Fishing Expedition - a Supervised Approach to Extract Patterns from a Compendium of Expression Profiles

Zhen Zhang[1], Grier Page[1], Hong Zhang[2]

[1]Dept. of Biometry and Epidemiology Medical University of SC; [2]BIOwulf Technologies

Reference databases of expression profiles from diverse mutations and chemical treatments of a single assay offers a bird's-eye view of changing expression patterns due to multiple perturbations. Such a compendium of expression profiles has been used to ascertain the roles of previously uncharacterized genes and infer the pathways through which their impact may take place. However, many genes have multiple molecular functions and are involved in different biological processes. The interaction patterns between genes and profiles from two-dimensional hierarchical clustering of such compendium of data could be very complex and often scattered. This makes it difficult to identify and extract all the genes and profiles whose variation in expression levels are closely associated with a particular target function. In this paper, we propose a supervised component analysis approach in which a small number of profiles and/or genes of known properties are used as "baits" to help to "fish out" other profiles and genes from a reference database that are relevant to a particular function of interest. The final cluster analysis and pattern match are then done using a much reduced data set. A data set from a study of 300 expression profiles of yeast mutant and chemical treatment is used to demonstrate the process. In particular, profiles of yeast mutants erg2D, and erg3D, and tet-ERG11 are used to reduce an originally selected data set of 136 profiles and 551 ORFs to that of 78 profiles and 200 ORFs. The reduced data set contains clusters of all important ergosterol biosynthesis related genes identifiable using the large data set. It also includes a few additional ORFs and another cluster of ORFs that are not easily identifiable using the large data set.