PAGE: phase-shifted analysis of gene expression

Elo Leung and Pierre R. Bushe

1Bioinformatics and Computational Biology, School of Computational Sciences, George Mason University, Manassas, VA 20110, USA and 2National Institute of Environmental Health Sciences, National Center for Toxicogenomics, Microarray Group, Research Triangle Park, NC 27709, USA

ABSTRACT
Summary: Grouping of gene expression patterns across biological experiments, treatments and time-series data is performed in q-intervals of measurements using phase-shifted analysis of gene expression (PAGE); a Java-based tool to find clusters of genes that share trends of expression profiles within the dataset. The patterns and genes within q-Clusters are visualized in trend plots and compared to determine biological relevance from the gene annotations.

Availability: PAGE is available at http://dir.niehs.nih.gov/microarray/software/page/

Contact: bushel@niehs.nih.gov

Supplementary information: The Supplementary data are available at http://dir.niehs.nih.gov/microarray/software/page/

1 INTRODUCTION

Many methods that identify the correlation between gene clusters are designed to handle series data, which is derived from a biological condition and time or dose treatment. These methods include Cross-Correlation Function, Needleman–Wunsch alignment algorithm, and q-Clustering (Ji and Tan, 2005). Cross-Correlation Function and Needleman–Wunsch use scoring functions that are based on every pair of genes over all time points in the dataset. Both methods are computationally expensive and cannot locate the time interval at which the gene has high correlation with another gene. The q-Clustering method overcomes these limitations by using q-consecutive intervals to determine the correlation across and within the genes. An algorithm called rhythmic analysis of gene expression (RAGE) was developed to extract and characterize gene expression profiles from genome-wide microarray data based on periodic biological processes (Langmead et al., 2003). Similar to the former two methods, the current q-Clustering implementation and RAGE are limited to gene expression data from a dose or time-series experiment. Many datasets such as those generated for toxicogenomics, pharmacogenomics or clinical/disease/genetic informatics involve gene expression measurements across biological type, dose and time (Hamadeh et al., 2004; Hubner et al., 2005). A method that computes the correlations between gene clusters should be able to analyze data from biological condition, treatments and time points simultaneously so that the patterns can be compared.

A Java-based software, phase-shifted analysis of gene expression (PAGE), was developed to analyze gene expression data across multiple biological conditions, treatments and time series. PAGE was applied to a trans-compound dataset to extract the phase-shifts of gene expression in response to seven different toxicants at various dose treatments and time points.

2 METHODS AND IMPLEMENTATION

The gene expression data for analysis with PAGE were generated using rat liver samples analyzed on the Agilent RatTox oligo array. The data are available for download at http://dir.niehs.nih.gov/microarray/datasets/home-pub.htm and will be submitted to the Chemical Effects in Biological Systems Database (http://cebs.niehs.nih.gov/) as a subset of the NCT Compendium dataset. Java 2 Standard Edition and JFreeChart were used for development of PAGE.

The 433 differentially expressed gene patterns were first selected on a per agent, dose and time point manner, ranked as informative using relevance analysis and then filtered for genes with profiles that contain missing values.

The PAGE method is based on the q-Clustering method designed to identify time-lagged patterns of genes in time-series gene expression data (see Ji and Tan, 2005 for details). Briefly, the PAGE method has the following three phases:

- Phase 1: Gene expression pattern matrix transformation into −1, 0, 1 to indicate the direction of expression change from each biological condition at fixed time points and treatments. All biological replicates are averaged if provided.
- Phase 2: Generate q-clusters that have similar patterns of expression of over q-consecutive conditions.
- Phase 3: Assign a significance score for each bicluster in all q-Clusters and identify the inhibition patterns of each q-Cluster.

A window size of 3 and default threshold of 1.0 was used for specifying the q-interval consecutive points and the bin cut-off for defining the upward and downward trends, respectively. As the normalization threshold approaches zero, the bin for the unchanged gene expression from one point to the next shrinks, in turn the upward and downward gene expression bins expand, and vice versa.

3 APPLICATION FEATURES

PAGE uses a line graph to dynamically illustrate the phase-shifted patterns of gene expressions based on the q-Cluster selected. Each
line shown on the line graph represents the trend of a bicluster whose score is equal to or below the maximum threshold value. The line graph can be zoomed in and can be exported in jpg format. Also, the genes associated with the trends shown on the line graph can be exported to a text file. Furthermore, all phase-shifted patterns in each of the q-Clusters can be exported as a tab-delimited text file.

4 RESULTS FROM APPLICATION

4.1 q-Clustering at the gene level

The q-Clusters generated from the differentially expressed genes contain expression profiles that have the same and opposite patterns of expression. The average of the gene expression profiles in each bicluster is shown in Figure 1 to illustrate the trend similarity across four of the seven toxicants, which were determined to have patterns in phase-shift. Figure 1a shows the upregulated trends of gene expression profiles from acetaminophen, allyl alcohol, carbon tetrachloride and methapyrilene treatments, whereas Figure 1b shows the downregulated trends from these treatments. As shown in Figure 1b, the downregulated trends across the patterns of acetaminophen, allyl alcohol, carbon tetrachloride and methapyrilene suggest similarity in biological response at the gene level. These results are in agreement with Waring et al. (2001) but they also reveal patterns of genes from the methapyrilene time-series experiment, which are phase-shifted with the patterns of genes from the acetaminophen dose–response experiment.

4.2 Pathway mapping

The union of the genes from the two q-Clusters in Figure 1 was obtained for pathway mapping using DAVID (Dennis et al., 2003). Results indicate that folate biosynthesis is upregulated and fatty acid metabolism is downregulated. These results suggest the potential response mechanism(s) since the genes in these q-Clusters correlate with the acetaminophen biological response as revealed previously by Heinloth et al. (2004).

5 REQUIREMENTS FOR PAGE

The gene expression data should be normalized and transformed to the user’s satisfaction. In addition, missing data need to be imputed or profiles containing them removed, and all treatments and time points need to be consecutive with at least one interval (two adjacent points). The software has been tested on Windows PCs running the XP operating system and requires JRE version 1.4.2 or later.

ACKNOWLEDGEMENTS

The authors thank Maribel Bruno and Frank G. Bottone, Jr for review of this manuscript and testing of the software. This research was supported by the Intramural Research Program of the NIH, National Institute of Environmental Health Sciences. Requests for software should be addressed to E.L.

Conflict of Interest: none declared.

REFERENCES