BIOINFORMATICS APPLICATIONS NOTE Vol. 17 no. 6 2001
Pages 564–565

MAPS: a microarray project system for gene expression experiment information and data validation

Pierre R. Bushel[1], Hisham Hamadeh[1], Lee Bennett[1], Stella Sieber[2], Karla Martin[1], Emile F. Nuwaysir[1,3], Kate Johnson[1,3], Kelli Reynolds[1], Richard S. Paules[2] and Cynthia A. Afshari[1,*]

[1]Laboratory of Molecular Carcinogenesis and [2]Laboratory of Environmental Carcinogenesis/Mutagenesis, National Institute of Environmental Health Sciences, PO Box 12233, Research Triangle Park, NC 27709, USA

Received on November 29, 2000; revised and accepted on January 21, 2001

ABSTRACT

Summary: MAPS is a MicroArray Project System for management and interpretation of microarray gene expression experiment information and data. Microarray project information is organized to track experiments and results that are: (1) validated by performing analysis on stored replicate gene expression data; and (2) queried according to the biological classifications of genes deposited on microarray chips.

Availability: MAPS is accessible at http://www.dir.niehs.nih.gov/microarray/software/maps/

Contact: bushel@niehs.nih.gov

INTRODUCTION

DNA microarray technology has risen to the forefront of gene expression analysis as a tool by which researchers can detect genome-wide differential expression of thousands of genes. Interest in identifying transcription signature patterns has led to an explosion in the use of microarray technology to simultaneously analyze changes in gene expression profiles (Hughes *et al.*, 2000).

Statistical methods and information systems have been developed to facilitate the analysis and management of complex and large-scale microarray gene expression results. For example, methods have been developed where computed confidence intervals are used to determine significantly changed genes from a ratio distribution of gene expression data (Chen *et al.*, 1997). Differentially expressed genes can be validated by comparing replicate measurements of microarrays and performing subsequent biological assays to confirm the biological significance of altered gene expression (Wittes and Friedman, 1999). In addition, the microarray database ArrayDB was designed to store, analyze and associate gene expression data with information from remote biological resources (Ermolaeva *et al.*, 1998).

MicroArray Project System (MAPS) is a relational database management system with a web interface to cDNA microarray project information, validate replicate gene expression experimental results, and query gene expression data based on gene classifications of interest. In cDNA microarray analysis, gene expression ratios are computed from pixel intensity values acquired from 16 bit gray scale images. Differentially expressed genes are identified based on confidence intervals determined from a distribution of the ratios of intensity values, with significantly changed genes being those that fall outside of the confidence limits (Chen *et al.*, 1997). For a given confidence level the probability of chance occurrences in replicate experiments is determined using a binomial distribution model (Casella and Berger, 1990). Loading microarray information and data into MAPS allows a user to query for gene expression results which meet defined search parameters, validation criteria, and gene classifications of interest which are defined in the database by grouping genes into functional categories. MAPS, in contrast to ArrayDB, provides the advantage of associating experimental and biological microarray information with statistically validated gene expression data.

IMPLEMENTATION AND WEB INTERFACE

MAPS is implemented with a Microsoft Access database jet engine (Microsoft). Over 40 tables are defined in the database to manage microarray project information, detail experimental design, track clones, sample preparation,

*To whom correspondence should be addressed.
[3]Present address: NimbleGen Systems, LLC, University Research Park, 505 South Rosa Road, Madison, WI 53719, USA.

