*Bioinformatics*, 34(18), 2018, 3225–3227
doi: 10.1093/bioinformatics/bty331
Advance Access Publication Date: 24 April 2018
Applications Note



Systems biology

# MetaBridge: enabling network-based integrative analysis via direct protein interactors of metabolites

**Samuel J. Hinshaw[1], Amy H.Y. Lee[1], Erin E. Gill[1] and Robert E.W. Hancock[1,*]**

[1]Centre for Microbial Diseases and Immunity Research, University of British Columbia, Vancouver, BC V6T 1Z4, Canada

*To whom correspondence should be addressed.

Associate Editor: Bonnie Berger

Received on February 26, 2018; revised on March 13, 2018; editorial decision on April 19, 2018; accepted on April 20, 2018

## Abstract

**Summary:** Here, we present MetaBridge, a tool that collates protein interactors (curated metabolite-enzyme interactions) that influence the levels of specific metabolites including both biosynthetic and degradative enzymes. This enables network-based integrative analysis of metabolomics data with other omics data types. MetaBridge is designed to complement a systems-biology approach to analysis, pairing well with network analysis tools such as NetworkAnalyst.ca, but can be used in any bioinformatics workflow.

**Availability and implementation:** MetaBridge has been implemented as a web tool at https://www.metabridge.org, and the source code is available at https://github.com/samhinshaw/metabridge_shiny (GNU GPLv3).

**Contact:** bob@hancocklab.com

## 1 Introduction

With the decreasing costs of high-throughput technologies, (Goodwin *et al.*, 2016), systems biology studies increasingly involve multiple high-throughput assays employing different omics methods (transcriptomics, proteomics, metabolomics, etc.), providing data based on different levels of biological information flow (Karahalil, 2016). Running concurrent high-throughput assays supplements the individual weaknesses of each platform such as resolving power and analyte detection thresholds. However, each omics data type is usually analyzed independently. While independent analyses of these data types can provide useful information, integrated analysis offers direct compensation for the limitations of individual platforms (Kaever *et al.*, 2014).

Deriving methods for integrative analysis of metabolomics data are particularly difficult, but can yield some of the most promising results (Patti *et al.*, 2012). Metabolomics data are challenging to integrate with other omics platforms due the relative sparsity of data and the indirect relationship between the metabolome and other omics outputs. In contrast to transcripts and proteins, which can be usually directly linked to a source gene, metabolites are often the end result of multiple distinct biochemical pathways (Trivedi *et al.*, 2017).

This challenge has been tackled using a number of different approaches (Wanichthanarak *et al.*, 2015). However, each approach suffers from various drawbacks, from statistical approaches that are blind to the underlying biology to command-line tools which require significant technical expertise.

MetaBridge is a biologically-informed tool that identifies connections between metabolites and their interacting proteins, such as synthetic or degradative enzymes. MetaBridge scaffolds metabolites onto known biological reactions, accounting for the complex relationships between metabolites and enzymes influencing their concentrations. MetaBridge is available as a web tool that does not require any specialized knowledge and can be readily accessed through any modern web browser.

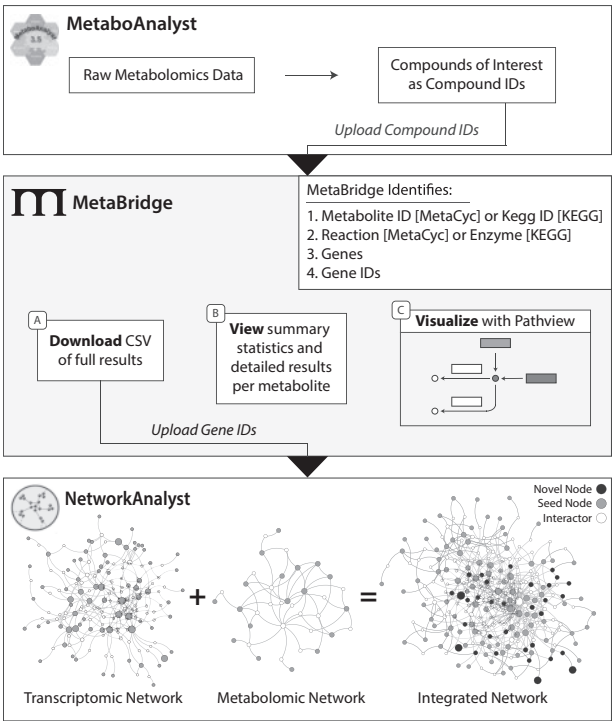


**Fig. 1.** Use of MetaBridge in an integrative analysis pipeline. The user begins by pre-processing their metabolomics data (MetaboAnalyst), identifying metabolites of interest and their IDs. The user then uploads these metabolite IDs to MetaBridge to identify directly interacting enzymes. The user receives a gene list, and can upload this list to e.g. NetworkAnalyst for network-based integration. Here, a metabolomic network (expanded by including first order interactors) has been integrated with a transcriptomic network, revealing novel nodes for investigation

## 2 Materials and methods

MetaBridge draws upon the curated biological pathway databases KEGG and/or MetaCyc to map metabolites to their directly interacting enzymes (Caspi *et al.*, 2016; Kanehisa *et al.*, 2017). MetaBridge is distributed in a visual interface using the interactive framework 'Shiny' developed by RStudio (Chang *et al.*, 2017).

