Bioinformatics Center
- Pathway Engineering -

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

Prof
MAMITSUKA, Hiroshi
(D Sc)

Assist Prof
TAKIGAWA, Ichigaku
(D Eng)

PD (JSPS)
HANCOCK, Timothy Peter
(Ph D)

PD
SHIGA, Motoki
(D Eng)

PD (JSPS)
WAN, Raymond
(Ph D)

PD (JSPS)
ZHU, Shanfeng
(Ph D)

Students

YABU-UCHI, Hiroaki (D1)
MATSUSHIMA, Yoshifumi (M1)
SATO, Yoshiko (M1)

Visitors

Ms LI, Li-Min	University of Hong Kong, China, 11 June–8 September 2007
Dr MATTHIAS, E Futschik	Humboldt University, Germany, 6 August 2007
Prof KASKI, Samuel	Helsinki University of Technology, Finland, 20 September 2007
Dr HORTON, Paul	National Institute of Advanced Industrial Science and Technology, Japan, 30 October 2007

Scope of Research

With the recent advancement of experimental techniques in molecular biology, research in modern life science is shifting to the comprehensive understanding of a biological mechanism consisting of a variety of molecules. Our focus is placed on molecular mechanisms in biological phenomena, represented by biological networks such as metabolic and signal transduction pathways. Our research objective is to develop techniques based on computer science and/or statistics to systematically understand biological entities at the cellular and organism level.

Research Activities (Year 2007)

Publications

Shiga M, Takigawa I, Mamitsuka H: A Spectral Clustering Approach to Optimally Combining Numerical Vectors with a Modular Network, *Proceedings of the Thirteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2007)*, 647-656 (2007).

Shiga M, Takigawa I, Mamitsuka H: Annotating Gene Function by Combining Expression Data with a Modular Gene Network, *Bioinformatics*, **23** (13) (*Proceedings of the Fifteenth International Conference on Intelligent Systems for Molecular Biology (ISMB 2007)*) , i468-i478 (2007).

Presentations

Random Field, Network Modularity, Spectral Clustering and Beyond, Mamitsuka H, Keynote Speech, *3rd International Conference on Intelligent Computing (ICIC 2007)*, Academic Exchange Center, Ocean University of China, Qingdao, China, 22 August 2007.

An Integrative Approach for Gene Annotation Based on Spectral Clustering and Network Modularity, Mamitsuka H, Invited Talk, AASBi (Association of Asian Societies of Bioinformatics) Symposium 2007, Biopolis, Singapore, 2 December 2007.

Clustering Genes with Expressions and Network Modularity

Recent progress in genome sciences has led to the development of DNA microarray technology which allows to monitor the expression of thousands of genes simultaneously. A current popular approach to annotate gene function from gene expression data is clustering genes by expression values based on the assumption that genes with similar expression patterns can be clustered into a group with the same gene function. However, microarray expression data is inevitably noisy, making the clustering result by the above methods unstable. A promising solution for this issue in current bioinformatics research is to combine microarray expression data with the existing knowledge of gene annotation derived from literature. This type of combination is of interest, since dynamic behavior of genes which would be observed from microarray data can be integrated with the literature-derived biological data which is obviously static information. However, existing approaches are not methodologically sophisticated enough in combining the two data, i.e. real-valued expression data and literature-derived data, especially gene networks. In addition, the focus of current approaches is placed on the rather local information, such as neighboring genes, of gene networks, and incorporating global information of gene networks might find more appropriate gene clusters. In light of the above, we develop two new methods based on the idea of network modularity which allow to consider the global property of gene networks. We emphasize that our methods are general for combining two different types of data: structured (microarray expressions) and unstructured (gene networks) data, and one method is probabilistic model learning and the other is based on spectral clustering. Interested readers should refer the publications raised on the left-hand side page. We show some examples of results obtained by applying one of our methods to real data. Figure 1 shows the clustering results obtained in three cases in which we use a. only gene expressions, b. both expressions and networks, and c. only networks. This figure shows that combining two data information sources is very useful for clustering genes, since the color distribution of the middle network is the closest to the standard color distribution. Figure 2 shows the enlargement of each of the squares in Figure 1. As shown in this figure, two colors, red and green, were merged in b. while more than two colors were merged in a. and the two colors were clearly separated in c. We then checked orange colored genes in b. by using the KEGG database and found that they all correspond to those categorized in "metabolism of cofactors and vitamins", more precisely those in the folate biosynthesis pathway.

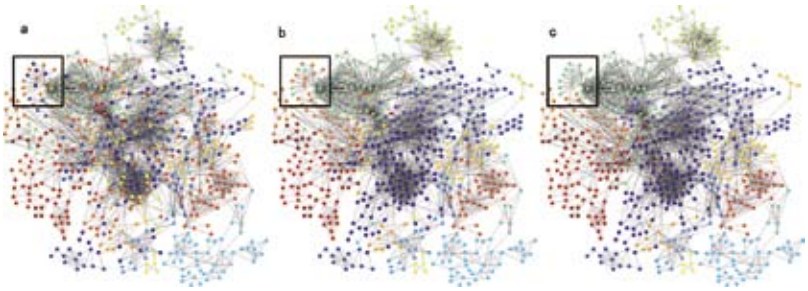

Figure 1. Clustering results where we used a. only gene expressions, b. both expressions and networks, and c. only networks.

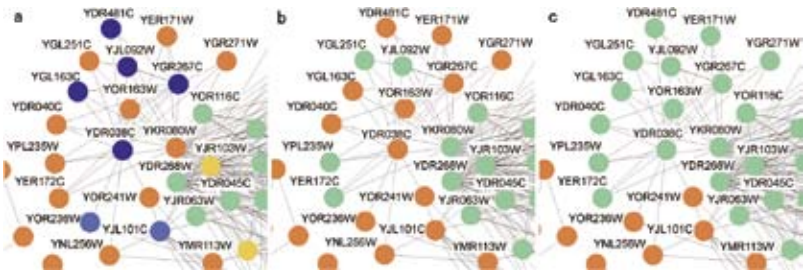

Figure 2. Enlargement of the corresponding squares in Figure 1.

Grants

Mamitsuka H, Integrative Data Mining Approaches for Unstructured Data in Life Sciences, Research Grant from BIRD (BioInformatics Research and Development) of JST (Japan Science and Technology Agency), 15 October 2007–30 September 2010.

Takigawa I, Large-Scale Biological Information Processing Based on Computational Geometric Structures and Adaptive Sampling, Grant-in-Aid for Young Scientist (B), 1 April 2006–31 March 2008.