

Application of Discretization and Association Algorithms in Analysis of High Throughput Gene Expression Profiles for Modelling Leukemia and Neurodegenerative Disease

* Bijlani RR1, * Cheng Y1, Brooks AI2,3, Pearce D2, Federoff HJ2,3, Ogihara M1.

University of Rochester and University of Rochester Medical Center; Depts. of Computer Science¹, Neurology: Division of Molecular Medicine & Gene Therapy², Center for Aging and Developmental Biology³. Rochester, NY 14642

A new method for building classifiers of expression data is proposed. The method first explores various normalizations of gene expression data, and then searches for genes whose expression is consistently at a certain level in one class of data while it is consistently not at that level in other classes. To find such genes, two standard concepts in data mining are used, discretization and association, with a few modifications. The classifiers that are built exhibit higher accuracy in distinguishing between ALL samples and AML samples, studied in Golub et al., Science 286:531-537, 1999. In addition, we are constructing similar classifiers to help interpret the molecular mechanisms of a mouse model of neurodegenerative disease.

* Authors contributed equally to the work