

Linear Models for Controlling Background Signal in Microarray Gene Expression Studies

Janet Holt, Ph.D.
University of Northern Illinois

Abstract:

The general linear modeling approach to oligonucleotide microarray data subsumes various statistical models for data analysis: (a) a covariate model, in which perfect match signal is some estimated linear function of mismatch signal and other variables, (b) a perfect match-only model, in which mismatch data is not utilized, and (c) a difference model of perfect match - mismatch (Chu et al., 2002). By decomposing the correlations among the variables in the structural model and making certain assumptions, the statistical model that most accurately reflects the actual gene expression level can be theoretically derived. Various linear models are compared using analytical methods. Results show that when modeling non-systematic variation, the covariate model provides maximum flexibility and often more accurately reflects the actual gene expression levels as compared to the difference model. Further, the PM-only model can be a useful alternative in limited circumstances.