Bioinformatics Center
– Pathway Engineering –

http://www.bic.kyoto-u.ac.jp/pathway/

Prof
MAMITSUKA, Hiroshi
(D Sc)

Assist Prof
TAKIGAWA, Ichigaku
(D Eng)

Assist Prof
SHIGA, Motoki
(D Eng)

PD (JSPS)
NGUYEN, Hao Canh
(Ph D)

Proj Res
NATSUME, Yayoi
(D Agr)

PD
HANCOCK, Timothy Peter
(Ph D)

Student

du VERLE, David (D2)

Visitors

Assoc Prof ZHU, Shanfeng	Fudan University, China, P. R., 18–30 January
Assist Prof WICKER, Nicolas	University of Strasbourg, France, 27 February–13 March
Assoc Prof SJOLANDER, Kimmen	University of California, Berkeley, USA, 13 July
Assoc Prof ZHU, Shanfeng	Fudan University, China, P. R., 6–19 June
Prof KNAPP, Ernst-Walter	Free University of Berlin, Germany, 23 July
Prof KRAMER, Stefan	Technical University of Munich, Germany, 25–29 July
Senior Lect LAWRENCE, Neil	University of Manchester, UK, 25–29 July
Assoc Prof XING, Poe	Carnegie Mellon University, USA, 26–29 July
Assoc Prof ESKIN, Eleazar	University of California, Los Angeles, USA, 25–29 July
Dr LANDSMAN, David	National Institute of Health, USA, 25–27 July
Prof SHARAN, Roded	Tel-Aviv University, Israel, 26–29 July
Assoc Prof ZHU, Shanfeng	Fudan University, China, P. R., 24–30 July
Dr BAIROCH, Amos	Swiss Institute of Bioinformatics (SIB), Switzerland, 8 October
Dr LANE, Lydie	Swiss Institute of Bioinformatics (SIB), Switzerland, 8 October
Prof VARNEK, Alexandre	University of Strasbourg, France, 28 October
Assoc Prof LI Jiuyong	University of South Australia, Australia, 2, 9 December

Scope of Research

Mechanisms of biological functions can be shown by pathways in biological networks, such as metabolic network, signal transduction pathways and protein-protein interactions. We are interested in mining pathways related with biological conditions, such as patients/normal, by not only using biological networks but also the data of showing genes/proteins activities such as gene expression. Our approach is to develop new machine learning/data mining techniques for our interest, especially focusing on integrating different types of biological data including networks and graphs in biology.

Selected Publications

Hancock T, Takigawa I, Mamitsuka H: Mining Metabolic Pathways through Gene Expression, *Bioinformatics*, **26 (17),** 2128-2135 (2010).

Hu X, Zhou W, Udaka K, Mamitsuka H, Zhu S: MetaMHC: A Meta Approach to Predict Peptides Binding to MHC Molecules, *Nucleic Acids Research*, **38,** W474-W479 (2010).

Hancock T, Mamitsuka H: Boosted Optimization for Network Classification, *Proceedings of the 13th International Conference on Artificial Intelligence and Statistics (AISTATS 2010)* (*JMLR: Workshop and Conference Proceedings)*, **Vol. 9,** 305-312 (2010).

KEYWORDS

Data Mining
Bioinformatics
Machine Learning
Computational Genomics
Systems Biology

Mining Metabolic Pathways through Gene Expression

Metabolic networks are maps of chemical reaction pathways that are known to occur within a cell. The activity of each metabolic pathway is controlled by the activation and interaction between genetic pathways. For many organisms the reaction structure and genetic dependencies of metabolism have been identified and are stored within the Kyoto Encyclopedia for Genes and Genomes (KEGG) database. The KEGG database reveals that metabolic networks are large and highly complex. This size and complexity is sufficient to hide the key genetic pathways which define the response of the metabolic network to external stimuli.