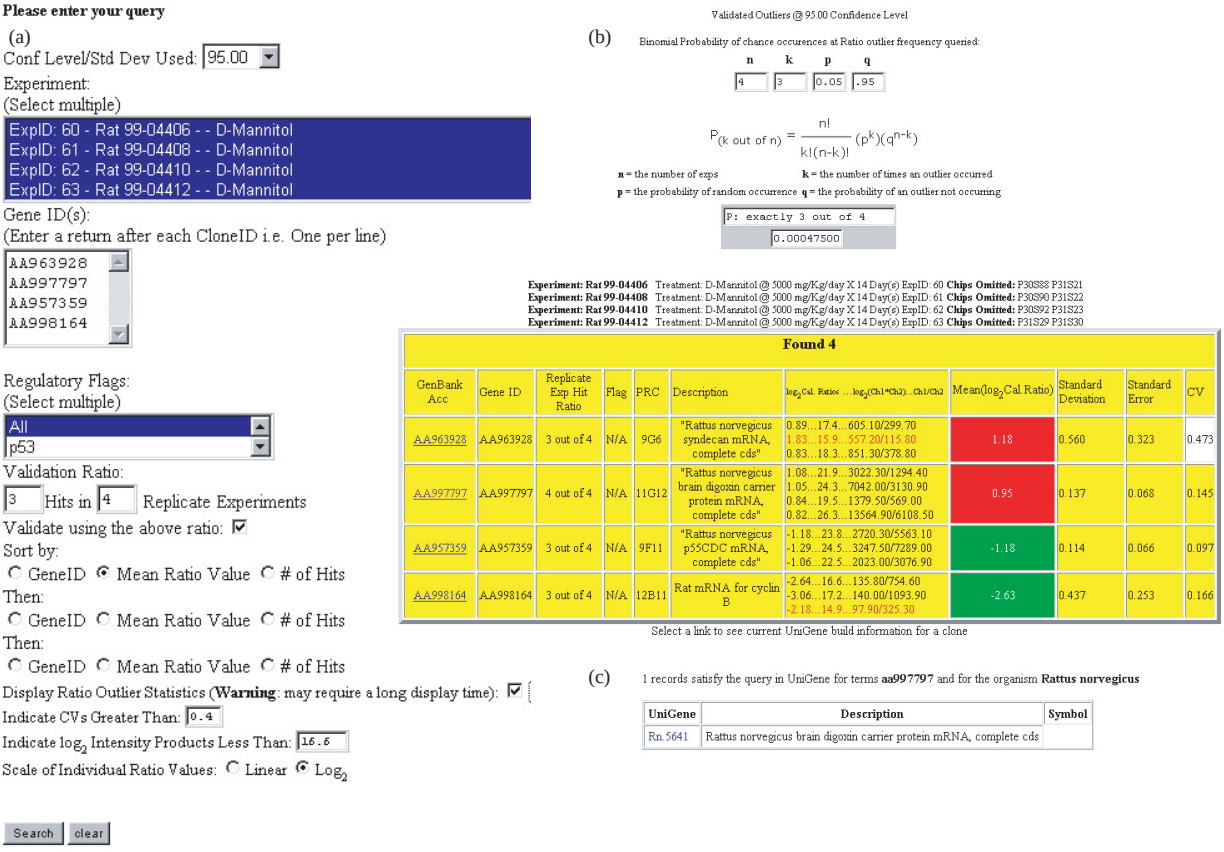

Fig. 1. MAPS differentially expressed gene validation. Screenshots from the MAPS gene expression validation pages illustrating the query form for detecting valid differentially expressed genes across multiple experiments (a), results of genes meeting specified search criteria with probability of chance occurrence (b) and display of the hyperlink to the cDNA clone current UniGene cluster (c).

labeling and hybridization, and survey the quality control and assurance of processed cDNA chips. Connection to the database is accomplished by open database connectivity, a ColdFusion application server (Allaire), and an IIS web server (Microsoft). Structured query language, proprietary ColdFusion tagging, and JavaScript are integrated in platform-independent web browser forms to provide users with seamless access to the database. Queried information and data are displayed in hypertext markup language tables and formatted to allow users to drill-down to detailed record information as well as hyperlink to interface with remote biological resources (Figure 1).

In the coming age of high-throughput genomics analysis using microarray technology scientists will rely heavily on robust database systems and statistical procedures to assist in the organization and interpretation of results. MAPS is a platform for scientists to not only manage information and analyze microarray gene expression data but also proceed towards discovering heuristic knowledge from biological information and data.

ACKNOWLEDGEMENTS

The authors would like to acknowledge our collaborators Raymond Stoll, Kerry Blanchard, and Supriya Jayadev for the study of rat liver gene expression, and thank Yidong Chen for ArraySuite data analysis software, Joseph Haseman for statistical consulting, Richard Lowry for JavaScript binomial probability calculations and Alex Merrick and Leping Li for review of the manuscript.

REFERENCES

Casella,G. and Berger,R.L. (1990) *Statistical Inference*. Duxbury Press, Belmont, CA.

Chen,Y., Dougherty,E.R. and Bittner,M.L. (1997) Ratio-based decisions and the quantitative analysis of cDNA microarray images. *J. Biomedical Optics*, **2**, 364–374.

Ermolaeva,O. *et al.* (1998) Data management and analysis for gene expression arrays. *Nature Genet.*, **20**, 19–23.

Hughes,T.R. *et al.* (2000) Functional discovery via a compendium of expression profiles. *Cell*, **102**, 109–126.

Wittes,J. and Friedman,H.P. (1999) Searching for evidence of altered gene expression: a comment on statistical analysis of microarray data. *J. Natl Cancer Inst.*, **91**, 400–401.