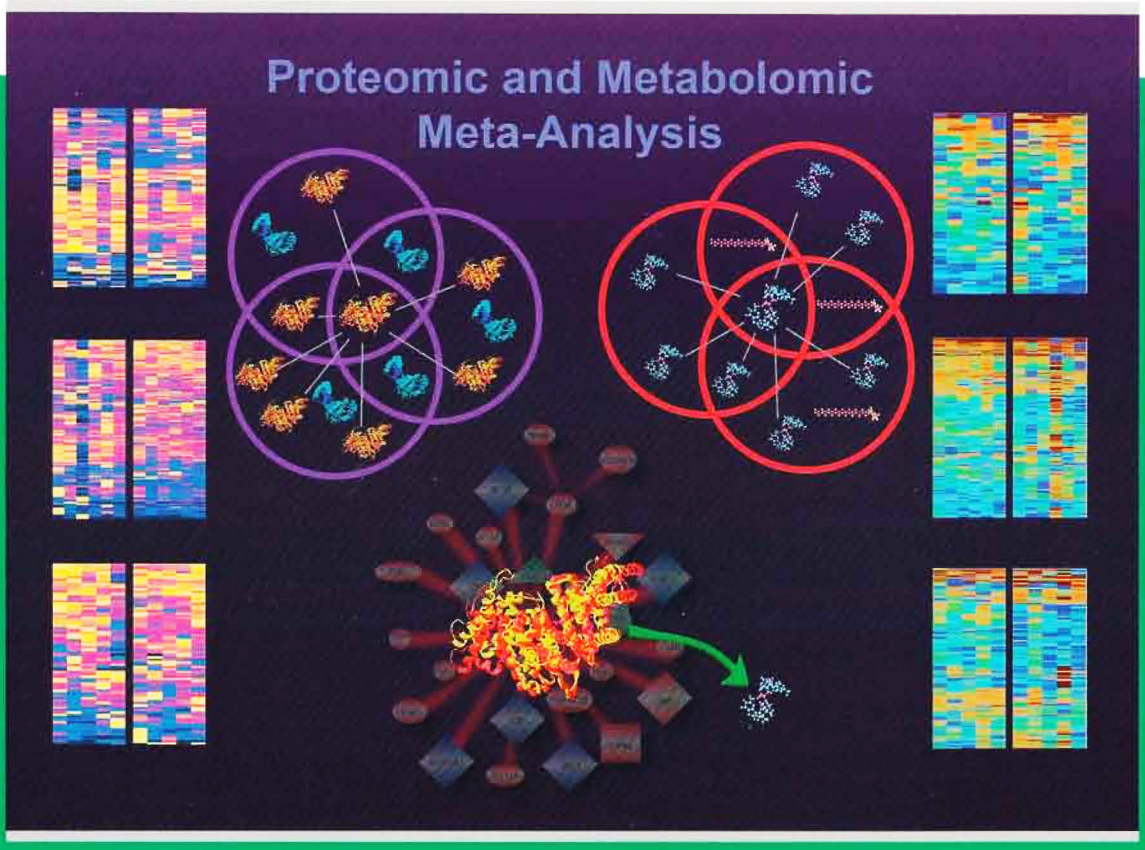
ISSN 0712-4813

Volume 26, Number 3, 2011

SPECTROSCOPY

BIOMEDICAL APPLICATIONS



IOS Press

26 (2011) 151–154
SPE-2011-0534

Meta-analysis of global metabolomics and proteomics data to link alterations with phenotype

Gary J. Patti [a], Ralf Tautenhahn [b], Bryan R. Fonslow [c], Yonghoon Cho [b], Adam Deutschbauer [d], Adam Arkin [d], Trent Northen [e] and Gary Siuzdak [b,*]

[a] *Departments of Chemistry, Genetics and Medicine, Washington University School of Medicine, St. Louis, MO, USA*
[b] *Center for Metabolomics and Mass Spectrometry, The Scripps Research Institute, La Jolla, CA, USA*
[c] *Department of Chemical Physiology, The Scripps Research Institute, La Jolla, CA, USA*
[d] *Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA*
[e] *Department of Bioengergy/GTL & Structural Biology, Life Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA*

Abstract. Global metabolomics has emerged as a powerful tool to interrogate cellular biochemistry at the systems level by tracking alterations in the levels of small molecules. One approach to define cellular dynamics with respect to this dysregulation of small molecules has been to consider metabolic flux as a function of time. While flux measurements have proven effective for model organisms, acquiring multiple time points at appropriate temporal intervals for many sample types (e.g., clinical specimens) is challenging. As an alternative, meta-analysis provides another strategy for delineating metabolic cause and effect perturbations. That is, the combination of untargeted metabolomic data from multiple pairwise comparisons enables the association of specific changes in small molecules with unique phenotypic alterations. We recently developed metabolomic software called metaXCMS to automate these types of higher order comparisons. Here we discuss the potential of metaXCMS for analyzing proteomic datasets and highlight the biological value of combining meta-results from both metabolomic and proteomic analyses. The combined meta-analysis has the potential to facilitate efforts in functional genomics and the identification of metabolic disruptions related to disease pathogenesis.

