Sequential analysis for microarray data based on sensitivity and meta-analysis

Guillemette Marot[1], Florence Jaffrézic[1], Jean-Louis Foulley[1], and Claus-Dieter Mayer[2]

[1] INRA, UR 337 Station de Génétique Quantitative et Appliquée, France
[2] Biomathematics and Statistics Scotland, UK

Sequential methods and adaptive designs have a long standing tradition in clinical trials in order to reduce sample size while keeping a reasonable statistical power. These methods are characterized by interim analyses at pre-defined stages and a stopping rule which determines at each stage whether to stop or continue sampling. The high cost of microarray experiments makes such sample size reduction very desirable and as technical limits (defined by the size of the hybridization chamber for example) often make the data available in a staggered fashion anyway, we propose a sequential approach for such experiments. This approach enables us to readdress the question of sample size after each stage, which is useful when there is no or little prior knowledge available, which would allow an accurate power calculation.
One interesting feature of the sequential approach for microarrays is that, in contrast to a univariate situation, the large number of variables (genes) tested simultaneously prevents the interim analysis from introducing a serious bias to the final p-values. Thus, results from different stages can be combined by meta-analysis methods and error rates (e.g. the false discovery rate) can be controlled by applying standard procedures (e.g. the Benjamini-Hochberg rule) to the p-values from the combined stages. We study two different meta-analysis approaches:
a) combining p-values from different stages by the inverse normal method
b) combining effect sizes calculated from moderated t-tests at each stage.
We suggest stopping rules based on either the estimated number of true positives or the estimated sensitivity, i.e. the expected proportion of true positives among the truly differentially expressed genes. The experiment is stopped if this sensitivity estimate exceeds a given threshold of, for example, 70%. We compare several mixture models that can be used for sensitivity estimation. Our sequential method was tested in an extensive simulation study and also applied to several real data sets. Our results show that the application of sequential methods is able to reduce sample-sizes and thus costs in microarray experiments. An R package implementing the method is available upon request from the first author.