Robust hyperparameter estimation protects against hypervariable genes and improves power to detect differential expression. (English)  

Summary: One of the most common analysis tasks in genomic research is to identify genes that are differentially expressed (DE) between experimental conditions. Empirical Bayes (EB) statistical tests using moderated genewise variances have been very effective for this purpose, especially when the number of biological replicate samples is small. The EB procedures can, however, be heavily influenced by a small number of genes with very large or very small variances. This article improves the differential expression tests by robustifying the hyperparameter estimation procedure. The robust procedure has the effect of decreasing the informativeness of the prior distribution for outlier genes while increasing its informativeness for other genes. This effect has the double benefit of reducing the chance that hypervariable genes will be spuriously identified as DE while increasing statistical power for the main body of genes. The robust EB algorithm is fast and numerically stable. The procedure allows exact small-sample null distributions for the test statistics and reduces exactly to the original EB procedure when no outlier genes are present. Simulations show that the robustified tests have similar performance to the original tests in the absence of outlier genes but have greater power and robustness when outliers are present. The article includes case studies for which the robust method correctly identifies and downweights genes associated with hidden covariates and detects more genes likely to be scientifically relevant to the experimental conditions. The new procedure is implemented in the limma software package freely available from the Bioconductor repository.

MSC:
62P10 Applications of statistics to biology and medical sciences; meta analysis  
62C12 Empirical decision procedures; empirical Bayes procedures  
62F35 Robustness and adaptive procedures (parametric inference)  
92D20 Protein sequences, DNA sequences

Keywords:  
empirical Bayes; outliers; robustness; gene expression; microarrays

Software:
limma; edgeR; Rsusbread; Bioconductor; diffHic; Voom; csaw; BRENTE

Full Text: DOI Euclid

References:


Edited by FIZ Karlsruhe, the European Mathematical Society and the Heidelberg Academy of Sciences and Humanities © 2021 FIZ Karlsruhe GmbH


This reference list is based on information provided by the publisher or from digital mathematics libraries. Its items are heuristically matched to zbMATH identifiers and may contain data conversion errors. It attempts to reflect the references listed in the original paper as accurately as possible without claiming the completeness or perfect precision of the matching.