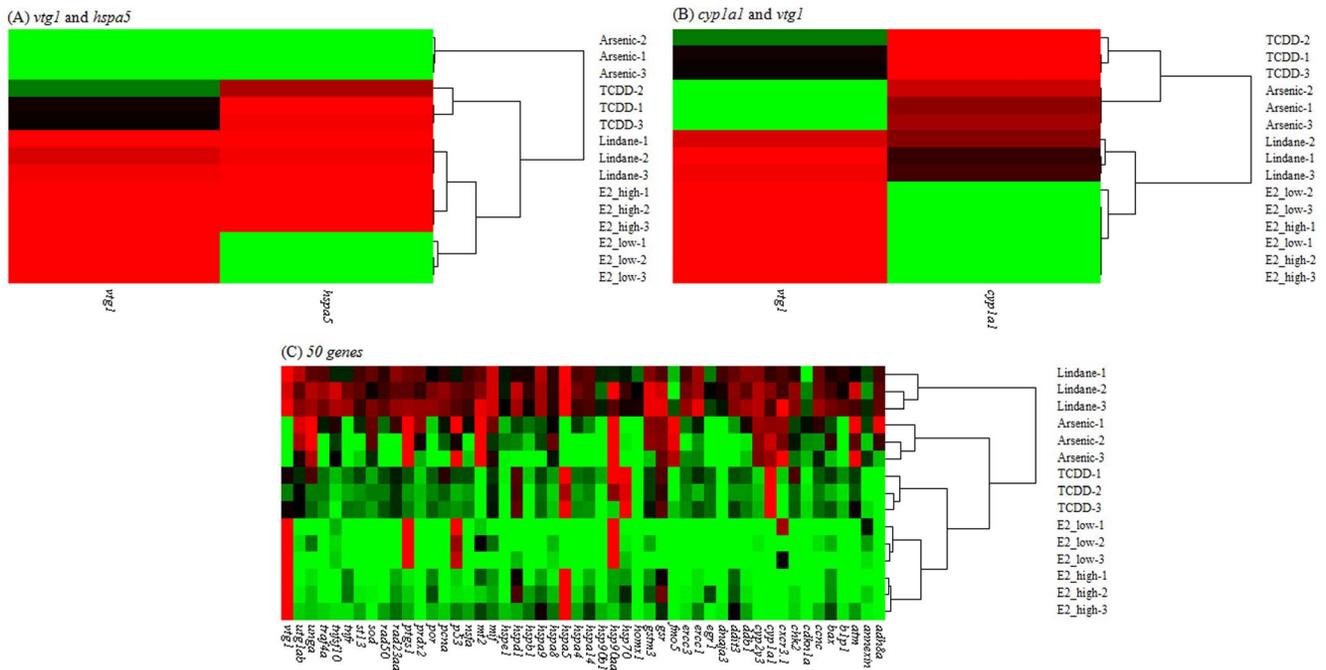


Figure 4. Clustering relationships among chemicals/concentrations using mRNA expression data of (A) *cyp1a1* and *vtg1*, (B) *vtg1* and *hspa5* and (C) 50 genes with the maximum RD. TCDD: 5 nM 2,3,7,8-tetrachlorodibenzo-*p*-dioxin; lindane: 100 µg/L lindane; arsenic: 15 mg/L arsenic; E2_low: 5 µg/L 17β-estradiol; E2_high: 50 µg/L 17β-estradiol. The full names of genes can be found in Table S1 or S2. doi:10.1371/journal.pone.0083954.g004

Table S4 Maximum mean relative distances (MMRDs) of gene sets with different amounts of genes among 5 chemicals/concentrations and the corresponding components of genes. (DOC)

Program S1
(ZIP)

