BIOINFORMATICS APPLICATIONS NOTE Vol. 26 no. 18 2010, pages 2342–2344
doi:10.1093/bioinformatics/btq418

Systems biology

Advance Access publication July 13, 2010

MetPA: a web-based metabolomics tool for pathway analysis and visualization

Jianguo Xia[1] and David S. Wishart[1,2,3,*]

[1]Department of Biological Sciences, [2]Department of Computing Sciences and [3]National Research Council, National Institute for Nanotechnology (NINT), University of Alberta, Edmonton, AB, Canada

Associate Editor: Alfonso Valencia

ABSTRACT

Summary: MetPA (Metabolomics Pathway Analysis) is a user-friendly, web-based tool dedicated to the analysis and visualization of metabolomic data within the biological context of metabolic pathways. MetPA combines several advanced pathway enrichment analysis procedures along with the analysis of pathway topological characteristics to help identify the most relevant metabolic pathways involved in a given metabolomic study. The results are presented in a Google-map style network visualization system that supports intuitive and interactive data exploration through point-and-click, dragging and lossless zooming. Additional features include a comprehensive compound library for metabolite name conversion, automatic generation of analysis report, as well as the implementation of various univariate statistical procedures that can be accessed when users click on any metabolite node on a pathway map. MetPA currently enables analysis and visualization of 874 metabolic pathways, covering 11 common model organisms.

Availability: Freely available at http://metpa.metabolomics.ca

Contact: david.wishart@ualberta.ca

Received on May 17, 2010; revised on June 22, 2010; accepted on July 8, 2010

1 INTRODUCTION

Over the past decade, pathway analysis has emerged as an invaluable aid to understanding the data generated from various 'omics' technologies. As a result, a number of robust software tools have been developed to support pathway analysis for genomics and proteomics studies. These tools combine powerful statistical analysis with visualization capacities to help researchers identify significant pathways involved in the conditions under study. The first pathway analysis tools were typically designed to take a list of differentially expressed genes (or gene products) and compare the number of differentially expressed genes detected in each pathway of interest with the number of genes expected to be found in the given pathway just by chance—a procedure known as over representation analysis (Draghici *et al.*, 2003; Khatri *et al.*, 2002). Second-generation pathway analysis tools typically use normalized gene expression data to calculate the expression of biological pathways in association with phenotypes—a procedure generally known as gene set enrichment analysis (GSEA) (Goeman *et al.*, 2004; Hummel *et al.*, 2008; Subramanian *et al.*, 2005). More recent tools consider both the significance of gene expression changes and their topological characteristics in order to better evaluate their impact on the pathways of interest (Draghici *et al.*, 2007; Glaab *et al.*, 2010; Tarca *et al.*, 2009). To date, essentially all pathway analysis tools have been designed and developed for the analysis of genomic or proteomic data, but not metabolomic data. Here, we wish to introduce MetPA (Metabolomics Pathway Analysis), a dedicated pathway analysis and visualization tool to facilitate the use of these relatively new and powerful methods in metabolomic studies.

*To whom correspondence should be addressed.

2 METHODS

2.1 Pathway analysis

Pathway analyses in MetPA are conducted through three routes. Pathway enrichment analysis supports both over-representation analysis as well as GSEA-based approaches. The available algorithms include Fishers' exact test, the hypergeometric test, global test (Goeman *et al.*, 2004) and GlobalAncova (Hummel *et al.*, 2008). MetPA's pathway topological analysis is based on the centrality measures of a metabolite in a given metabolic network. Centrality is a local quantitative measure of the position of a node relative to the other nodes, and is often used to estimate a node's relative importance or role in network organization (Aittokallio and Schwikowski, 2006). Since metabolic networks are directed graphs, MetPA uses relative betweenness centrality and out degree centrality measures to calculate compound importance. The pathway impact is calculated as the sum of the importance measures of the matched metabolites normalized by the sum of the importance measures of all metabolites in each pathway. Finally, MetPA provides a number of univariate analyses performed at the compound level to provide a more detailed view of the distribution of individual metabolite concentrations with regard to phenotypes. They include the *t*-test, one-way analysis of variance (ANOVA), and linear regression.

2.2 Pathway library construction and visualization

The pathway data used in MetPA were downloaded as KGML files from the KEGG database (Kanehisa *et al.*, 2008). Chemical compounds and pathway topology information were parsed into graph models using the *KEGGgraph* package (Zhang and Wiemann., 2009). The current library contains 874 metabolic pathways from 11 model organisms including humans, mouse, Drosophila, Arabidopsis, *Escherichia coli*, etc.

