

# Bioinformatics Framework for the Analysis and Interpretation of Metabolomic Data

Alla Karnovsky<sup>1</sup>, Jing Gao<sup>2</sup>, Glenn Tarcea<sup>2</sup>, Christopher Beecher<sup>3</sup>, Charles Burant<sup>4</sup>, Barbara Mirel<sup>2</sup>, H.V. Jagadish<sup>2</sup>, Gilbert S. Omenn<sup>1,2,4</sup>

<sup>1</sup>Center for Computational Medicine and Biology, <sup>2</sup>National Center for Integrative Biomedical Informatics, <sup>3</sup>Michigan Center for Translational Pathology,

<sup>4</sup>Departments of Internal Medicine and Human Genetics, University of Michigan, Ann Arbor, MI 48109-2218

## BACKGROUND AND MOTIVATION

Metabolomics is a rapidly emerging field that is joining other high-throughput "omics", such as proteomics and transcriptional profiling. It promises to be a powerful systems approach for studying metabolic profiles pertinent to a variety of normal and disease states. Transcriptional profiling and proteomics have established data analysis tools; a metabolomics analytical toolkit is yet to be developed. We are creating a set of tools that will allow the user to examine experimental metabolomic data in the context of human metabolic networks and to combine them with other high throughput data.

A number of public sources contain information about human metabolic networks consisting of compounds, chemical reactions, pathways, enzymes and genes: KEGG (Kanehisa et al., 2008), BIGG (Duarte et al., 2007), EHMN (Ma et al., 2007). These networks are indispensable for the interpretation of experimental metabolomic data (Beecher, 2003). In this project, we used KEGG and EHMN data to trace the connections between metabolites and genes. Compounds, reactions, enzymes, genes and the relationships between them provide an initial framework for the analysis of metabolomic data. The Michigan Molecular Interactions database (MIMI) developed by the National Center for Integrative Biomedical Informatics (NCIBI) integrates protein interactions data from a number of public sources and thus can supply broader context for the analysis of the experimental data (Tarcea et al., 2009).

The data are stored locally in a Microsoft SQL Server database. They can be accessed either via a web-based query interface, or via Metscape, our new plug-in for Cytoscape (<http://www.cytoscape.org/>). Users can upload a list of metabolites or genes to the web interface, identify reactions, genes and pathways that are associated with those, and explore their relationships. In a parallel workflow users can import normalized experimental metabolite data directly into Metscape and display them within the network of metabolic reactions (Fig. 1).

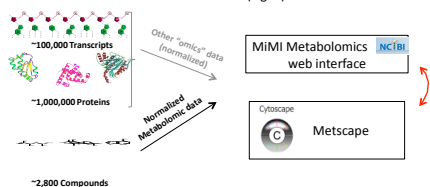


Figure 1. Suggested workflows. Users can browse the pathways, genes, enzymes and metabolites using web interface or Metscape (plug-in for Cytoscape) or upload a list of metabolites or genes of interest to MiMI-Metabolome web interface. Metscape currently allows users to upload normalized experimental metabolomic data. Web interface and Metscape are accessing the same database and will be tightly interconnected in the future. Our goal is to be able to bring normalized expression profiling and proteomics data into Metscape and display them side by side with metabolomic data (grey arrow).

## LINKS

MiMI Metabolome can be accessed at the following URL: <http://mimi.ncibi.org/MimiWebBeta/upload-page-metab.jsp> Cytoscape can be obtained from <http://www.cytoscape.org/> A beta version of Metscape can be downloaded via Cytoscape plug-in manager (under Network and Attribute I/O category).

## RESULTS

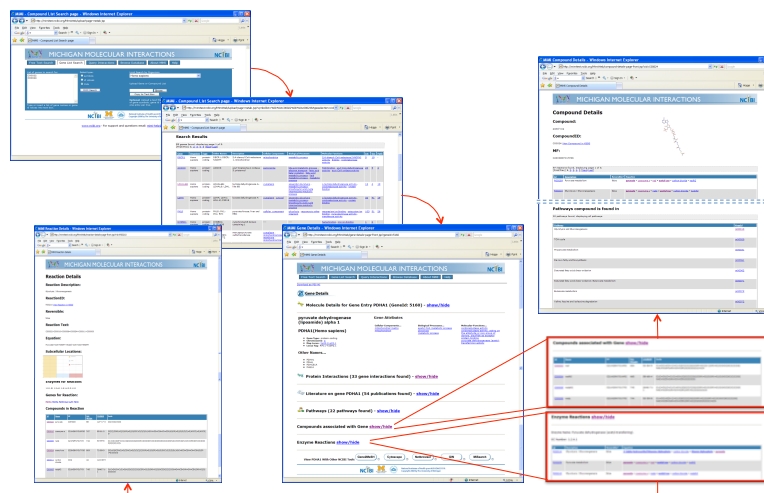


Figure 2. MiMI Metabolome web interface. The web tools allow the user to search the database for genes, reactions and pathways associated with a compound or gene. MiMI Metabolome is integrated with Michigan Molecular Interactions database (MIMI) that consolidates the wealth of information about protein-protein interactions.

