Pathway Miner: extracting gene association networks from molecular pathways for predicting the biological significance of gene expression microarray data

Ritu Pandey, Raghavendra K. Guru and David W. Mount*

Bioinformatics Core, Arizona Cancer Center, University of Arizona, Tucson, AZ 85724, USA

ABSTRACT

Summary: We have developed a web-based system (Pathway Miner) for visualizing gene expression profiles in the context of biological pathways. Pathway Miner catalogs genes based on their role in metabolic, cellular and regulatory pathways. A Fisher exact test is provided as an option to rank pathways. The genes are mapped onto pathways and gene product association networks are extracted for genes that co-occur in pathways. The networks can be filtered for analysis based on user-selected options.

Availability : Pathway Miner is a freely available web accessible tool at http://www.biorag.org/pathway.html

Contact: mount@email.arizona.edu

The recent explosion in the number of gene expression experiments and the need to extract useful biological information from the resulting volumes of data has presented a new challenge, that is, to identify the most significant changes in gene expression and to interpret the results in terms of biological relationships. There are a number of web resources that provide data annotation services for gene expression datasets (Guffanti *et al*., 2002). Other tools mine gene ontology attributes [GoFish (Berriz *et al*., 2003), ChipInfo (Zhong *et al*., 2003)] and pathways to provide a framework for visualizing the data (Dahlquist *et al*., 2002; Bouton and Pevsner, 2002) organization of expression profiles (Grosu *et al*., 2002) or interactive modeling (Toyoda and Konagaya, 2003) of metabolic pathways. All of these resources have their own unique capabilities and are individually focused on specific functions or features.

We describe Pathway Miner, a novel tool that mines gene associations and networks in biological pathway information represented in the currently available resources. An integrated database at http://www.biorag.org contains metabolic, cellular and regulatory pathways for human and mouse gene products from three different open source pathway resources—KEGG at http://www.genome.ad.jp, BioCarta at http://www.biocarta.com and GenMAPP at http://www.genmapp.org. Pathway Miner is an interactive user-friendly application that allows users to selectively view, analyze, interpret and download pathway information and networks extracted from the underlying database for high-throughput analysis of gene expression data. The relational database that stores the functional relationships resides on a 900 MHz processor Sun Fire 280R Unix Server that uses 700 GB RAID system for data storage. The website is driven by the Apache server. MySQL is used as the relational database management system.

Pathway Miner is a web tool and uses a client application to access the pathway database, extract and display the network. The client is written using the Java Swing API and runs on platforms with Java run-time environment version 1.4 or higher.

The tool provides two options to analyze genes in the dataset: (1) to search genes based on their associations in metabolic and/or cellular and regulatory pathways from the pathways resources and (2) in addition to the above, perform a statistical test and rank significant pathways based on their P-values from each of the three resources. A one sided Fisher exact test is implemented in Pathway Miner. This test uses the script by Øyvind Langsrud at http://www.matforsk.no/ola/fisher.htm. Both the options include a fold change cutoff filter that is used for removing less interesting data points. Interactive HTML outputs of organized pathway profiles are produced with options for studying the pathway maps and exploring a graph-based gene association network that is extracted from each of three internet pathway resources separately. The HTML outputs are arranged in multiple ways to (1) view sample-specific pathway profiles, (2) compare profiles between samples, (3) reveal pathways that are the most highly represented in the datasets or (4) list gene products that participate in single or multiple pathways. Expression values from as many as four samples can be

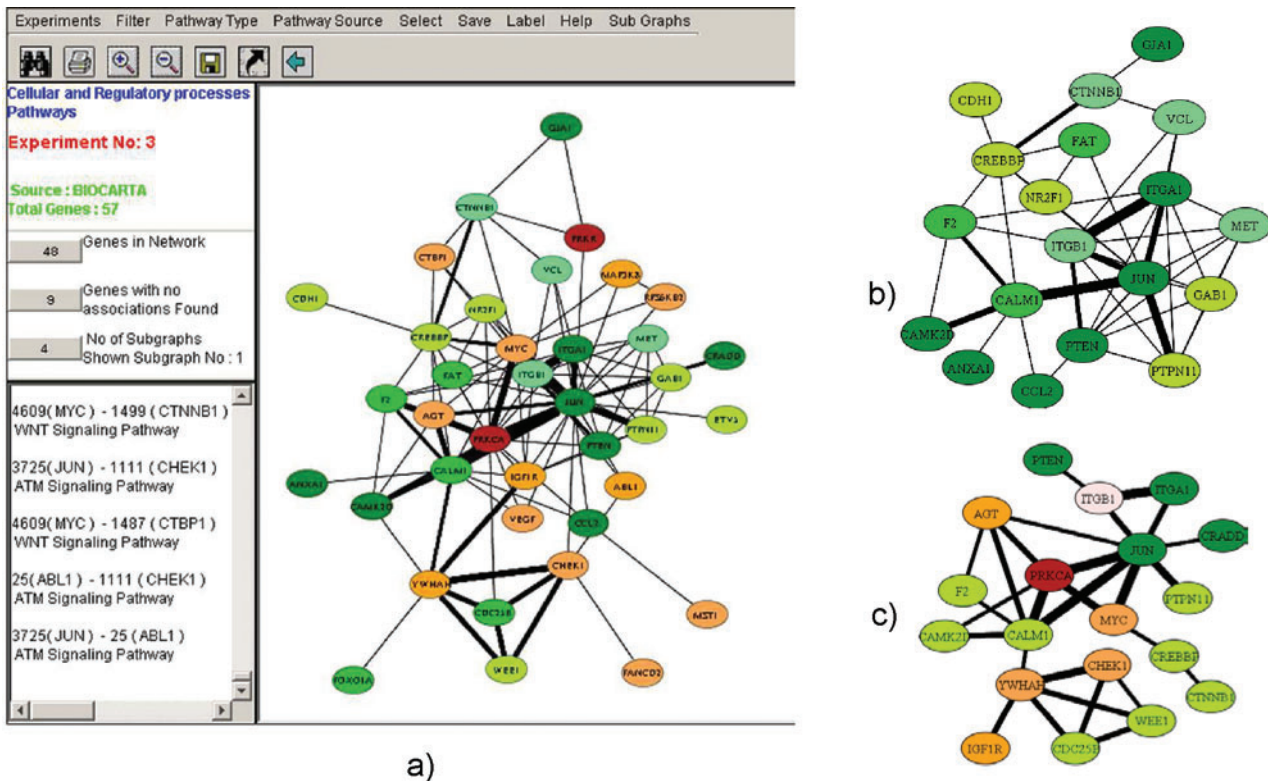*To whom correspondence should be addressed.

