

# Annotating Gene Functions with Integrative Spectral Clustering on Microarray Expressions and Sequences

Limin Li<sup>1</sup>, Motoki Shiga<sup>2</sup>, Wai-ki Ching<sup>1</sup>, Hiroshi Mamitsuka<sup>2</sup>

<sup>1</sup>Department of Mathematics, The University of Hong Kong

<sup>2</sup>Bioinformatics Center, Kyoto University

Annotating genes is a fundamental issue in the post-genomic era. A typical procedure for this issue is first clustering genes by their features and then assigning functions of unknown genes by using known genes in the same cluster. A lot of genomic information are available for this issue, but two major types of data which can be measured for any gene are microarray expressions and sequences, both of which however have their own flaws. Thus a natural and promising approach for gene annotation is to integrate these two data sources, especially in terms of their costs to be optimized in clustering. We develop an efficient gene annotation method with three steps containing spectral clustering over the integrated cost, based on the idea of network modularity. We rigorously examined the performance of our proposed method from three different viewpoints. All experimental results indicate the performance advantage of our method over possible clustering/classification-based approaches of gene function annotation, using expressions and/or sequences.