Metabolic pathways are presented as a network of chemical compounds with metabolites as nodes and reactions as edges. The graph generation and manipulation were implemented using Graphviz (http://www.graphviz.org) and ImageMagick (http://www.imagemagick.org). This visualization system supports lossless zooming, dragging and linking operations based on Ajax (Asynchronous JavaScript with XML) technology (Berger *et al.*, 2007). All relevant information can be obtained by clicking on the corresponding graphical elements.

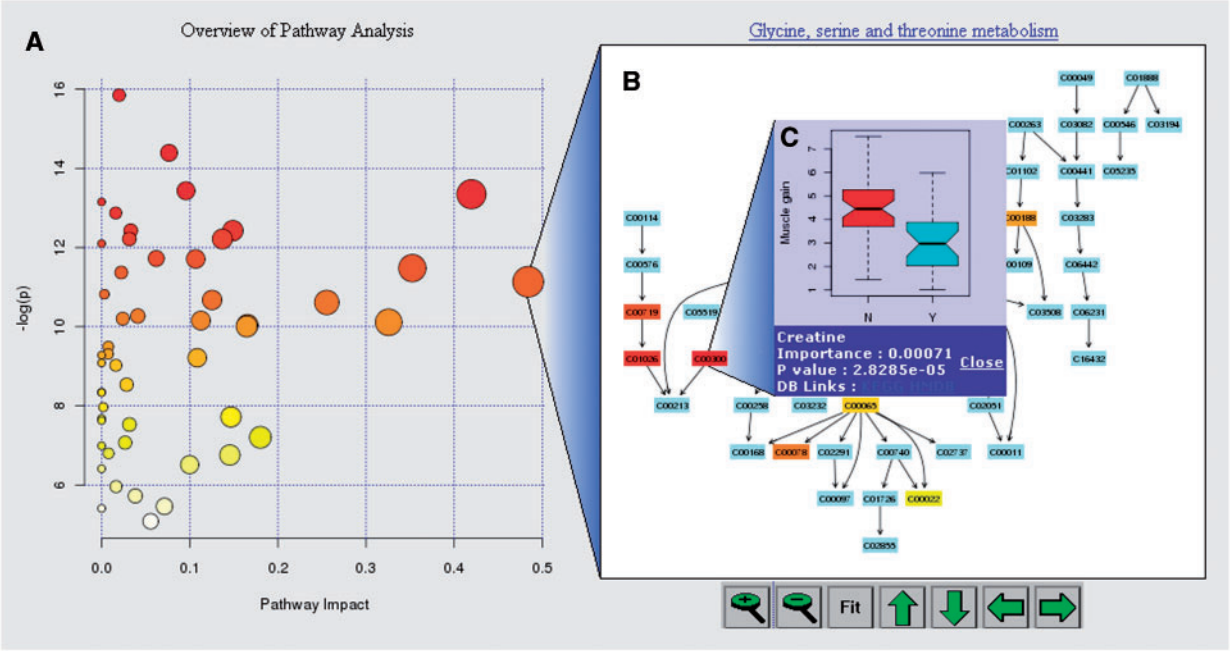

Fig. 1. Screenshot illustration of MetPA's data visualization features—(**A**) metabolome view, (**B**) pathway view and (**C**) compound view.

2.3 Implementation of web application

MetPA's web interface was implemented using the JSF or Java Server Faces (http://java.sun.com/javaee/javaserverfaces) framework. The pathway analysis algorithms were implemented in the R (version 2.10.0) programming language (http://www.r-project.org/). The communication between R and Java was established through the *Rserve* TCP/IP server (http://www.rforge.net/Rserve/). The web application is hosted on GlassFish (version 3) using a Linux operating system (Fedora Core 12). The server is equipped with two Intel Core 2 Quad processors (3.0 GHz each) and 8 GB of physical memory. The web application is platform independent and has been successfully tested on Mozilla Firefox 3.0+, Safari 4.0+, Google-Chrome 5.0+, Opera 10.0+ and Internet Explorer 8.0.