References

1. Harada M (1995) Minamata disease: Methylmercury poisoning in Japan caused by environmental pollution. *Crit Rev Toxicol* 25: 1–24.
2. van der Oost R, Beyer J, Vermeulen NPE (2003) Fish bioaccumulation and biomarkers in environmental risk assessment: a review. *Environ Toxicol Phar* 13: 57–149.
3. Flouriot G, Pakdel F, Ducouret B, Valotaire Y (1995) Influence of xenobiotics on rainbow trout liver estrogen receptor and vitellogenin gene expression. *J Mol Endocrinol* 15: 143–151.
4. Wu JP, Chang LW, Yao HT, Chang H, Tsai HT, et al. (2009) Involvement of oxidative stress and activation of aryl hydrocarbon receptor in elevation of CYP1A1 expression and activity in lung cells and tissues by arsenic: an *in vitro* and *in vivo* study. *Toxicol Sci* 107: 385–393.
5. Amin RP, Hamadeh HK, Bushel PR, Bennett L, Afshari CA, et al. (2002) Genomic interrogation of mechanism(s) underlying cellular responses to toxicants. *Toxicology* 181–182: 555–563.
6. Bartosiewicz M, Penn S, Buckpitt A (2001) Applications of gene arrays in environmental toxicology: fingerprints of gene regulation associated with cadmium chloride, benzo(a)pyrene, and trichloroethylene. *Environ Health Perspect* 109: 71–74.
7. Hamadeh HK, Bushel PR, Jayadev S, Martin K, DiSorbo O, et al. (2002) Gene expression analysis reveals chemical-specific profiles. *Toxicol Sci* 67: 219–231.
8. Hook SE, Skillman AD, Small JA, Schultz IR (2006) Gene expression patterns in rainbow trout, *Oncorhynchus mykiss*, exposed to a suite of model toxicants. *Aquat Toxicol* 77: 372–385.
9. Lam SH, Mathavan S, Tong Y, Li H, Karuturi RKM, et al. (2008) Zebrafish whole-adult-organism chemogenomics for large-scale predictive and discovery chemical biology. *PLoS Genetics* 4: e1000121.
10. Garcia-Reyerer N, Poynton HC, Kennedy AJ, Guan X, Escalon BL, et al. (2009) Biomarker discovery and transcriptomic responses in *Daphnia magna* exposed to munitions constituents. *Environ Sci Technol* 43: 4188–4193.

Author Contributions

Conceived and designed the experiments: CL HX SHL ZG. Performed the experiments: HX. Analyzed the data: CL HX SHL ZG. Contributed reagents/materials/analysis tools: CL HX SHL ZG. Wrote the paper: CL HX SHL ZG.

11. Osborn HL, Hook SE (2013) Using transcriptomic profiles in the diatom *Phaeodactylum tricornutum* to identify and prioritize stressors. *Aquat Toxicol* 138–139: 12–25.
12. Lam SH, Winata CL, Tong Y, Korzh S, Lim WS, et al. (2006) Transcriptome kinetics of arsenic-induced adaptive response in zebrafish liver. *Physiol Genomics* 27: 351–361.
13. Cuesta A, meseguer J, Esteban MÁ (2008) Effects of the organochlorines *p,p'*-DDE and lindane on gilthead seabream leucocyte immune parameters and gene expression. *Fish Shellfish Immun.* 25: 682–688.
14. Mattingly C, Toscano WA (2011) Posttranscriptional silencing of cytochrome P4501A1 (CYP1A1) during zebrafish (*Danio rerio*) development. *Dev Dynam* 222: 645–654.
15. Yamaguchi A, Ishibashi H, Kohra S, Arizono K, Tominaga N (2005) Short-term effects of endocrine-disrupting chemicals on the expression of estrogen-responsive genes in male medaka (*Oryzias latipes*). *Aquat Toxicol* 72: 239–249.
16. Xu H, Lam S.H, Shen Y, Gong Z (2013) Genome-wide identification of molecular pathways and biomarkers in response to arsenic exposure in zebrafish liver. *PLoS ONE* 8: e68737.
17. Di Giulio R.T, Washburn P.C, Wenning R.J, Winston G.W, Jewell C.S (1989) Biochemical responses in aquatic animals: A review of determinants of oxidative stress. *Environ Toxicol Chem* 8: 1103–1123.
18. Lackner R (1998) “Oxidative stress” in fish by environmental pollutants. *Fish Ecotoxicol* 86: 203–224.
19. Franco R, Sánchez-Olea R, Reyes-Reyes E.M, Panayiotidis M.I (2009) Environmental toxicity, oxidative stress and apoptosis: Ménage à trois. *Mutat Res-Rev Mutat* 674: 3–22.
20. Roberts R.A, Nebert D.W, Hickman J.A, Richburg J.H, Goldsworthy T.L (1997) Perturbation of the mitosis/apoptosis balance: A fundamental mechanism in toxicology. *Toxicol Sci* 38: 107–115.
21. Murata M, Midorkawa K, Koh M, Umezawa K, Kawanishi S (2004) Genistein and daidzein induce cell proliferation and their metabolites cause oxidative

