Identification of differentially expressed metabolites in Desulfovibrio and Communities using meta-analysis and the METLIN database

Richard Baran[1], Cho, Kevin[2], Benjamin Bowen[1], Paul Adams[1], Adam A. Arkin[1], Adam M. Deutschbauer[1], Linh Hoang[2], Ewa Kalisiak[2], Nitin S. Baliga[3], Zhengjiang Zhu[2], Ralf Tautenhahn[2], Trent R. Northen[1], **Siuzdak, Gary**[2]* (siuzdak@scripps.edu)

Microbial Biotechnology siuzdak@scripps.edu

[1]*Lawrence Berkeley National Lab, Berkeley CA;* [2]*The Scripps Center for Mass Spectrometry, La Jolla CA;* [3]*Institute for Systems Biology, Seattle WA*

http://enigma.lbl.gov *http://masspec.scripps.edu/*

Project Goals: Our objectives are to examine microorganisms and their communities and measure both their chemical input and output as a read out of specific biochemical activity. This effort will build on our previous experience of complex microbial systems by examining how bacteria communities rely on symbiotic relationships for survival and reproductive success, and how these relationships consequently affect their biochemical capabilities.

Abstract: Metabolomics has emerged as a powerful tool to interrogate cellular biochemistry at the systems level by tracking alterations in the levels of small molecules. The success of metabolomics over the past decade has largely relied on advances in mass spectrometry instrumentation, coupled with developments in bioinformatic tools such as XCMS-Online, it has now become relatively routine to comprehensively compare the levels of metabolite peaks. To facilitate metabolite identification within the ENIGMA program we have developed a freely accessible metabolite database called METLIN (http://metlin.scripps.edu) which contains tandem mass spectral data from thousands of metabolites. This repository allow investigators to compare MS2 data from their research samples to MS2 data from model compounds catalogued in the database and thereby improve the speed, efficiency, and cost effectiveness of untargeted studies. One approach to define cellular dynamics with respect to alterations in of small molecules has been to consider metabolic flux. While flux measurements have proven effective for model organisms, acquiring multiple time points at appropriate temporal intervals for many sample types is challenging. As an alternative, meta-analysis provides another strategy for delineating metabolic cause and effect perturbations. This combination of metabolomic data from multiple genotypes or environmental conditions 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 utility 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 specific phenotypes. In particular, we present this approach to characterize strains of *Desulfovibrio alaskensis* and microbial communities. The output of these experiments include the identification of novel endogenous metabolites as well as proteins uniquely associated to a range of pathways including glycolysis, the citric acid cycle, the urea cycle, and select amino acid metabolism.

Informatics Development. Our **XCMS**/**Metlin** platform consists of a continuously evolving technology that has become the world standard for metabolomics data analysis. **Metlin (http://metlin.scripps.edu/)** the largest tandem mass spectrometry metabolite database, has had over 10 million hits containing over 45,000 structures, and MS/MS data on over 5000 metabolites. We are currently developing Metlin with the addition of ENIGMA bacterial metabolites as well as extending its functionality within **XCMS**. XCMS is the most cited metabolomics software tool with over 50,000

downloads. **XCMS ONLINE** is being developed (**https://xcmsonline.scripps.edu/**) to facilitate the analysis of microbial studies and already has over 1500 users in 36 countries. **XCMS**/**Metlin** is being designed with an easy to use command driven interface with direct connection to **Metlin** for molecule identification. XCMS ONLINE is also being developed for bacterial analysis to allow for second-order ("meta") analysis (**Figure 1**). MetaXCMS (Anal. Chem. 2011, **Nature Protocols** in press) allows for data reduction when analyzing multiple sample groups to provide information on key proteins/metabolites related to a phenotype (such as *Desulfovibrio desulfuricans* metal reduction).

Proteomics/Metabolomics Analytical Development. Our mass spectrometry based technology developments include quantitative **QqQ MRM** metabolomics methods, **LC/MS/MS** shotgun high throughput protein analysis (**Nature 2010**) as well as novel **CESI-MS** (capillary electrophoresis mass spectrometry). In particular, we are applying these methods to genetically characterized strains of *Desulfovibrio desulfuricans* including the G100 wild type and the G20 mutant. The output of these experiments include the identification of novel endogenous metabolites (agmatine, cytosine, and 7-hexadecenoic acid) and 66 proteins uniquely present in the G100 wild type organism. These approaches are being developed for ENIGMA are allowing us to examine glycolysis, the citric acid cycle, the urea cycle, and select amino acid metabolism among other pathways.

Community Flux by Pulse labeling with a Biological Event. Although very biologically informative, isotopic pulse labeling flux experiments are extremely difficult in the context of mass spectrometry. Especially when examining systems that are largely unknown, stable isotopes cause a shift in mass readout that complicates databases and limits analysis by current bioinformatic software. Therefore, as an alternative to stable isotopes, we are pulse labeling with a biological event to examine temporal changes with respect to that event. These experiments are being performed with different populations of the community where the primary purpose is to identify the key biochemical perturbations that occur as a response to that particular event.

References

Patti G.J., Yanes O., Siuzdak G. **Metabolomics: leading from behind**
Nature Reviews, in press, **2012.**

Patti G.J., Yanes O., Shriver L., Courade J-P., Tautenhahn R., Manchester M., Siuzdak G.
Metabolomics Implicates Altered Sphingolipid Metabolites
Nature Chemical Biology, in press, **2012**.

Patti G.J., Tautenhahn R., Siuzdak G.
Meta-Analysis of Untargeted Metabolomic Data from Multiple Profiling Experiments
Nature Protocols, in press, **2012.**

Patti G.J, Tautenhahn R., Fonslow B.R., Cho Y., Deutschbauer A., Arkin A., Northen T., Siuzdak G.
Meta-Analysis of Global Metabolomics and Proteomics Data to Link Alterations with Phenotype
Spectroscopy, in press (cover article), **2012.**

Tautenhahn R., Patti G., Kalisiak E., Miyamoto T., Schmidt M., Fan Y.L., McBee J., Baliga N.S., Siuzdak G.
metaXCMS: Second-Order Analysis of Untargeted Metabolomics Data
Analytical Chemistry, **2011**, 83, 696-700.

Yanes O., Tautenhahn R., Patti G., Siuzdak G.
Expanding Coverage of the Metabolome for Global Metabolite Profiling
Analytical Chemistry, **2011**, 83, 2152-2161.

Greving M.P., Patti G.J., Siuzdak G.
Nanostructure-Initiator Mass Spectrometry (NIMS) Metabolite Analysis and Imaging
Analytical Chemistry, **2011**, 83(1), 2-7.

Long J.Z., Cisar J.S., Miliken D., Niessen S., Wang C., Trauger S.A., Siuzdak G., Cravatt B.F.
Metabolomics annotates ABHD3 as a regulator of medium-chain phospholipids
Nature Chemical Biology, in press, **2011.**

This work conducted by ENIGMA (Ecosystems and Networks Integrated with Genes and Molecular Assemblies) was supported by the Office of Science, Office of Biological and Environmental Research, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.