Fig. 1. Examples from Pathway Miner. (**a**) Default layout of the graph network. (**b**) Network filtered for down-regulated genes. (**c**) Network filtered for nodes with chosen edge strength. Nodes are labeled using the gene names.

compared on any pathway map from KEGG, BioCarta or GenMAPP and their corresponding graph networks can be extracted. Starting with an organized profile of genes and pathways based on the user dataset, an association graph network for the genes whose proteins function together in one or more pathway is produced. This graph network is drawn as a two-dimensional layout using the Neato program in the Graphviz software (www.research.att.com/sw/tools/graphviz/) and is displayed in a network browser that runs as a Java applet. Single network or multiple subnetworks of gene associations are produced where the nodes in the graph represent genes. Nodes are colored based on the expression values of up- and down-regulated genes (Fig. 1a). The edges indicate relationships between the genes in a pathway. The edges are given weights based on the number of pathways in which the associating nodes (gene products) co-occur together in the selected resource (KEGG, BioCarta or GenMAPP). Networks can be produced for multiple experiments for comparing the expression patterns of genes. These networks can be filtered based on various criteria such as displaying only the nodes that reflect up- or down-regulated genes (network of down-regulated genes shown in Fig. 1b) or those that have edge strength of a certain value (Fig. 1c). The network browser provides options for performing several other operations for analyzing the gene network. The details for Pathway Miner

and network browser are available at http://www.biorag.org/help.html#pathway and http://www.biorag.org/applets/network_browser_help.html

Gene association networks are very helpful for analyzing pathway relationships among genes that are found to be co-regulated in expression data, and for analyzing up- and down-regulated pathways that have several participating genes. Most importantly, the tool can assist in finding all pathways that are perturbed by genes whose expression is varying under given conditions. Other tools that independently map expression data to individual pathway maps are available. However, Pathway Miner has the unique features of extracting information from multiple pathways in which gene associations are found and presenting a global picture of the behavior of genes and pathways. We expect that Pathway Miner will be a beneficial tool for biologists.

ACKNOWLEDGEMENTS

REFERENCES

Berriz,G.F., White,J.V., King,O.D. and Roth,F.P. (2003) GoFish finds genes with combinations of Gene Ontology attributes. *Bioinformatics*, **19**, 788–789.

Bouton,C.M. and Pevsner,J. (2002) DRAGON view: information visualization for annotated microarray data. *Bioinformatics*, **18**, 323–324.

Dahlquist,K.D., Salomonis,N., Vranizan,K., Lawlor,S.C. and Conklin,B.R. (2002) GenMAPP, a new tool for viewing and analyzing microarray data on biological pathways. *Nat. Genet.*, **31**, 19–20.

Grosu,P., Townsend,J.P., Hartl,D.L. and Cavalieri,D. (2002) Pathway processor: a tool for integrating whole-genome expression results into metabolic networks. *Genome Res.*, **12**, 1121–1126.

Guffanti,A., Reid,J.F., Alcalay,M. and Simon,G. (2002) The meaning of it all: web-based resources for large-scale functional annotation and visualization of DNA microarray data. *Trends Genet.*, **18**, 589–592.

Toyoda,T. and Konagaya,A. (2003) KnowledgeEditor: a new tool for interactive modeling and analyzing biological pathways based on microarray data. *Bioinformatics*, **19**, 433–434.

Zhong,S., Li,C. and Wong,W.H. (2003) ChipInfo: software for extracting gene annotation and gene ontology information for microarray analysis. *Nucleic Acids Res.*, **31**, 3483–3486.