- DNA damage in relation to isoflavone-induced cancer of estrogen-sensitive organs. *Biochemistry* 43: 2569–2577.
22. Soto AM, Sonnenschein C (2010) Environmental causes of cancer: endocrine disruptors as carcinogens. *Nat Rev Endocrinol* 6: 363–370.
 23. Simic M.G (1991) DNA damage, environmental toxicants, and rate of aging. *J Environ Sci Heal C* 9: 113–153.
 24. Hartwig A, Schwerdtle T (2002) Interactions by carcinogenic metal compounds with DNA repair processes: toxicological implications. *Toxicol Lett* 127: 47–54.
 25. Caino M.C, Oliva J.L, Jiang H, Penning T.M, Kazanietz M.G (2007) Benzo[*a*]pyrene-7,8-dihydrodiol promotes checkpoint activation and G₂/M arrest in human bronchoalveolar carcinoma H358 cells. *Mol Pharmacol* 71: 744–750.
 26. Pomati F, Castiglioni S, Zuccato E, Fanelli R, Vigetti D, et al. (2006) Effects of a complex mixture of therapeutic drugs at environmental levels on human embryonic cells. *Environ Sci Technol* 40: 2442–2447.
 27. Gupta S.C, Sharma A, Mishra M, Mishra R.K, Chowdhuri D.K (2010) Heat shock proteins in toxicology: How close and how far? *Life Sci* 86: 377–384.
 28. Lee S, Lee S, Park C, Choi J (2006) Expression of heat shock protein and hemoglobin genes in *Chironomus tentans* (Diptera, chironomidae) larvae exposed to various environmental pollutants: a potential biomarker of freshwater monitoring. *Chemosphere* 65: 1074–1081.
 29. Handzel Z.T (2000) Effects of environmental pollutants on airways, allergic inflammation, and the immune response. *Rev Environ Health* 15: 325–336.
 30. Khalaf H, Salste L, Karlsson P, Ivarsson P, Jass J, et al. (2009) In vitro analysis of inflammatory responses following environmental exposure to pharmaceuticals and inland waters. *Sci Total Environ* 407: 1452–1460.
 31. Catchen J.M, Conery J.S, Postlethwait J.H (2009) Automated identification of conserved synteny after whole-genome duplication. *Genome Res* 19: 1497–1505.
 32. Xu H, Li Z, Li M, Wang L, Hong Y (2009) Boule is present in fish and bisexually expressed in adult and embryonic germ cells of medaka. *PLoS ONE* 4: e6097.
 33. Harley K.G, Marks AR, Chevrier J, Bradman A, Sjödin A, et al. (2010) PBDE concentrations in women's serum and fecundability. *Environ Health Perspect* 118: 699–704.
 34. Meeker JD, Stapleton HM (2013) House dust concentrations of organophosphate flame retardants in relation to hormone levels and semen quality parameters. *Environ Health Perspect* 118: 318–323.
 35. Nelson JW, Hatch EE, Webster TF (2009) Exposure to polyfluoroalkyl chemicals and cholesterol, body weight, and insulin resistance in the general U.S. population. *Environ Health Perspect* 118: 197–202.
 36. Stapleton HM, Eagle S, Anthopolos R, Wolkin A, Miranda ML (2011) Associations between polybrominated diphenyl ether (PBDE) flame retardants, phenolic metabolites, and thyroid hormones during pregnancy. *Environ Health Perspect* 119: 1454–1459.
 37. Vested A, Ramlar-Hansen CH, Olsen SF, Bonde JP, Kristensen SL, et al. (2013) Associations of in utero exposure to perfluorinated alkyl acids with human semen quality and reproductive hormones in adult men. *Environ Health Perspect* 121: 453–458.
 38. Ariese F, Kok SJ, Verkaik M, Gooijer C, Velthorst NH, et al. (1993) Synchronous fluorescence spectrometry of fish bile: A rapid screening method for the biomonitoring of PAH exposure. *Aquat Toxicol* 26: 273–286.
 39. Lucarelli F, Authier L, Bagni G, Marrazza G, Baussant T, et al. (2003) DNA biosensor investigations in fish bile for use as a biomonitoring tool. *Anal Lett* 36: 1887–1901.
 40. Schnurstein A, Braunbeck T (2001) Tail moment *versus* tail length—Application of an *in vitro* version of the comet assay in biomonitoring for genotoxicity in native surface waters using primary hepatocytes and gill cells from zebrafish (*Denio rerio*). *Ecotox Environ Sate* 49: 187–196.
 41. Thomas M, Florion A, Chrétien D, Terver D (1996) Real-time biomonitoring of water contamination by cyanide based on analysis of the continuous electric signal emitted by a tropical fish: *Apteronotus albifrons*. *Water Res* 30: 3083–3091.
 42. Vindimian E, Namour P, Migeon B, Garric J (1991) In situ pollution induced cytochrome P450 activity of freshwater fish: barbell (*Barbus barbus*), chub (*Leuciscus cephalus*) and nase (*Chondrostoma nasus*). *Aquat Toxicol* 21: 255–266.
 43. Zeng Z, Shan T, Tong Y, Lam SH, Gong Z (2006) Development of estrogen-responsive transgenic medaka for environmental monitoring of endocrine disrupters. *Environ Sci Technol* 39: 9001–9008.
 44. Zheng W, Xu H, Lam SH, Luo H, Karuturi RK, et al. (2013) Transcriptomic analyses of sexual dimorphism of the zebrafish liver and the effect of sex hormones. *PLoS One* 8: e53562.
 45. Ankley GT, Miller DH, Jensen KM, Villeneuve DL, Martinović D (2008) Relationship of plasma sex steroid concentrations in female fathead minnows to reproductive success and population status. *Aquat Toxicol* 88: 67–74.