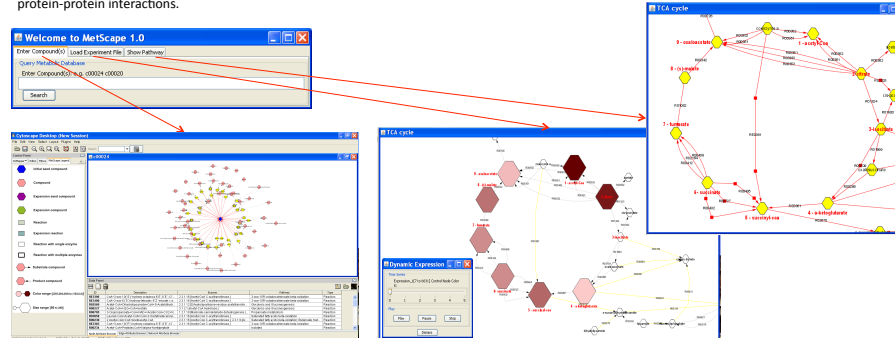


Figure 3. Metscape. The plug-in has three tabs that allow users to upload one or more compounds of interest, upload file containing experimental data, or display pathway specific networks. Currently there are two different styles for visualizing compound networks: in one both compounds and reactions are represented by nodes, in another the compounds are nodes and reactions are edges. Users can expand and collapse compound network. Metscape allows users to display the entire metabolic network, or pathways-specific networks. Users can apply pathway filter to a network and create sub networks from resulting subsets of compounds. Up to two different types of experimental measurements can be uploaded. The first is represented by changing node color; the second one is represented by changing node size.

## METHODS

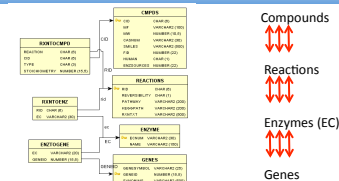


Figure 4. Partial schema of the Microsoft SQL server database and the relationships between metabolites and genes. Compounds, reactions, enzymes, genes and the relationships between them provide an initial framework for the data analysis.

## SUMMARY AND FUTURE PLANS

We developed a set of tools that has core functionality for visualizing experimental metabolomic data in context of human metabolism. Since metabolites are linked to genes, we can take advantage of multiple data sources (e.g. GO, enzyme databases) to enhance our current tools.

- We plan to fully implement the interconnectivity between MiMI Metabolome web interface and Metscape.
- We are currently working on enhancing the existing pathway layouts.
- We plan to use dynamic expression plug-in for Cytoscape ([http://chianti.ucsd.edu/cyto\\_web/plugins/index.php](http://chianti.ucsd.edu/cyto_web/plugins/index.php)) to visualize normalized expression profiling data side by side with metabolomic data.
- We are working on enhancing the compound search.

## ACKNOWLEDGEMENTS

This work is supported by NIH grant 1U54 DA021519 for the National Center for Integrative Biomedical Informatics.

## REFERENCES

- Beecher C., "The Human Metabolome" in "Metabolic Profiling: Its Role in Biomarker Discovery and Gene Function Analysis" eds. Harrigan G & Goodacre R, Kluwer Academic Publishers (Boston), 2003, 311 – 319.
- Duarte NC, Becker SA, Jamshidi N, Thiele I, Mo ML, Vo TD, Srivas R, Palsson BØ. Global reconstruction of the human metabolic network based on genomic and bibliomic data. Proc Natl Acad Sci U S A. 2007 Feb 6;104(6):1777-82.
- Gao J, Ade AS, Tarcea VG, Weymouth TE, Mirel BR, Jagadish HV, States DJ. Integrating and annotating the interactome using the MIMI plugin for cytoscape. Bioinformatics. 2009 Jan 1;25(1):137-8.
- Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, Katayama T, Kawashima S, Okuda S, Tokimatsu T, Yamanishi Y. KEGG for linking genomes to life and the environment. Nucleic Acids Res. 2008 Jan;36(Database issue):D480-4.
- Ma H, Sorokin A, Mazein A, Selkov A, Selkov E, Demin O, Goryanin I. The Edinburgh human metabolic network reconstruction and its functional analysis. Mol Syst Biol. 2007;3:135.
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Research. 2003 Nov; 13(11):2498-504
- Tarcea VG, Weymouth T, Ade A, Bookvich A, Gao J, Mahavisno V, Wright Z, Chapman A, Jayapandian M, Ozgur A, Tian Y, Cavalcoli J, Mirel B, Patel J, Radev D, Athey B, States D, Jagadish HV. Michigan molecular interactions 2: from interacting proteins to pathways. Nucleic Acids Res. 2009 Jan;37 (Database issue):D642-6.