



Characterizing Unknown Genes Through Metabolomics

Tiffany Y. Liang¹, Savannah Sanchez², Jason Rostron², Jeremy Frank², Daniel Cuevas¹, Daniel Garza³, Anca Segal², Forest Rohwer², Robert Edwards^{2,3}

¹Bioinformatics and Medical Informatics Research Center, ²Biology Department, ³Computer Science Department, San Diego State University, San Diego, CA 92182



SAN DIEGO STATE UNIVERSITY

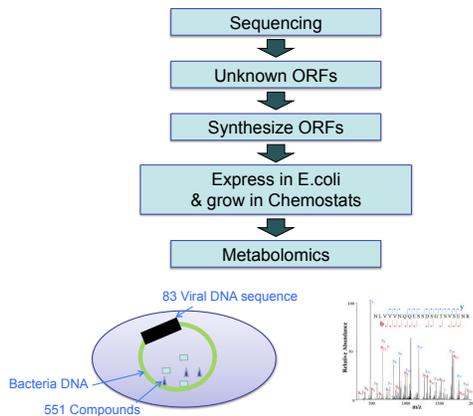
Abstract

Viruses are thought to be the most diverse biological entities on earth. However, they are also the least characterized regarding their genetic, taxonomic, and functional diversity. In metagenomic analyses of viral communities from various environments, most sequences are unrelated to any known sequences. About 90% of the viral sequences found in marine environments are unknown. The goal of this study is to characterize the function of unknown viral genes that alter host metabolism.

Open reading frames (ORFs) are predicted from the metagenomes and genes synthesized and cloned into *E. coli*. The metabolites produced by the bacterial clones are then identified by gas chromatography-coupled time-of-flight (GC/TOF) mass spectrometry. In total, 551 metabolites were found, however 84% of them were not identified based on their mass.

We are identifying the specific metabolites produced or affected by the over expression of phage proteins to suggest the physiological roles for these proteins that could be tested experimentally. We are building a systematic analysis tool that can process any kind of metabolomics data for downstream analysis of metabolomics and related data sets.

Experiment Setup



Identify Over Expressed and Under Expressed Targets

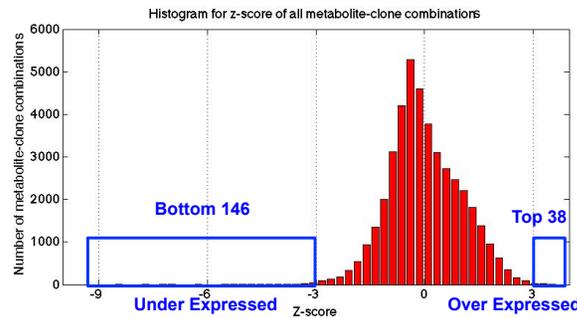


Figure1. Histogram for Z-score of each metabolite-clone combination. Data is normalized within each metabolite.

Clusters Of Over Expressed Metabolites

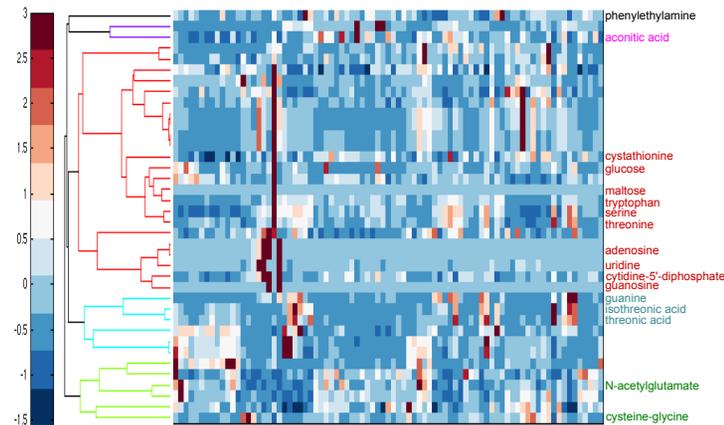


Figure2. Heatmap for metabolites in the most over expressed 38 metabolites-clone combinations. Clusters generated by hierarchical clustering with correlation distance.

Clusters Of Under Expressed Metabolites

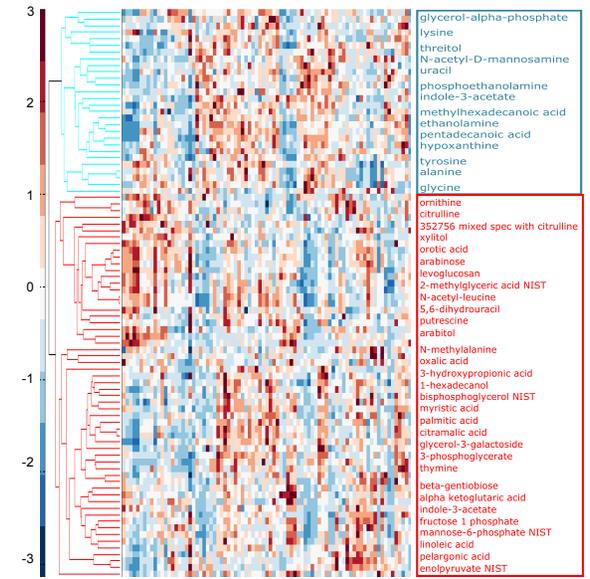


Figure3. Heatmap for metabolites in the most under expressed 146 metabolites-clone combinations. Clusters generated by hierarchical clustering with correlation distance.

Future Direction

- Based on clusters and mass spectrometry data, we will identify unknown metabolites
- For known metabolites in each cluster, we will identify metabolic networks, and further investigate unknown genes' function
- Build linear model using input and output metabolites within the system

Acknowledgements

Thanks for the help from Prof. James Nulton and Prof. Barbara Bailey.

Funding: NSF, Roche, CRI, Agouron Institute, CIFAR

Contact Information

Tiffany Y. Liang, LiangTiff@gmail.com