BIOINFORMATICS APPLICATIONS NOTE Vol. 26 no. 24 2010, pages 3140–3142
doi:10.1093/bioinformatics/btq602

Systems biology

Advance Access publication November 15, 2010

iMAT: an integrative metabolic analysis tool

Hadas Zur[1,*], Eytan Ruppin[1,2,*,†] and Tomer Shlomi[3,*,†]

[1]The Blavatnik School of Computer Science, [2]Sackler Faculty of Exact Sciences, The Sackler School of Medicine, Tel Aviv University, Tel Aviv 69978 and [3]Computer Science Department, Technion - Israel Institute of Technology, Haifa 32000, Israel

Associate Editor: John Quackenbush

ABSTRACT

Summary: iMAT is an Integrative Metabolic Analysis Tool, enabling the integration of transcriptomic and proteomic data with genome-scale metabolic network models to predict enzymes' metabolic flux, based on the method previously described by Shlomi *et al.* The prediction of metabolic fluxes based on high-throughput molecular data sources could help to advance our understanding of cellular metabolism, since current experimental approaches are limited to measuring fluxes through merely a few dozen enzymes.

Availability and Implementation: http://imat.cs.tau.ac.il/

Contact: zurhadas@post.tau.ac.il; ruppin@post.tau.ac.il; tomersh@cs.technion.ac.il

Supplementary information: Supplementary data are available at *Bioinformatics* online.

Received on July 18, 2010; revised on September 20, 2010; accepted on October 18, 2010

1 INTRODUCTION

Modern genomic sequencing technologies have enabled the rapid reconstruction of metabolic networks, giving rise to more than 50 highly curated metabolic reconstructions published to date (Oberhardt *et al.*, 2009), spanning all three domains of life including Eukaryota, Bacteria and Archaea. Additional computational methods to automatically reconstruct metabolic network models have recently resulted in draft network reconstructions for 160 microbial species (Henry *et al.*, 2010). Such reconstructed metabolic network models have been commonly used for metabolic phenotype prediction, metabolic engineering, studies of network evolution and biomedical applications (Oberhardt *et al.*, 2009). These studies employ various constraint-based modeling (CBM) methods to analyze the network function by solely relying on simple physical–chemical constraints (Price *et al.*, 2004).

Utilizing gene and protein expression to predict metabolic flux is a challenging task due to the complex mapping between the two. Previous studies have found a strong qualitative correspondence between gene expression and measured (Daran-Lapujade *et al.*, 2004) as well as predicted (Famili *et al.*, 2003) metabolic fluxes in microbes. However, the correlation between expression and metabolic flux is generally moderate and in some cases significant transcriptional changes do not reflect changes in flux, and vice-versa, significant changes in measured flux may not reflect transcriptional changes (Ovacik and Androulakis, 2008). These discrepancies may result from post-transcriptional regulatory processes that effect the actual levels of enzymes translated and from metabolic regulation, representing the effect of metabolite concentrations on the actual enzyme activity through allosteric and mass action effects (Rossell *et al.*, 2006).

Several CBM methods for analyzing and predicting metabolic flux distributions based on gene expression data have been suggested previously. The methods of Åkesson *et al.* (2004) and Becker and Palsson (2008) use gene expression data to identify genes that are absent or likely to be absent in certain contexts and search for metabolic states that prevent (or minimize) the flux through the associated metabolic reactions. Shlomi *et al.* (2008) consider data on both lowly and highly expressed genes in a given context as cues for the likelihood that their associated reactions carry metabolic flux, and employ constraint-based modeling (CBM) to accumulate these cues into a global, consistent prediction of the metabolic state. The latter method was shown to accurately predict human tissue metabolism, based on tissue-specific gene and protein expression data. Its application has demonstrated that in many cases, the activity of genes responsible for metabolic diseases is not directly manifested in enzyme-expression data, though can still be correctly predicted by expression integration with the metabolic network. The implementation of the method based on Shlomi *et al.* (2008) involves solving multiple, complex Mixed-Integer Linear Programming (MILP) optimization problems, requiring extensive parallel computing resources, and hence has not been readily accessible for the research community since its publication.

