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)

Assist Prof*
SHIGA, Motoki
(D Eng)

*Bioinformatics Center, Laboratory of Advanced Data Mining for Bioinformatics

Program-Specific Res
KAYANO, Mitsunori
(D Sc)

Program-Specific Res
NATSUME, Yayoi
(D Agr)

Program-Specific Res
WAN, Raymond
(Ph D)

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

Students

MATSUSHIMA, Yoshifumi (M2)
SATO, Yoshiko (M2)

Visitors

Prof WONG, Limsoon	National University of Singapore, Singapore, 14 January 2008
DU VERLE, David	Pierre & Marie Curie University, France, 21 May 2008/1 September 2008
Dr VO, Anh Ngoc	University of Melbourne, Australia, 22 June–20 July 2008
KOH, Chuan Hock	National University of Singapore, Singapore, 7 November 2008
Assoc Prof SJÖLANDER, Kimmen	University of California, Berkeley, USA, 10–11 November 2008

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 2008)

Publication

Hashimoto K, Takigawa I, Shiga M, Kanehisa M, Mamitsuka H: Mining Significant Tree Patterns in Carbohydrate Sugar Chains. *Bioinformatics*, **24 (16)** (*Proceedings of the Seventh European Conference on Computational Biology (ECCB 2008)*, Cagliari, Sardinia-Italy, September 2008), i167-i173, 2008.

Presentations

Data-integrative Informatics for Chemical Genomics, Mamitsuka H, Keynote Speech, *Workshop: Systems Biology –From Molecules to Life–,* La Trobe University, Melbourne, Australia, 27 May 2008.

Clustering Numerical Vectors with a Modularity Network, Mamitsuka H, Invited Talk, *Foundation of Computational Mathematics (FoCM) 2008*, City University of Hong Kong, China, 22 June 2008.

Mining Significant Patterns from Trees, Mamitsuka H, Invited Talk, *Hayama Seminar*, Sokendai, Hayama, Japan, 14 October 2008.

Efficiently Finding Significant Substructural Patterns Conserved in Glycans, Takigawa I, *Annual Conference of the*

Efficiently Mining Significant Substructural Patterns Conserved in Glycans

Glycans or carbohydrate sugar chains are one of the four fundamental macromolecular components of all cells along with nucleic acids, proteins, and lipids. Most of glycans are on the outer surface of cellular and secreted macromolecules, and modulate or mediate a wide variety of events in cell-cell and cell-matrix interactions crucial to the development and function of a complex multicellular organism. The structure of a glycan is a branched (or a linear) chain of monosaccharides attached to one another via glycosidic linkages. These "tree-shaped" sequence structures potentially lead to considerably more linkage variation in contrast to a linear sequence such as nucleic acids and proteins. However, naturally occurring ones contain relatively few of the possible monosaccharide units in a limited number of combinations. Thus, as with sequence motifs in nucleic acids or proteins, conserved structural patterns in glycans can be a key for analyzing functions of glycans. With the rapid increase in glycan structural data, mining conserved patterns from large-scale data is vital to make better understandings of glycans.

The techniques for finding all frequent subtree patterns in a given set of trees have been developed in the field of data mining. However, complete enumeration of frequent subtrees practically generates a huge number of outputs, which makes subsequent biological analysis difficult. The output patterns, hence, should be summarized as more compact patterns keeping characteristic information. For example, CMTreeMiner can retrieve closed and maximal subtree patterns. In our preliminary trials with real glycan data, the closed patterns produced still large outputs and the maximal patterns contained only too specific patterns. Moreover, very frequent patterns were often not biologically significant because such patterns were likely to be too simple and could be ubiquitous simply by chance. In order to handle these two types of problems, we first introduce a new concept, α-closed frequent subtrees, as parametric summarization of patterns controlled by α, and develop an efficient method for mining all these subtree patterns from given trees. Then we rank these obtained patterns according to not the frequencies but the *p*-values in a significance test against random control data. We experimentally verified the effectiveness of this approach using real structures of glycans retrieved from KEGG GLYCAN database. As shown in Table 1, we examined the top ranking subtrees obtained by our method and confirmed that those subtrees are significant motifs in glycobiology such as typical core structures, known extension patterns, and functionally important subtree motifs.

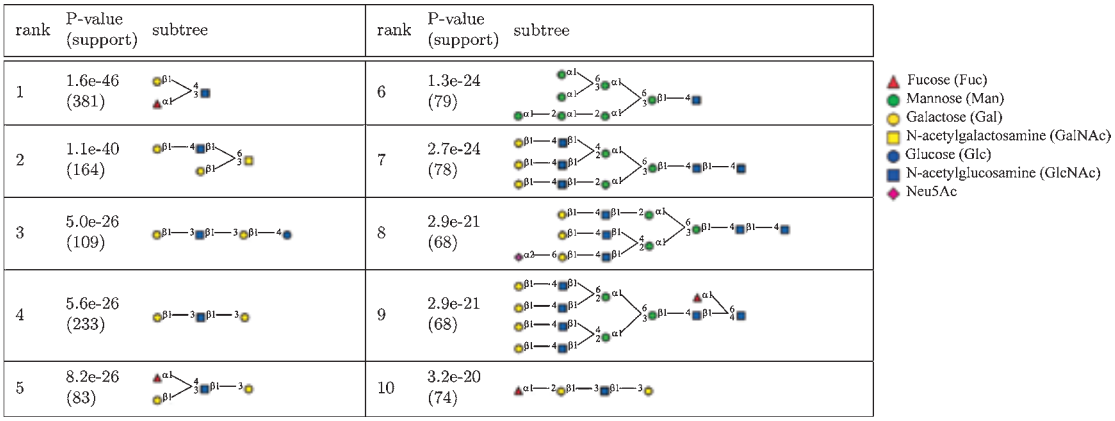

Figure 1. Top ten significant patterns in glycans.

Japanese Society for Bioinformatics (JSBi2008), Senri Life Science Center, Osaka, Japan, 15 December 2008.

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, Multifaceted Exploration of Nonhomogeneous and Ambiguous Data by Combining Partial Similarities, Grant-in-Aid for Young Scientist (B), 1 April 2008–31 March 2011.

Shiga M, Integrative Data Mining Based on Structure Analysis of Biological Networks, Grant-in-Aid for Young Scientist (B), 1 April 2008–31 March 2010.