Pre-processing must be conducted externally to obtain a list of metabolites of interest and corresponding metabolite IDs. We recommend MetaboAnalyst (Xia *et al.*, 2015) for metabolomics analysis and pre-processing, as shown in Figure 1.

The user uploads a CSV spreadsheet of their metabolites as KEGG, HMDB, PubChem, or CAS IDs. The user selects the column of their spreadsheet containing the metabolite IDs, and chooses the database to map against.

MetaBridge maps the provided metabolite IDs to their MetaCyc Object IDs (Fig. 1). Then, using pathway-tools, (Karp *et al.*, 2016) MetaBridge identifies all of the reactions in which the metabolite participates. Next, MetaBridge identifies all of the genes that encode for enzymes in these reactions. Finally, MetaBridge identifies the corresponding gene names and Ensembl gene IDs for each of the MetaCyc Gene IDs. Note that MetaCyc is not species-specific.

If the user did not upload KEGG IDs, MetaBridge converts the metabolite IDs to KEGG IDs (Fig. 1). Then MetaBridge identifies the KEGG-annotated interacting enzymes by EC number. Next, MetaBridge identifies the set of human genes that code for these enzymes. For each metabolite uploaded, MetaBridge also identifies the KEGG pathways in which the metabolite participates.

The user can download the complete results of this MetaBridge mapping as a CSV spreadsheet containing each metabolite, the reactions or enzymes for each metabolite and the genes and gene IDs of each reaction or enzyme. Additionally, a summary table is displayed, showing for each metabolite: (i) how many unique reactions in which each metabolite participates (MetaCyc) or how many unique enzymes with which each metabolite interacts (KEGG), and (ii) how many unique genes code for the enzymes in each reaction. The user can select any summarized metabolite to see its full mapping details. If the user has mapped metabolite interactors via KEGG, they have the option to visualize their results using Pathview (Luo and Brouwer, 2013).

We recommend integrative network analysis for comparative analysis with other omics types. Specifically, as demonstrated in Figure 1, NetworkAnalyst (Xia *et al.*, 2014) allows for generation of protein-protein interaction networks from these lists of enzymes. In our experience, these networks interconnect well with protein–protein interaction networks from other omics datasets.

## 3 Conclusion

MetaBridge enables a molecular interaction-based approach to assessing metabolomics data that draws on our previous success in utilizing high-quality curated biological data to facilitate biological inference (Yeung *et al.*, 2015). MetaBridge uses high-quality curated molecular interactions to identify directly interacting enzymes, enabling subsequent analysis alongside transcriptomic and proteomic data and thus biologically-relevant data integration.

MetaBridge is designed to be open and workflow-agnostic, compatible with other analysis techniques. MetaBridge is accessible in any web browser, and one can get started in seconds at metabridge.org.

## Acknowledgements

The authors thank Arjun Baghela and Dr. Daniel Pletzer for their beta-testing of MetaBridge. They thank Dr. Jianguo Xia for his advice in developing MetaBridge.

## Funding

Support from the Canadian Institutes for Health Research (FDN-154287) is gratefully acknowledged. S.J.H. is a recipient of The University of British Columbia 'Four Year Doctoral Fellowship'. Salary support from REWH from a 'Canada Research Chair in Health and Genomics' and 'UBC Killam Professorship' is gratefully acknowledged.

*Conflict of Interest*: none declared.

## References

Caspi,R. *et al.* (2016) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res.*, **44**, D471–D480.

Chang,W. *et al.* (2017) Shiny: Web Application Framework for R. https://shiny.rstudio.com/.

Goodwin,S. *et al.* (2016) Coming of age: ten years of next-generation sequencing technologies. *Nat. Rev. Genet.*, **17**, 333–351.

Kaever,A. *et al.* (2014) Meta-analysis of pathway enrichment: combining independent and dependent omics data sets. *PloS One*, **9**, e89297.

Kanehisa,M. *et al.* (2017) KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.*, **45**, D353–D361.

Karahalil,B. (2016) Overview of systems biology and omics technologies. *Curr. Med. Chem.*, **23**, 4221–4230.

Karp,P.D. *et al*. (2016) Pathway Tools version 19.0 update: software for pathway/genome informatics and systems biology. *Brief. Bioinform*., **17**, 877–890.

Luo,W. and Brouwer,C. (2013) Pathview: an R/Bioconductor package for pathway-based data integration and visualization. *Bioinformatics*, **29**, 1830–1831.

Patti,G.J. *et al*. (2012) Metabolomics: the apogee of the omics trilogy. *Nat. Rev. Mol. Cell Biol*., **13**, 263–269.

Trivedi,D.K. *et al*. (2017) Metabolomics for the masses: the future of metabolomics in a personalized world. *New Horizons Trans. Med*., **3**, 294–305.

Wanichthanarak,K. *et al*. (2015) Genomic, proteomic, and metabolomic data integration strategies. *Biomarker Insights*, **10s4**, BMI.S29511.

Xia,J. *et al*. (2014) NetworkAnalyst–integrative approaches for protein-protein interaction network analysis and visual exploration. *Nucleic Acids Res*., **42**, W167–W174.

Xia,J. *et al*. (2015) MetaboAnalyst 3.0—making metabolomics more meaningful. *Nucleic Acids Res*., **43**, W251–W257.

Yeung,A.T.Y. *et al*. (2015) Conditional-ready mouse embryonic stem cell derived macrophages enable the study of essential genes in macrophage function. *Sci. Rep*., **5**, 8908.