Here, we present an integrative metabolic analysis tool (iMAT) that is a web-based implementation based on the method of Shlomi *et al*. The new tool will serve the community by enabling the prediction of the metabolic state of an organism in a specific condition given pertaining gene and protein expression data. We provide below a high-level description of the iMAT server with an illustrative example of applying it to a toy model.

2 TOOL DESCRIPTION

The usage of iMAT is straightforward. The input is gene and/or protein expression data for a certain organism. The output is a visualization map of the organism's metabolic state, showing the most likely predicted metabolic fluxes across its reactions. iMAT supports the integration of functional data with an array of different existing metabolic CBM models, including: (i) a highly curated metabolic network model of human metabolism (accounting

*To whom correspondence should be addressed.

†The authors wish it to be known that, in their opinion, the last two authors should be regarded as joint First Authors.

for ∼1500 genes) by Duarte *et al.* (2007), enabling the prediction of metabolic activity under various tissues and cell types; (ii) common model organisms such as *Escherichia coli* and *Saccharomyces cerevisiae* (accounting for ∼1300 and ∼800 genes, respectively); (iii) an array of automatically reconstructed networks for 160 bacteria (Henry *et al.*, 2010) (for a list of supported organisms, see the iMAT web site); and (iv) user submitted models in either SBML or matlab format. For any of the organisms in this list, iMAT enables the prediction of metabolic activity under various environmental and genetic conditions.

The gene and/or protein expression data submitted to iMAT should be in the form of discrete tri-valued expression states, representing either low, moderate or high expression in the condition studied. If continuous data is submitted, iMAT will perform discretization automatically. Various parameters can be tuned to control the discretization of the raw input values (see Section 2, Supplementary Material).

Given the target species metabolic model and gene or protein expression data, iMAT predicts a flux activity state for each reaction in the model, reflecting the presence or absence of its associated metabolic flux. For some of the reactions, the flux activity state can be uniquely determined to be active or inactive, with associated confidence estimations. For others, the activity state cannot be uniquely determined because of potential alternative flux distributions with the same overall consistency with the expression data due to isozymes or alternative pathways. In cases where the predicted flux activity of reactions deviates from the given expression state of the corresponding enzyme-coding gene, the corresponding gene is considered to be post-transcriptionally up- or downregulated.

iMAT provides as output the predicted flux activity state and the corresponding confidence values over all network reactions in both tabular and network visualization forms. The network visualization displays the relevant transcriptomic and proteomic data given as input, as well as the predicted metabolic flux, superimposed on top of the organism's metabolic network, employing the publicly available Cytoscape software (Cline *et al.*, 2007). To further facilitate the interpretation of the predicted flux activities, iMAT performs a pathway enrichment analysis, reporting the significant active and inactive pathways comprising the metabolic profile signature of the biological experiment studied. In addition, iMAT reports predicted post-transcriptionally up- and downregulated genes.

3 TOY MODEL EXAMPLE

We describe the application of iMAT on a small toy model (Fig. 1). The toy model is comprised of 10 metabolites and 13 reactions, including 7 exchange reactions that enable the uptake of substrates and the secretion of metabolic byproducts. The predicted flux is consistent with the expression high/low state of four of the five reactions. One reaction (M6->M9) is predicted to be inactive though its corresponding gene is highly expressed, reflecting the potential effect of post-transcriptional regulation. Of the seven metabolites that can be transported across the membrane boundary in the toy model (M1-3, M5, M7-9), iMAT predicts the uptake of one metabolite (M1) and the secretion of two others (M7 and M8). The reaction M1->M4 is predicted to be active with low confidence, since an alternative flux distribution through M1->M10->M4 in which it is inactivated achieves the same level of consistency with the expression data.

