Using Functional Genomic Units to Corroborate User Experiments with the Rosetta Compendium

Simon M. Lin[1], Xuejun Liao[2], Patrick McConnell[1], Korkut Vata[3], Lawrence Carin[2], Pascal Goldschmidt[3]

[1]Duke Bioinformatics Shared Resource, Duke University Medical Center; [2]Department of Electrical and Computer, Duke University; [3]Department of Cardiology, Duke University Medical Center

The Rosetta data set opens the possibility of comparing a user microarray data set with a reference profile from the compendium (see [1]). However, explaining this comparison in terms of individual genes could be a daunting job because of the sheer number of genes. Thus, we postulate a new strategy of modeling microarray data in terms of functional genomic units. A functional genomic unit is a group of genes carrying out common biological function. We explored the possibility of defining the functional genomic units from the Gene Ontology (GO) annotation of the yeast genome. To visualize the tree structure of the GO, we have written a yeast genomic knowledge browser in Java, and integrated it with the microarray data. The pitfall of using the GO is that only a portion of the genes in the genome are functionally known or inferred. Thus, we further investigated an unsupervised learning method to identify those functional genomic units in the yeast genome. We have applied the established technology of digital signal processing, Independent Component Analysis (ICA), to the Rosetta data set. The functional genomic units identified by ICA helped us to corroborate our own microarray experiment with the Rosetta data set. To further demonstrate the utility of the Rosetta compendium, we have designed an experiment to investigate the yeast cells transfected with human Rac1, a small GTPase protein of the Rho family. In conclusion, we demonstrated that our designed experiment obtained from the Affymetrix platform could be corroborated by the Rosetta compendium generated by cDNA arrays.