LMMNorm: a package for the normalization of microarrays using linear mixed models

Philippe Haldermans[1,*] **, Ziv Shkedy**[1]

1. Interuniversity Institute for Biostatistics and statistical Bioinformatics (I-BioStat), Center for Statistics, Hasselt University, Agoralaan 1,B-3590 Diepenbeek,Belgium

* Contact author: philippe.haldermans@uhasselt.be

Keywords: microarrays, normalization, linear mixed model, graphical user interface

During the last years a large expansion in the research on microarrays has been observed. Many techniques for the normalization of these microarrays have been proposed and implemented, for instance nonlinear normalization (Park *et al.* (2003)) and print-tip dependent normalization (Dudoit *et al.* (2002)) . Several of these normalization methods can be written in terms of a linear mixed model as described in Haldermans *et al.* (2007) . This enables us to use an objective selection criteria, such as Akaike Information Criterion (AIC), to determine which is the best normalization method for a given array.

The `LMMNorm` package normalizes cDNA microarray data using linear mixed models (LMM) as scatterplot smoothers of the MA-plot. This can be done since the normalization models, i.e. global, linear and nonlinear can be formulated as LMM:

$$\mathbf{M}(\mathbf{A}) = \begin{cases} \beta_{\mathbf{0}} & \text{global,} \\ \beta_{\mathbf{0}} + \beta_{\mathbf{1}}\mathbf{A} & \text{linear,} \\ \beta_{\mathbf{0}} + \beta_{\mathbf{1}}\mathbf{A} + \mathbf{Zb} & \text{nonlinear.} \end{cases} \tag{1}$$

Other normalization models, such as print-tip specific normalization (Dudoit *et al.*, 2002) or normalization with non-constant variance models can be expressed as a LMM as well. These models were implemented using the R function `lme()`. The `LMMNorm` package contains functions that automatically create the requested models, for instance the construction of the design matrix $\mathbf{Z}$ in the nonlinear model for the random effects, and compares the models using a selection criteria like AIC to produce the most appropriate model.

To facilitate the use of these functions for unexperienced users, a graphical user interface was developed. This offers the users a point and click environment, which allows them to use the statistical methodology without the need for a thorough knowledge of the model framework. It provides plots before and after normalization making it easy to see the effect of the normalization method. The GUI makes extensive use of the interface implemented in the `gWidgets` package.

References

[1] Sandrine Dudoit, Yee Hwa Yang, Matthew J. Callow, and Terence P. Speed, *Statistical methods for identifying differentially expressed genes in replicated cdna microarray experiments*, Statistica Sinica **12** (2002), no. 1, 111–139.

[2] Philippe Haldermans, Ziv Shkedy, Suzy Van Sanden, Tomasz Burzykowski, and Aerts Marc, *Using linear mixed models for normalization of cdna microarrays*, Statistical Applications in Genetics and Molecular Biology **6** (2007), Article 19.

[3] Taesung Park, Sung-Gon Yi, Sung-Hyun Kang, SeungYeoun Lee, Yong-Sung Lee, and Richard Simon, *Evaluation of normalization methods for microarray data*, BMC Bioinformatics **4** (2003), 33.