Pathways of coordinated gene expression determine which metabolic compounds can be synthesized, and thus can define the function of metabolic networks. Additionally, microarray experiments have allowed researchers to measure gene expression under various experimental conditions. Since their inception, bioinformatics methods have tried and combine microarray expression and metabolic networks to identify the pathways that drive an observed response. In this research we have developed a framework to identify the functional metabolic pathways within gene expression.

Our approach is a combination of probabilistic models for pathway ranking, clustering and classification and is freely available as an R package: ***PathRanker***.

PathRanker first employs a non-parametric pathway extraction method to identify the most highly correlated paths through the metabolic network. Figure 1 presents an image the top 1000 pathways for yeast in *stationary phase* compared with control observations. In Figure 1 each row is a pathway and each column is a gene. The left plots show the probability and *p*-value of each path, and the bottom bar plot displays the probability of a gene over all paths. Clearly visible in Figure 1 is a block of pathways defines yeast's stationary phase.

PathRanker can then extract the defining structure within the top ranked pathways using Markov clustering and classification algorithms. The result of 3M Markov clustering is presented in Figure 2. The pathways in Figure 2 summarize 93% of yeast's stationary phase pathways. The analysis of these identified pathways can be performed at multiple resolutions including interacting genes, reactions, compounds and pathways (Figure 2). We also confirm the importance of these pathways from a biological perspective.

Finally we would like to emphasize that integrating probabilistic path ranking with clustering or classification is a reasonable combination in terms of machine learning, because path ranking generates tons of pathways which might be redundant because of the complex nature of metabolic networks while a large number of pathways can be summarized by clustering or classification.

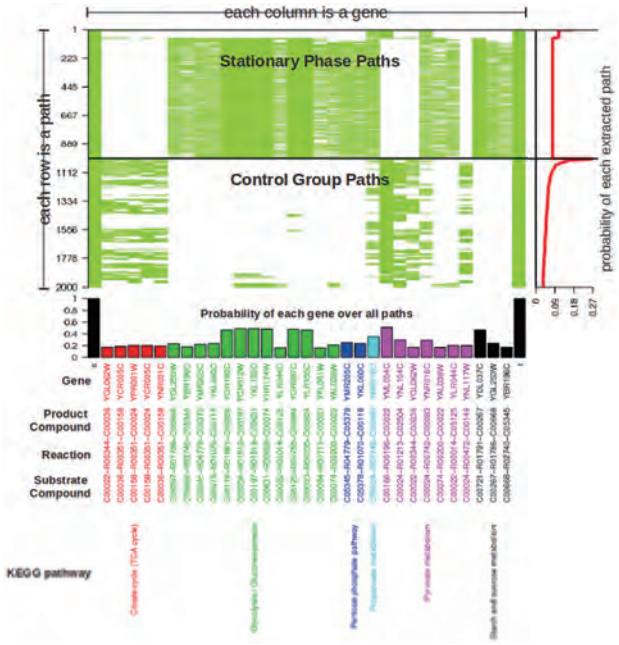

Figure 1.

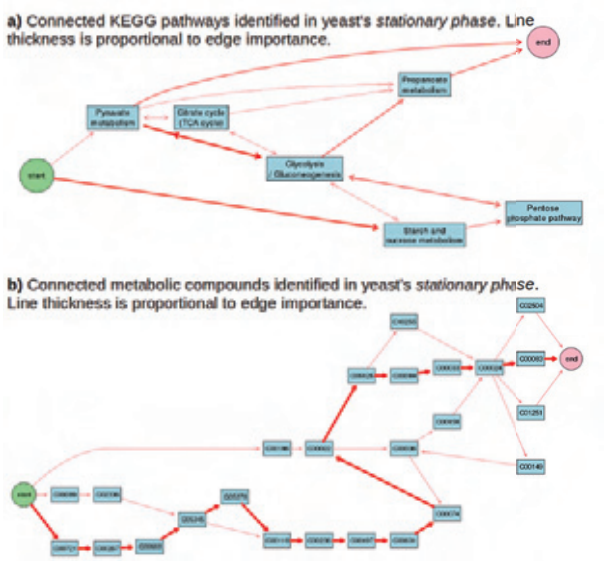

Figure 2.