

**Smyth, Gordon K.**

**Linear models and empirical Bayes methods for assessing differential expression in microarray experiments.** (English) [Zbl 1038.62110](#)  
Stat. Appl. Genet. Mol. Biol. 3, No. 1, Article 3 (2004).

Summary: The problem of identifying differentially expressed genes in designed microarray experiments is considered. *I. Lönnstedt* and *T. Speed* [Stat. Sin. 12, 31–46 (2002; [Zbl 1004.62086](#))] derived an expression for the posterior odds of differential expressions in a replicated two-color experiment using a simple hierarchical parametric model. The purpose of this paper is to develop the hierarchical model of Lönnstedt and Speed into a practical approach for general microarray experiments with arbitrary numbers of treatments and RNA samples. The model is reset in the context of general linear models with arbitrary coefficients and contrasts of interest. The approach applies equally well to both single channel and two color microarray experiments. Consistent, closed form estimators are derived for the hyperparameters in the model. The estimators proposed have robust behavior even for small numbers of arrays and allow for incomplete data arising from spot filtering or spot quality weights.

The posterior odds statistic is reformulated in terms of a moderated  $t$ -statistic in which posterior residual standard deviations are used in place of ordinary standard deviations. The empirical Bayes approach is equivalent to shrinkage of the estimated sample variances towards a pooled estimate, resulting in far more stable inference when the number of arrays is small. The use of moderated  $t$ -statistics has the advantage over the posterior odds that the number of hyperparameters which need to be estimated is reduced; in particular, knowledge of the non-null prior for the fold changes are not required. The moderated  $t$ -statistic is shown to follow a  $t$ -distribution with augmented degrees of freedom. The moderated  $t$  inferential approach extends to accommodate tests of composite null hypotheses through the use of moderated  $F$ -statistics. The performance of the methods is demonstrated in a simulation study. Results are presented for two publicly available data sets.

**MSC:**

- [62P10](#) Applications of statistics to biology and medical sciences; meta analysis
- [92C40](#) Biochemistry, molecular biology
- [62C12](#) Empirical decision procedures; empirical Bayes procedures

Cited in <b>1</b> Review Cited in <b>79</b> Documents
--

**Full Text:** [DOI Link](#)