3 EXAMPLE ANALYSIS

MetPA accepts either a list of significant compound names, or a compound concentration table with binary, multi-group or continuous phenotype labels. In the latter case, it is advisable to first normalize the concentration data, i.e. using MetaboAnalyst (Xia *et al.*, 2009). As an example, we present the analysis on urinary metabolite concentration data (log-normalized) from cancer patients experiencing either muscle gain (Y) or muscle loss (N) monitored over a three-month period. The purpose is to investigate if certain metabolic pathways are significantly different between the two groups of patients. The first step is to convert the compound names of the uploaded data to the compound names used in the pathway library. MetPA uses compound names, synonyms and database IDs data from the HMDB (Wishart *et al.*, 2009) to perform compound name mapping. The next step is to specify the parameters for the pathway analysis—i.e. the pathway library, the algorithm for pathway enrichment analysis, as well as the algorithm for topological analysis. In this case, we select the 'Homo sapiens' library and use the default 'Global Test' and 'Relative Betweenness Centrality' for pathway enrichment analysis and pathway topological analysis, respectively. The result is presented in two parts—the graphical output (shown in Fig. 1) and a table containing all the analysis results. Users can intuitively explore the results by pointing and clicking on various hyperlinked nodes. For example, let us look at the 'Glycine, serine and threonine metabolism' pathway, which is the top pathway from the pathway topological analysis and is also significant in the pathway enrichment analysis (4.65E-5 after adjustment of multiple testing). Clicking the circle on the 'metabolome view' (Fig. 1A) on the left panel launches the corresponding 'pathway view' (Fig. 1B) on the right. It is interesting to see that many of these significantly changed amino acids are in key positions for this pathway. Further checking (by clicking on each metabolite node) indicates that all the nine matched amino acids show higher concentration values in the muscle loss group, with Creatine being the most significant (Fig. 1C). It is interesting to see that the most significant pathway identified from the enrichment analysis is 'Galactose metabolism' (highlighted as the dark red circle on the top left corner of the 'metabolome view'). Further checking indicates only three downstream peripheral compounds are involved, with 'Myoinositol' being most significant. It is less likely that this pathway is strongly associated with muscle change. The results from this analysis are currently under discussion with experts in the field.

4 CONCLUSIONS

The growing interest in metabolomics and systems biology has increased the need for computational and visual tools for pathway analysis. MetPA is a full-featured, easy-to-use pathway analysis and visualization environment that combines advanced statistical enrichment analysis with pathway topological characteristics to help

researchers identify the most relevant pathways involved in the conditions under study.

Funding: Alberta Ingenuity Fund (AIF), now part of Alberta Innovates - Technology Futures.

Conflict of Interest: none declared.

REFERENCES

Aittokallio,T. and Schwikowski,B. (2006) Graph-based methods for analysing networks in cell biology. *Brief. Bioinform.*, **7**, 243–255.

Berger,S.I. *et al.* (2007) AVIS: AJAX viewer of interactive signaling networks. *Bioinformatics*, **23**, 2803–2805.

Draghici,S. *et al.* (2003) Global functional profiling of gene expression. *Genomics*, **81**, 98–104.

Draghici,S. *et al.* (2007) A systems biology approach for pathway level analysis. *Genome Res.*, **17**, 1537–1545.

Glaab,E. *et al.* (2010) TopoGSA: network topological gene set analysis. *Bioinformatics*, **26**, 1271–1272.

Goeman,J.J. *et al.* (2004) A global test for groups of genes: testing association with a clinical outcome. *Bioinformatics*, **20**, 93–99.

Hummel,M. *et al.* (2008) GlobalANCOVA: exploration and assessment of gene group effects. *Bioinformatics*, **24**, 78–85.

Kanehisa,M. *et al.* (2008) KEGG for linking genomes to life and the environment. *Nucleic Acids Res.*, **36**, D480–D484.

Khatri,P. *et al.* (2002) Profiling gene expression using onto-express. *Genomics*, **79**, 266–270.

Subramanian,A. *et al.* (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA*, **102**, 15545–15550.

Tarca,A.L. *et al.* (2009) A novel signaling pathway impact analysis. *Bioinformatics*, **25**, 75–82.

Wishart,D.S. *et al.* (2009) HMDB: a knowledgebase for the human metabolome. *Nucleic Acids Res.*, **37**, D603–D610.

Xia,J. *et al.* (2009) MetaboAnalyst: a web server for metabolomic data analysis and interpretation. *Nucleic Acids Res.*, **37**, W652–W660.

Zhang,J.D. and Wiemann,S. (2009) KEGGgraph: a graph approach to KEGG PATHWAY in R and bioconductor. *Bioinformatics*, **25**, 1470–1471.