

*Bayesian Characterizations of Natural Variation in Gene Expression*

Bhattacharjee, M., Sillanpaa, M.J., Arjas, E.

University of Helsinki, Helsinki, Finland

We present a new latent variable based Bayesian clustering method for classifying genes into categories of interest. The approach is integrated in the sense that normalization and classification can be carried out jointly along with estimation of uncertainty, with the consequence that it will be unnecessary to carry out a large number of testing of hypotheses. Possible distortion in measuring the actual expression level due to factors like environmental and experimental conditions, dye, etc. is attempted to be incorporated into the normalisation part of the model. The observed expression is treated as a black box for the different effects which are considered jointly in a nested common structure. The residual is then classified into different categories, which is of interest to us here. The approach is very general in the sense that it is easily customisable for different needs and can be modified with availability of additional knowledge/information.

A preliminary and an extended version of the model were applied to the expression data provided by Pritchard et al. (2001). The classification categories of interest here are variational categories of genes in normal circumstances. Our findings support the hypothesis that, apart from the fact that there are several sources of variation affecting the observed expression of the genes, some genes by nature exhibit highly varied expression. Several other Bayesian approaches have been presented recently for microarray data analysis (Medvedovic 2000; Keller et al. 2000; Long et al. 2001; Baldi and Long 2001; Newton et al. 2001; Ibrahim et al. 2002; Dror et al. 2002; Parmigiani et al. 2002). However, our approach differs substantially from the others.