MetaNetter

Jourdan, Fabien; Breitling, Rainer; Barrett, Michael P.; Gilbert, David

Published in:
Bioinformatics.

DOI:
10.1093/bioinformatics/btm536

Publication date:
2008

BIOINFORMATICS APPLICATIONS NOTE Vol. 24 no. 1 2008, pages 143–145
doi:10.1093/bioinformatics/btm536

Data and text mining

MetaNetter: inference and visualization of high-resolution metabolomic networks

Fabien Jourdan[1,*], Rainer Breitling[2], Michael P. Barrett[3] and David Gilbert[4]

[1]UMR1089 Xénobiotiques INRA-ENVT, 180 chemin de Tournefeuille, St-Martin-du-Touch, BP3 31931 Toulouse, France, [2]Groningen Bioinformatics Centre, University of Groningen, 9751NN Haren, The Netherlands, [3]Institute of Biomedical and Life Sciences and [4]Bioinformatics Research Centre, University of Glasgow, Glasgow G12 8QQ, UK

Received and revised on August 9, 2007; accepted on October 17, 2007

Advance Access publication November 14, 2007

Associate Editor: Thomas Lengauer

ABSTRACT

Summary: We present a Cytoscape plugin for the inference and visualization of networks from high-resolution mass spectrometry metabolomic data. The software also provides access to basic topological analysis. This open source, multi-platform software has been successfully used to interpret metabolomic experiments and will enable others using filtered, high mass accuracy mass spectrometric data sets to build and analyse networks.

Availability: http://compbio.dcs.gla.ac.uk/fabien/abinitio/abinitio.html

Contact: Fabien.Jourdan@toulouse.inra.fr

Supplementary information: http://compbio.dcs.gla.ac.uk/fabien/abinitio/doc/Supplementary.pdf

1 INTRODUCTION

Metabolomics aims at the identification and quantification of all metabolites that are present in a biological sample. Various spectrometric technologies are capable of identifying thousands of metabolites. Recently, ultra high-resolution mass spectrometry (FTICR-MS or Orbitrap) has been successfully used in metabolomic studies (Aharoni *et al.*, 2002). Such high-resolution data has also been used to predict *ab initio* biochemical interactions between metabolites (Breitling *et al.*, 2006). Moreover, perturbation studies allow the use of correlation analysis to infer/confirm links between metabolites whose abundance correlates across various conditions.

The combination of these two inference methods generates networks containing hundreds of nodes (metabolites) and hundreds of predicted edges (biochemical reactions and/or high correlations). To analyse, explore and interpret these two kinds of relations, powerful visualization tools are required.

No currently available software allows inference and visualization of such high-resolution metabolomic networks directly from raw data. In this article, we present a new plugin for Cytoscape (Shannon *et al.*, 2003) for this purpose. Inference requires a list of potential biochemical transformations (e.g. Supplementary Material Table 1). The definition of this list may relate to experimentation (i.e. the organism or perturbation under study), hence, we provide facilities to edit/select putative biochemical transformations. The plugin also allows the extraction of parts of the network that contain a selected subset of reactions. Finally, to enrich the visual exploration, it is possible to highlight local topological properties of the network (e.g. degree or clustering index).

*To whom correspondence should be addressed.

2 PROGRAM OVERVIEW

2.1 Input

Ab inito inference (Breitling *et al.*, 2006) requires a list of masses (one mass per line) and a transformation list (Supplementary Material Table 1). To add correlation links, it is necessary to add quantitative measures for each mass. Thus, in the import format, each line begins with a mass followed by tabulated quantitative values if available. Of course before using MetaNetter, careful pre-processing is essential. This can be done by using standard software that accompanies Orbitrap and FTICR-MS machines, or any of a plethora of more advanced methods currently under development.

2.2 *Ab initio* network inference

A metabolic network is a combination of biochemical transformations, which turn one molecule (substrate) into another (product) under the action of a given enzyme. Given two molecules, A (substrate) and B (product), of molecular weight w_A and w_B, we can compute the molecular weight difference $w_X = |w_A - w_B|$ corresponding to a specific transformation. For instance, a carboxylation reaction will be associated to a mass difference of 43.98983, which is the molecular weight of CO_2. The user can define a list of possible transformations according to biological knowledge or general textbook information. We have provided a default list as a guide. A simple example involving the glycolytic subnetwork is also provided (Supplementary Material). The *ab initio* process (Breitling *et al.*, 2006) involves finding whether the weight difference between any two metabolites fits with a transformation in the list. To account for the limited accuracy of mass spectrometry data, an accuracy threshold (p.p.m. value) is used.

Thus, the two parameters of this method are the p.p.m. value and the transformation list. The plugin permits both

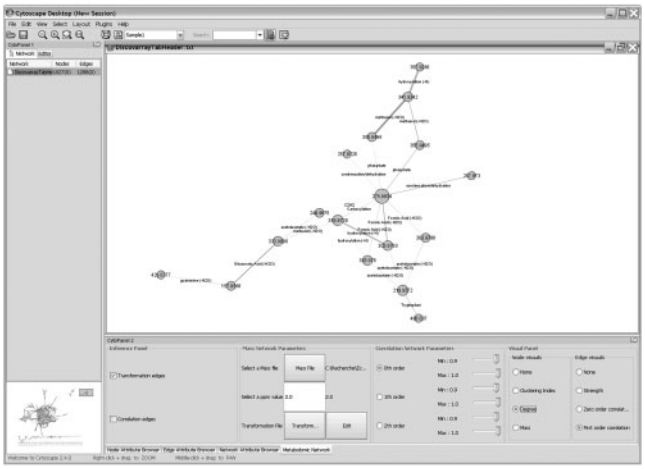

Fig. 1. *Ab initio* network built using various transformation types. Node size is proportional to the degree, edge stroke width is proportional to first order correlation and their colour is mapped to the transformation type.

parameters to be user defined. Particular care was taken to introduce flexibility into the transformation list (editing/selecting and loading/saving). To allow a first use of MetaNetter, we propose using a non-exhaustive transformation list (Breitling *et al*., 2006).

