BIOINFORMATICS APPLICATIONS NOTE Vol. 21 no. 12 2005, pages 2921–2922
doi:10.1093/bioinformatics/bti436

Gene expression

twilight; a Bioconductor package for estimating the local false discovery rate

Stefanie Scheid* and Rainer Spang

Max Planck Institute for Molecular Genetics, Department of Computational Molecular Biology, Ihnestrasse 63-73, D-14195 Berlin, Germany

Received on January 31, 2005; revised and accepted on April 5, 2005
Advance Access publication April 7, 2005

ABSTRACT

Summary: twilight is a Bioconductor compatible package for analysing the statistical significance of differentially expressed genes. It is based on the concept of the local false discovery rate (FDR), a generalization of the frequently used global FDR. twilight implements the heuristic search algorithm for estimating the local FDR introduced in our earlier work. In addition to the raw significance measures, it produces diagnostic plots, which provide insight into the extent of differential expression across genes.

Availability: http://www.bioconductor.org

Contact: stefanie.scheid@molgen.mpg.de

Supplementary information: Please visit our software webpage on http://compdiag.molgen.mpg.de/software

INTRODUCTION

The false discovery rate (FDR) as introduced by Benjamini and Hochberg (1995) is a widely used error measure for multiple testing issues. In the context of differential gene expression, the FDR is defined as the expected proportion of genes falsely called differentially expressed among all genes called differentially expressed. There exist several approaches to control or estimate the FDR [for an overview see Reiner *et al*. (2003)]. A shortcoming of the FDR is that it does not refer to single genes but to a list of genes. Efron *et al*. (2001) introduced the local FDR, an analogous measure of uncertainty refering to single genes. It is defined as the probability that a gene is truly not differentially expressed given an observed test statistic or P-value.

In addition to its gene-by-gene interpretation, the local FDR provides an overview over the whole experiment. For ease of interpretation, we plot P-values versus one minus the local FDR (Fig. 1). The plot describes the course of gene expression from clear induction to clear non-induction. In between, a twilight zone spreads out where it is impossible to distinguish between induction and non-induction. We understand induction as the effect on gene expression that is caused by molecular differences between the examined conditions.

In our earlier work (Scheid and Spang, 2004), we proposed a penalized stochastic search algorithm to estimate the local FDR. In a nutshell, the algorithm works as follows: starting with a set of observed P-values, we successively remove P-values until the set of remaining P-values follows a uniform distribution. The set represents genes that are not differentially expressed. Given its uniform P-value density $\widehat{f_0}$, the percentage $\widehat{\pi_0}$ of P-values in the uniform part and the observed overall density $\widehat{f}$, the local FDR is estimated as $\widehat{\mathrm{fdr}}(p) = \widehat{\pi_0}\widehat{f_0}(p)/\widehat{f}(p)$ for each P-value p. We showed in simulations that our method estimates the local FDR accurately, and compares well with the previous methods. It outperforms its competitors when estimating the overall percentage π_0 of non-induced genes.

The procedure relies on the assumptions that gene-expression levels are independent of each other and P-values follow a uniform distribution under no differential expression. To our knowledge, the assumption of independence is common to all local FDR methods. We do not need any further assumptions. Our method, in particular, is not based on any distributional model on the mixture density f or its components, different from the works of, for instance, Efron (2004) and Liao *et al*. (2004).

IMPLEMENTATION

The algorithm is implemented in the R package twilight [R Development Core Team (2004)]. Time-consuming calculations are, however, written in C. The package is available from the Bioconductor project, a collection of R packages for genomic data (Gentleman *et al*., 2004). Package twilight contains a manual describing technical aspects in greater detail. We provide standard statistical tests on the difference of means for two-sample designs as well as correlation tests. The currently available version of twilight has changed and offers more tests than before. However, for estimating the local FDR, the main function *twilight* only needs a set of P-values as input. These P-values can be derived from any appropriate test. The local FDR estimation is not limited to gene-expression data but applies to a wide range of statistical hypothesis testing.

For illustration, we apply function *twilight* to the dataset of Golub *et al*. (1999). It comprises expression data from 72 Affymetrix HU6800 microarrays. After normalization, we compute P-values for a two-sample t-test on 47 acute lymphoblastic leukemia samples versus 25 acute myeloid leukemia samples. Function *twilight* invokes the local FDR estimation on the set of P-values. For each gene, an estimated value of the local FDR is returned. The estimator's variability is assessed on 100 bootstrap samples of the input P-values. Bootstrap means and bootstrap confidence intervals are returned.

Figure 1 displays the bootstrap mean of the estimated local FDR as a function of the P-values. The dashed lines denote the lower and upper bounds of the 95% bootstrap confidence interval. One observes how the local FDR varies along the range of P-values. We

*To whom correspondence should be addressed.