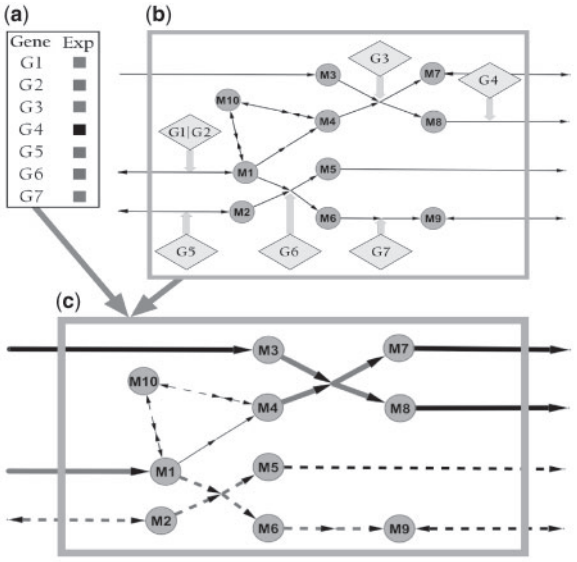

Fig. 1. (**a**, **b**) An illustrative example of applying iMAT to a toy metabolic network (shown in **b**) given gene expression data [where green (red) denotes high (low) expression and black denoting an intermediate level, as depicted in sub-figure (**a**)]. Circular nodes represent metabolites, edges represent biochemical reactions and diamond-shaped nodes represent enzyme-coding genes. iMAT's output is an optimal flux distribution (**c**) that is the most consistent with the input expression data. Reactions associated with highly, lowly or moderately expressed genes are colored in green, red or black, respectively (**c**). Solid (dashed) edges represent reactions predicted to be active (inactive). Reactions whose flux activity state is uniquely determined to be active or inactive (across the entire space of alternative optimal flux distributions) are marked with thick edges.

ACKNOWLEDGEMENTS

We are grateful to Bernhard Palsson for providing us with metabolic network layouts, and to Christopher Henry for allowing us the use of the automatic model reconstructions. We thank Ephi Sachs, Naor Yehodaey and especially Alon Ardenboim for taking part in the iMAT server development. Special thanks go to Tomer Benyamini for his invaluable help along the way. The development of this tool has been made possible thanks to the generous support of the Edmond J. Safra Bioinformatics Program at Tel-Aviv University.

Funding: E.R. and T.S. research in modeling metabolism is supported by grants from the Israel Science Foundation (ISF).

Conflict of Interest: none declared.

REFERENCES

Åkesson,M. *et al.* (2004) Integration of gene expression data into genome-scale metabolic models. *Metab. Eng.*, **6**, 285–293.

Becker,S. and Palsson,B. (2008) Context-specific metabolic networks are consistent with experiments. *PLoS Comput. Biol.*, **4**, e1000082.

Cline,M. *et al.* (2007) Integration of biological networks and gene expression data using Cytoscape. *Nat. Protoc. Electronic Edn*, **2**, 2366.

Daran-Lapujade,P. *et al.* (2004) Role of transcriptional regulation in controlling fluxes in central carbon metabolism of Saccharomyces cerevisiae. A chemostat culture study. *J. Biol. Chem.*, **279**, 9125–9138.

Duarte,N. *et al.* (2007) Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc. Natl Acad. Sci. USA*, **104**, 1777.

Famili,I. *et al.* (2003) Saccharomyces cerevisiae phenotypes can be predicted by using constraint-based analysis of a genome-scale reconstructed metabolic network. *Proc. Natl Acad. Sci. USA*, **100**, 13134–13139.

Henry,C.S. *et al.* (2010) High-throughput generation and optimization of genome-scale metabolic models, submitted.

Oberhardt,M.A. *et al.* (2009) Applications of genome-scale metabolic reconstructions. *Mol. Syst. Biol.*, **5**, 320.

Ovacik,M. and Androulakis,I. (2008) On the potential for integrating gene expression and metabolic flux data. *Curr. Bioinform.,* **3**, 142–148.

Price,N. *et al.* (2004) Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nat. Rev. Microbiol.*, **2**, 886–897.

Rossell,S. *et al.* (2006) Unraveling the complexity of flux regulation: a new method demonstrated for nutrient starvation in Saccharomyces cerevisiae. *Proc. Natl Acad. Sci. USA*, **103**, 2166.

Shlomi,T. *et al.* (2008) Network-based prediction of human tissue-specific metabolism. *Nat. Biotechnol.*, **26**, 1003–1010.