2.3 Correlation links

If the concentration of two metabolites is correlated across multiple samples, it is possible to infer a link between them [see Steuer (2006) for a detailed discussion on the interpretation of this kind of links in metabolomics]. This correlation can be quantified using the Pearson correlation coefficient and is widely applied to gene expression data (de la Fuente *et al*., 2004). Predictive quality can be enhanced by using partial correlation (Keurentjes *et al*., 2006), between two metabolites controlling for the effect of a third one (first-order correlation), another pair (second-order correlation) and so on. Correlations can also be used to confirm edges built from *ab initio* connectivity networks. For these reasons, the plugin allows the computation of zero-order correlation (Pearson coefficient) and first-order correlation. It is then possible to add correlation edges to the *ab initio* network, or to map correlation on the stroke width of edges (Fig. 1). This facility is also available in other programs like VANTED (Junker *et al*., 2006b), but these do not allow combining correlation and *ab initio* edges.

2.4 Visualization

Cytoscape is software dedicated to the visualization of biochemical networks (Shannon *et al*., 2003). We chose it since it is a stand-alone software, multi-platform and publicly available. Moreover, Cytoscape is designed in a modular way, allowing for the development of plugins such as the one presented here: MetaNetter. Once the user has defined the parameters of the *ab initio* computations, a view of the resulting network appears automatically (using one of the Cytoscape layout algorithms). Cytoscape also provides mapping tools (VizMapper) that allow the choice of labels to be displayed, defining colour mappings to nodes and edges (for instance, each occurrence of carboxylation may be coloured in red, etc.) and filtering on attribute values (for instance, selecting molecular weights above a given threshold). It is also possible to define external links to nodes. By right-clicking a node, the user can open a web page related to the selected node. We have, e.g. used it to link each node to reference pages in PubChem corresponding to masses of these nodes. In dealing with high-accuracy mass spectra, it is possible to identify small lists of putative chemical compounds for any given mass. By choosing within this list, the user can annotate the network. Annotation is generally a challenge in metabolomics since many metabolites are not yet identified. Network connectivity provides powerful clues to metabolite identity; the annotation of one node can support the identification of its neighbour.

Recently, much effort has been devoted to the computation of the local and global topological properties of biochemical networks (Jeong *et al*., 2000), such as node degree distribution or clustering index. Within the MetaNetter plugin, it is possible to compute and visualize these two metrics on nodes (metabolites). Moreover, since they will be considered as Cytoscape attributes on nodes, it is possible to export these values into spreadsheets suitable for further distribution analysis. Most of the topological properties used for biochemical network analysis are related to nodes (Junker *et al*., 2006a). We also propose to compute a topological property on edges. This metric, called Strength (Auber *et al*., 2003), is an extension of the clustering index to edges; it will highlight edges that are within a dense part of the network.

Cytoscape allows networks to be exported in graph format so the plugin can be used as a module of an external graph analysis tool. For instance, small connected subgraphs (e.g. Fig. 1) can be extracted and compared to textbook metabolic pathways. Graphic exports are also available for illustration purposes.

3 SUMMARY

In this article, we present a Cytoscape plugin dedicated to the inference and visualization of high-resolution mass spectrometry data sets. The inference is achieved using a defined list of putative biochemical transformations. The flexibility provided by this tool allows for isolation of transformation types in order to facilitate a focused analysis. A wide range of correlation analysis tools allows an even stronger inference of network connections than possible with mass difference analysis alone. In addition to the rich interactions provided by Cytoscape, the plugin offers a convenient way to visualize some topological properties of the networks.

ACKNOWLEDGEMENTS

Dave Watson and Muhammad Anas Kamleh kindly generated Orbitrap data sets. Alain Paris kindly advised on correlations.

Conflict of Interest: none declared.

REFERENCES

Aharoni,A. *et al*. (2002) Nontergeted metabolome analysis by use of fourier transform ion cyclotron mass spectrometry. *Omics*, **6**, 217–234.

Auber,D. *et al*. (2003) Multiscale visualization of small world networks. In: North,S.C. and Munzner,T. (eds.) *IEEE Symposium on Information Visualization*. pp. 75–81.

Breitling,R. *et al*. (2006) *Ab initio* prediction of metabolic networks using fourier transform mass spectrometry data. *Metabolomics*, **2**, 155–164.

de la Fuente,A. *et al*. (2004) Discovery of meaningful associations in genomic data using partial correlation coefficients. *Bioinformatics*, **20**, 3565–3574.

Jeong,H. *et al*. (2000) The large-scale organization of metabolic networks. *Nature*, **407**, 651.

Junker,B. *et al*. (2006a) Exploration of biological network centralities with centibin. *BMC Bioinformatics*, **219**.

Junker,B. *et al*. (2006b) VANTED: a system for advanced data analysis and visualization in the context of biological networks. *BMC Bioinformatics*, **7**, 109 EPub.

Keurentjes,J. *et al*. (2006) The genetics of plant metabolism. *Nat. Gene.*, **38**, 842–849.

Shannon,P. *et al*. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.*, **13**, 2498–2504.

Steuer,R. (2006) On the analysis and interpretation of correlations in metabolomic data. *Brief. Bioinformatics*, **7**, 151–158.