Keywords: Metabolomics, meta-analysis, proteomics, untargeted, XCMS

Our understanding of molecular biology has greatly evolved since Beadle and Tatum [2,10] first proposed the one gene-one enzyme hypothesis. Largely due to advances in the development of multicellular model organisms, there are now many examples in which nonlethal mutations or knockdowns of specific genes lead to unique phenotypes that are challenging to predict and understand based on our current knowledge of biochemical pathways alone [3–5]. Additionally, emerging metabolomic and proteomic technologies have demonstrated that defects in a single enzyme most frequently lead to a cascade of alterations at both the protein and metabolite levels (Fig. 1). It is in this context that the field of systems biology has excelled in identifying perturbations related to physiological function at the pathway level. Systems biology offers an opportunity to define the complex interactions between genes, proteins and metabolites and their resulting effect on phenotype. It has thus become a major goal of functional genomics and mechanistic disease studies to globally identify changes in protein activity and metabo-

*Corresponding author. E-mail: siuzdak@scripps.edu.

0712-4813/11/$27.50 © 2011 – IOS Press and the authors. All rights reserved

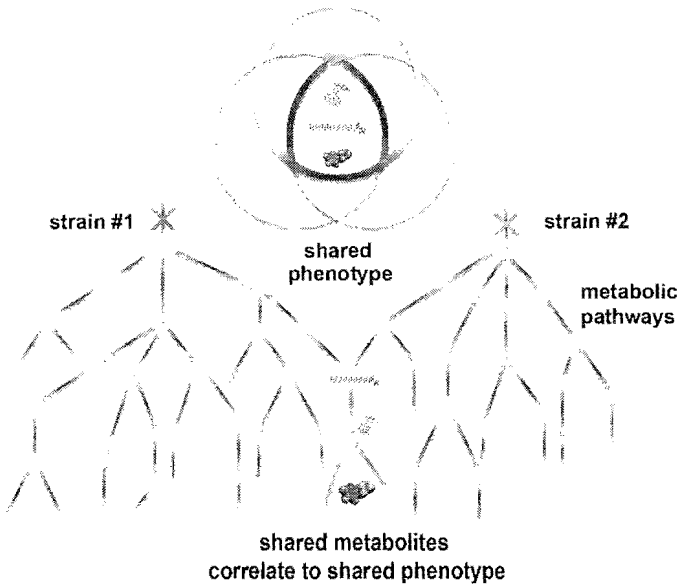


Fig. 1. Combined meta-analysis of metabolomic and proteomic data across different cell strains enables the association of specific pathways with phenotype. (Colors are visible in the online version of the article; http://dx.doi.org/10.3233/SPE-2011-0534.)

lite levels resulting from a genetic defect, and then to associate a specific subset of these changes with disease pathology.

Although rapid advances in mass spectrometry and bioinformatic tools have greatly facilitated global characterization of protein and metabolite alterations, defining subsets of these changes associated with a unique phenotype or disease state has proven difficult. A major challenge in interpreting the changes detected in metabolomic and proteomic data is distinguishing alterations that are the cause of metabolic dysfunction from those alterations that are a downstream cascade effect. Here, time-course experiments have been particularly useful for chronicling patterns of dysregulation and are often used as the basis for mathematical modeling of the perturbed system [1,6]. For metabolomic analysis, the flux of metabolites through a pathway can be quantitatively tracked by measuring the conversion of isotopically labeled substrates [12]. Introduction of these time-dependent omic analyses allows for the temporal correlation of specific changes in proteins and metabolites with the progression of a phenotype or disease symptom. While these data have the potential to provide significant biochemical insight, their application to mammals and clinical studies are often limited. Additionally, depending on the number of metabolic alterations, correlating a unique subset of molecular alterations with a specific phenotype can be challenging.

Recently, we developed metabolomic software called metaXCMS [8,11] as part of an alternative strategy for correlating changes in metabolite levels with a phenotype. metaXCMS is an extension of the original software XCMS [9] that was designed to globally compare untargeted metabolomic data from a pair of sample groups (e.g., wild type versus mutant, healthy versus disease, etc.). Unlike XCMS which is limited to comparing only 2 sample groups simultaneously, metaXCMS allows for the comparative analysis of multiple sample groups and thereby facilitates meta-analysis of metabolomic data. By designing intelligent comparisons, meta-analysis has been integrated in the metabolomic workflow as a

strategy for data reduction. Specifically, metabolite alterations that are not shared among variations of a phenotype can be excluded from data analysis as a source of "biological noise" that is not directly contributing to the phenotype of interest. As an example, 3 different models of pain were investigated with different pathogenic etiologies representing an inflammatory model, an acute heat model, and a model of spontaneous arthritis [11]. An analysis of altered metabolites in the three models relative to each respective control resulted in a sum total of 1825 peaks of statistical significance for the 3 pairwise comparisons. Only 3 out of the 1825, however, were shared between the models. One of the shared alterations was determined to be histamine, a well-established chemical mediator of pain.