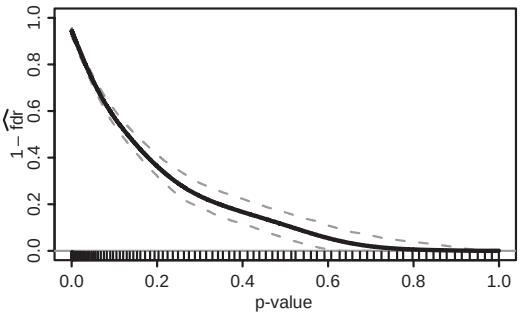

Fig. 1. Graphical output of package twilight with 100 bootstraps on biological data: P-values versus bootstrap mean of 1−estimated local FDR with 95% bootstrap confidence interval. Bottom ticks denote 1% quantiles of P-values.

follow its course from clear differential expression at the left side of the plot, starting with $1 - \widehat{\text{fdr}} = 0.95$, to clear non-induction on the right side where $1 - \widehat{\text{fdr}} = 0$. Between these bounds, we observe a broad twilight zone where the local FDR decreases rather slowly. For example, genes with P-values up to 0.12 have a probability $>50\%$ of being differentially expressed. We conclude from Figure 1 that the comparison of the two distinct leukemiae exhibits a large amount of differential expression. Based on the plot, genes with local FDR lower than a certain threshold can be chosen for further examination.

RUNTIME COMPARISON

We compare twilight with two local FDR estimators implemented in R, i.e. package locfdr and function localFDR. Package locfdr is based on methods in Efron (2004). For a set of input test statistics such as differences in means, the author assumes that the statistic's null distribution f_0 is normal. Location and scale parameters are estimated from the observed values. Function localFDR fits the piece-wise mixture model of Liao *et al*. (2004) to a set of P-values. The authors assume that the mixture distribution decomposes into a uniform distribution f_0 and a beta distribution f_1.

We examine CPU times on a Linux machine with 0.5 Gb memory and AMD Athlon XP 2400 + processor. The results are summarized in Table 1. locfdr is restricted in its applicability due to its distributional assumptions. Since it does not use permutations at all, it clearly outperforms both localFDR and twilight. Among the two permutation based programs twilight is the faster one. Bootstrap estimates of the local FDR are computationally expensive. Parallel computation on a Linux cluster is possible. Bootstraps are distributed on the cluster by using the functionality of package *snow* available on http://www.r-project.org. The CPU times for twilight with 100 bootstrap samples on the single machine and on a cluster of 20 comparable machines are shown in Table 1. With the cluster, the computation lasts 69 s and is faster than twilight without bootstrapping on a single machine (102 s).

Table 1. Runtime comparison on biological data using twilight, locfdr and localFDR with default values. In addition, twilight was run with $B = 100$ bootstrap samples on a single machine and on a cluster of 20 machines.

Application	CPU time (s)
locfdr	0.09
twilight	102.35
localFDR	943.94
twilight, $B = 100$	757.56
twilight, $B = 100$, 20 processors	68.72

ACKNOWLEDGEMENTS

This work was done within the context of the Berlin Center for Genome Based Bioinformatics (BCB), part of the German National Genome Network (NGFN), and supported by BMBF grants 031U109/209 and 01GR0455.

REFERENCES

Benjamini,Y. and Hochberg,Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B*, **57**, 289–300.

Efron,B. (2004) Large-scale simultaneous hypothesis testing: the choice of a null hypothesis. *J. Am. Stat. Soc.*, **99**, 96–104.

Efron,B. *et al*. (2001) Empirical Bayes analysis of a microarray experiment. *J. Am. Stat. Assoc.*, **96**, 1151–1160.

Gentleman,R. *et al*. (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.*, **5**, R80.

Golub,T.R. *et al*. (1999) Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science*, **286**, 531–537.

Liao,J.G. *et al*. (2004) A mixture model for estimating the local false discovery rate in DNA microarray analysis. *Bioinformatics*, **20**, 2694–2701. [http://www.geocities.com/jg_liao/software/]

R Development Core Team (2004) *R: A Language and Environment for Statistical Computing*. Manual of the R Foundation for Statistical Computing, Vienna, Austria.

Reiner,A. *et al*. (2003) Identifying differentially expressed genes using false discovery rate controlling procedures. *Bioinformatics*, **19**, 368–375.

Scheid,S. and Spang,R. (2004) A stochastic downhill search algorithm for estimating the local false discovery rate. *IEEE/ACM Trans. Comp. Biol. Bioinf.*, **1**, 98–108.