Science of the Total Environment 408 (2010) 5824–5827

Contents lists available at ScienceDirect

Science of the Total Environment

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

A microarray data analysis method to evaluate the impact of contaminants on wild animals

Kei Nakayama *, Itsuki C. Handoh, Shin-Ichi Kitamura, Eun-Young Kim, Hisato Iwata, Shinsuke Tanabe

Center for Marine Environmental Studies (CMES), Ehime University, 2-5 Bunkyo-cho, Matsuyama 790-8577, Japan

ARTICLE INFO

Article history:
Received 26 April 2010
Received in revised form 30 August 2010
Accepted 31 August 2010
Available online 24 September 2010

Keywords:
Toxicogenomics
Persistent organic pollutants
Common cormorant
Euclidian distance
Principal component analysis

ABSTRACT

Here we propose a novel microarray data analysis method applicable to evaluation of the chemical effects on wild animals. First, we analyzed correlations between log-transformed hepatic 2,3,7,8-tetrachlorodibenzo-*p*-dioxin toxic equivalent (TEQ) levels and probe signals detected in wild cormorant liver to screen contaminant-responsive genes. Second, principal component analysis (PCA) was conducted using the screened probes. Third, these probes were divided into two groups based on our PCA result. Finally, we calculated Euclidian distance of signals, which is equivalent to variance of gene expressions, in each probe set, and analyzed the relationship between log-transformed hepatic TEQ levels and Euclidian distances. A probe set whereby the calculated Euclidian distance was positively correlated with TEQ levels, could indicate genes that were directly affected by dioxins or other persistent organic pollutants (POPs), hence they can be used as biomarkers. By contrast, there were a number of probes whereby the Euclidian distance was negatively correlated with TEQ levels. In the latter probe group, the smaller Euclidian distances in highly contaminated individuals could point to changes in physiological activities of wild cormorants. Therefore, our microarray data analysis method will provide new insights into POPs-responsive genes in field-collected samples for toxicogenomics studies.

1. Introduction

Microarray technology has been used in the field of (eco) toxicogenomics, and has been applied to a wide variety of organisms (Ankley et al., 2006; Lettieri, 2006). In most of toxicogenomics studies, the target species are largely limited to model animals in laboratory exposure tests, whereas some attempts have recently been made to employ the microarray technology in environmental monitoring and risk assessment.

Nakayama et al. (2006, 2008) constructed a microarray platform for wild common cormorants (*Phalacrocorax carbo*), and investigated the potential effects of persistent organic pollutants (POPs) and emerging POPs. Williams et al. (2003) developed a gene chip for European flounder (*Platichthys flesus*), and compared the gene expression profiles in the fish collected from a chemically contaminated site, to those from a non-contaminated reference site. Since microarray is a powerful tool to understand physiological status of organisms, and their responses to external stimuli associated with a chemical exposure, the technology could enable us to evaluate overall effects of environmental contaminants in wild animals, as well as in experimental animals.

A number of factors in the complex ecosystems may influence gene expression pattern as well as physiological status in wild animals. Our microarray study on wild cormorants succeeded in screening 99 potential POPs-responsive genes (Nakayama et al., 2006). Yet, definitive cause-and-effect relationships associated with POPs need to be clarified. Additionally, the exact functions of some genes, even annotated genes, remain unclear, which makes it difficult to evaluate the chemical effects on wildlife. Thus, in order to isolate POPs-responsive genes, it is essential to develop a suitable method for analysis of microarray data from field-collected samples. Kishi et al. (2006, 2008), Kishi et al. (2008) defined the degree of physiological activity by calculating the Euclidian distance and Pearson correlation from hundreds to thousands of gene expressions in Japanese medaka (*Oryzias latipes*). The distance and correlation reflect the magnitude of changes in the mRNA expression levels and the similarity in expression patterns, respectively. Combining these parameters, the authors estimated the physiological dysfunction that could be caused by the chemical exposures. We believe that this approach may maximally utilize the biological information obtained from the microarray experiment, and enables us to estimate physiological status from the gene expression profile.

In the present study, we constructed a new microarray platform for common cormorants, and analyzed the gene expression profiles in the liver tissues. To evaluate the effects of POPs on gene expressions, we modified the data analysis method that was set forth by Kishi et al. (2006), and verified whether the modified method was useful for a toxicogenomics study on wild cormorants.

* Corresponding author. Tel.: +81 89 927 8132; fax: +81 89 927 8133.
E-mail address: kei-n@agr.ehime-u.ac.jp (K. Nakayama).

0048-9697/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.scitotenv.2010.08.065

2. Materials and methods

2.1. Samples

We have employed the same common cormorant liver samples as those in our previous report (Nakayama et al., 2006, 2008). Seventeen cormorants were collected from Lake Biwa, Japan in May 2001 ($n=13$; 6 males and 7 females) and May 2002 ($n=4$; 3 males and 1 female). Hepatic total TEQ levels of the samples were also reported in our previous study (Nakayama et al., 2006). The samples were selected by hepatic TEQ levels, and their contaminant levels were uniformly distributed from low levels to high levels, which were judged on the basis of our previous studies (Kubota et al., 2006; Nakayama et al., 2006). Liver samples were immediately removed after the measurement of biometric data, and then the samples were frozen in liquid nitrogen and stored at −80 °C until RNA isolation.

2.2. Microarray experiment

The sequences of 6930 expression sequence tags (ESTs) in the cormorant cDNA library (Nakayama et al., 2006) were re-analyzed, and then 1955 tentative consensus sequences (TCs) were obtained. For each sequence, four to eight 60-mer oligonucleotide probes were designed (eArray, Agilent Technologies, Palo Alto, CA). The number of probes per TCs depends on the gene length. A 8×15 K format oligoarray was constructed from 12,410 probes in total (Agilent Technologies).