Metabolites are the substrates and products of biochemical pathways and therefore provide a direct functional readout of cellular activity that is independent of regulatory processes that complicate other omic measurements such as epigenetic factors and post-translational modifications. Thus, for untargeted profiling, metabolite analysis facilitates identifying perturbations in global metabolism. In a therapeutic context, however, pharmacological intervention is most commonly at the level of proteins. Given that many metabolites can be biosynthesized from a number of unique substrates, identifying a protein target for therapeutic manipulation may require additional pathway mapping and metabolomic analysis. An increased level of pyruvate, for example, could be the result of dysregulated pyruvate biosynthesis from glycolysis, lactate, oxaloacetate, malate, formate, acetyl-CoA, acetyl phosphate or nicotinamide metabolism. Additionally, a complete list of pathway connectivities for many metabolites has yet to be established, further complicating the deconvolution of shared metabolic hubs.

As a complement to the metabolite profiling, a meta-comparison of the same sample groups by proteomics has the potential to validate the metabolic alterations determined to be associated with the phenotype from the metabolomic results. Furthermore, the proteomic data may ultimately provide higher pathway specificity. Continuing with the example described above, each of the biochemical conversions listed for pyruvate is catalyzed by a different protein and therefore have the potential to be resolved by proteomic analysis. To accomplish meta-analysis of proteomic data, liquid chromatography/mass spectrometry (LC/MS) data may be input into metaXCMS by using the same workflow as previously demonstrated [11]. Altered peaks determined to be shared among sample groups can subsequently be identified as peptides and searched in protein databases.

The relatively high-throughput capacity of LC/MS [7] combined with advances in the manipulation of genetically tractable model organisms has enabled the characterization of metabolites and proteins on a global scale for an exceedingly large number of unique phenotypes. Surprisingly, global analysis of genetically engineered multicellular organisms has revealed changes in the metabolites and proteins of pathways that are unexpected based on our current understanding of chemical physiology. These large metabolomic and proteomic datasets offer exciting opportunities for meta-analysis and their synergy has great potential for elucidating the relationship between metabolic function and phenotype as well as identifying new molecules in unexplored metabolome and proteome space.

Acknowledgements

This work was supported by the California Institute of Regenerative Medicine (TR1-01219), the National Institutes of Health (R24 EY017540-04, P30 MH062261-10 and P01 DA026146-02) and NIH/NIA L30 AG0 038036 (G.J.P.). Financial support was also received from the Department of Energy (Grants ENIGMA DE-AC0205CH11231).

References

[1] D. Amador-Noguez et al., Systems-level metabolic flux profiling elucidates a complete, bifurcated tricarboxylic acid cycle in *Clostridium acetobutylicum*, *J. Bacteriol.* **192**(17) (2010), 4452–4461.
[2] G.W. Beadle and E.L. Tatum, Genetic control of biochemical reactions in neurospora, *Proc. Natl. Acad. Sci. USA* **27**(11) (1941), 499–506.
[3] M.Q. Dong et al., Quantitative mass spectrometry identifies insulin signaling targets in *C. elegans*, *Science* **317**(5838) (2007), 660–663.
[4] S. Kang, A.G. Bader and P.K. Vogt, Phosphatidylinositol 3-kinase mutations identified in human cancer are oncogenic, *Proc. Natl. Acad. Sci. USA* **102**(3) (2005), 802–807.
[5] C. Kenyon et al., A *C. elegans* mutant that lives twice as long as wild type, *Nature* **366**(6454) (1993), 461–464.
[6] Y.K. Kwon et al., A domino effect in antifolate drug action in *Escherichia coli*, *Nat. Chem. Biol.* **4**(10) (2008), 602–608.
[7] G.J. Patti, Separation strategies for untargeted metabolomics, *J. Sep. Sci.* (2011), in press.
[8] G.J. Patti, R. Tautenhahn and G. Siuzdak, Meta-analysis of untargeted metabolomic data from multiple profiling experiments, *Nat. Protoc.* (2011), in press.
[9] C.A. Smith et al., XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification, *Anal. Chem.* **78**(3) (2006), 779–787.
[10] E.L. Tatum and G.W. Beadle, Genetic control of biochemical reactions in neurospora: an "aminobenzoicless" mutant, *Proc. Natl. Acad. Sci. USA* **28**(6) (1942), 234–243.
[11] R. Tautenhahn et al., metaXCMS: second-order analysis of untargeted metabolomics data, *Anal. Chem.* **83**(3) (2011), 696–700.
[12] N. Zamboni et al., (13)C-based metabolic flux analysis, *Nat. Protoc.* **4**(6) (2009), 878–892.