GENE CLUSTERING AS A METHOD TO IMPROVE PREDICTION OF MISSING VALUES IN MICROARRAYS

G. Feten[1], T. Almøy[1], A.H. Aastveit[1]

[1]*Norwegian University of Life Sciences, Ås, Norway*

Email: *guri.feten@umb.no*

Microarray experiments, both in gene expression studies and in comparative genomic hybridization studies, generate data sets with multiple missing intensity values. A method for clustering the genes according to what is believed to be their optimal prediction method is introduced, based on the assumption that there exists no uniform optimal method for prediction of all genes. A comparative study of this clustering method and an adjusted version of the local least squares imputation method is performed on data from gene expression studies and comparative genomic hybridization studies as well as simulated data. This study shows that there is no optimal clustering method for every data set. Clustering based on the Pearson correlation coefficients, is preferable for data with one or several groups of correlated genes, while for data including a group of uncorrelated or weakly correlated genes the method above seems to be best.