To prepare Cy3-labeled complement RNA (cRNA), total RNA was extracted from the liver tissue (≤400 mg) with TRIzol reagent (Invitrogen, Carlsbad, CA, USA), and the quality of each RNA sample was examined by denaturing agarose gel electrophoresis. We used 200 ng of the total RNA added One-Color Spike-Mix (Agilent Technologies) to synthesize Cy3-labeled cRNA using Low RNA Fluorescent Linear Amp Kit (Agilent Technologies). Cy3-labeled cRNA was purified using RNeasy Mini Kit (QIAGEN K. K., Tokyo, Japan). Hybridization and washing steps were conducted following the procedure provided by Agilent Technologies. The hybridized microarray slides were scanned by an Agilent DNA Microarray Scanner. Initial quality control of data from each slide was performed using Spike-In Control RNA. The intensities of spots were digitalized using Feature Extraction ver 9.5.3.1 (Agilent Technologies). Using the software, we then obtained more reliable data sets, through systematic removal of the data flagged as one of the following descriptions: *not positive and significant* or *not above background*, each of which indicates that the quantified values of these data are inaccurate.

2.3. Data analyses

The digitalized fluorescent intensities were normalized per chip by 50th percentile using GeneSpring GX 7.3 (Agilent Technologies). The data from 17 individuals were compared, and commonly available 8198 probes were extracted. The relationships between hepatic TEQ levels and gene expressions were analyzed by the Spearman rank-sum test. Here we used TEQ value as a representative of POPs concentration, since TEQ levels in cormorant were positively correlated with the concentrations of other POPs (Nakayama et al., 2006). The absolute values of correlation coefficients between TEQ levels and gene expressions ($|r|$) were used to create a few statistically meaningful data sets from all probes. For each data set containing the probes above a certain threshold based on $|r|$, we calculated the Euclidian distance (d), in gene expression (x_i) between N probes, following Kishi et al. (2006),

$$d = \sqrt{\sum_{i=1}^{N} (x_i - \overline{x})^2},$$

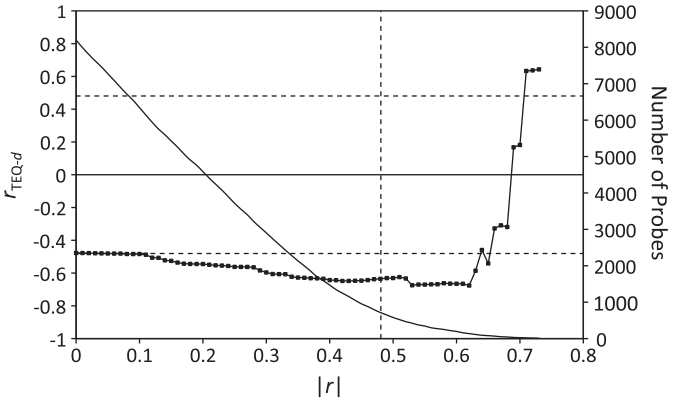

Fig. 1. Response of the correlation coefficients (r_{TEQ-d}) between log-transformed hepatic TEQ levels and Euclidian distances (d) (closed square) to the threshold $|r|$, where r is the correlation coefficient between log-transformed TEQ and gene expression levels ($n=17$). The numbers of probes used are also plotted (solid line). Dotted lines represent the statistical significance level of $p<0.05$ for a correlation coefficient.

where $\overline{x}$ is the mean gene expression. The relationships between log-transformed TEQ levels and Euclidian distances were also analyzed by the Spearman rank-sum test (r_{TEQ-d}).

We then applied a principal component analysis (PCA) to the gene expression values that were significantly correlated with log-transformed TEQ levels (the R project language, http://www.r-project.org/). PCA is an effective method for reducing high dimensionality and capture variations in gene expression values (Alter et al., 2000; Tomfohr et al., 2005). Given the sample size of 17, the absolute value of the threshold correlation coefficient ($|r|$) is 0.48. Prior to PCA, the gene expression values were normalized. It is important to note that our PCA employed both the above-mentioned Euclidian distance (variance of each sample), and gene expression covariance between samples. We computed the eigenvectors and eigenvalues of the variance-covariance matrix. The first and second principal components (PC1 and PC2) explained 50.2% and 13.4% of the total variance, respectively, and were well separated from the higher modes. For convenience, we have taken PC1 to separate the probes ($n=714$) into two groups, and calculated the Euclidian distance in each group. Subsequently, a linear regression analysis (SPSS 15.0J, SPSS, Chicago, IL, USA) was applied to evaluate a relationship between log-transformed hepatic TEQ level and the Euclidian distance.

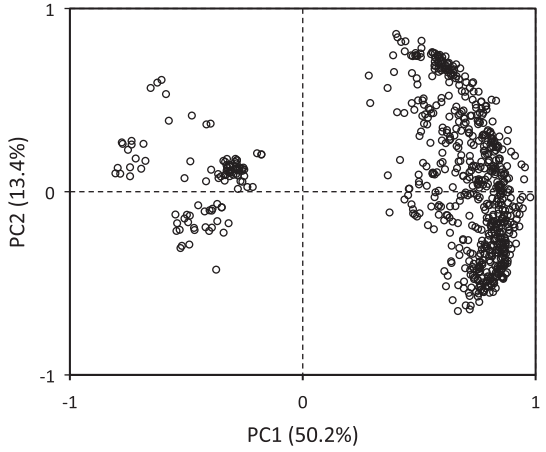

Fig. 2. Principal component scores of the normalized gene expression of the selected 714 probes whose gene expressions were significantly correlated with the log-transformed hepatic TEQ levels ($n=17$). The horizontal and vertical axes represent the first and second principal components, respectively.