BIOINFORMATICS APPLICATIONS NOTE
Vol. 1 no. 1 2004
Pages 1–2

limmaGUI: A graphical user interface for linear modeling of microarray data

James M. Wettenhall and Gordon K. Smyth

Division of Genetics and Bioinformatics, Walter and Eliza Hall Institute of Medical Research, 1G Royal Pde, Parkville, 3050, Australia

ABSTRACT

Summary: *limmaGUI* is a Graphical User Interface (GUI) based on R-Tcl/Tk for the exploration and linear modeling of data from two-color spotted microarray experiments, especially the assessment of differential expression in complex experiments. *limmaGUI* provides an interface to the statistical methods of the *limma* package for R, and is itself implemented as an R package. The software provides point and click access to a range of methods for background correction, graphical display, normalization, and analysis of microarray data. Arbitrarily complex microarray experiments involving multiple RNA sources can be accomodated using linear models and contrasts. Empirical Bayes shrinkage of the gene-wise residual variances is provided to ensure stable results even when the number of arrays is small. Integrated support is provided for quantitative spot quality weights, control spots, within-array replicate spots and multiple testing. *limmaGUI* is available for most platforms on the which R runs including Windows, Mac and most flavours of Unix.

Availability: http://bioinf.wehi.edu.au/limmaGUI

Contact: wettenhall@wehi.edu.au

BACKGROUND

While a wealth of advanced software for statistical analysis of microarrays is available in R packages on *Bioconductor* (http://www.bioconductor.org), these packages have sophisticated command-driven interfaces tuned to users from mathematical or computing backgrounds. Most biologists instead use commercial software packages, incorporating possibly less cutting edge statistical methods, because they are easier to use. One of the most commonly used Bioconductor packages is the *limma* (Linear Models for Microarray Analysis) package which provides data analysis and normalization for cDNA microarray data and analysis of differential expression for multi-factor designed experiments. The core of the *limma* package is an implementation of the Empirical Bayes linear modeling approach of Smyth (2004). *LimmaGUI* provides a point and click interface to the core functionality of the *limma* package. The GUI makes extensive use of the R-Tcl/Tk interface implemented in the *tcltk* package (Dalgaard, 2001).

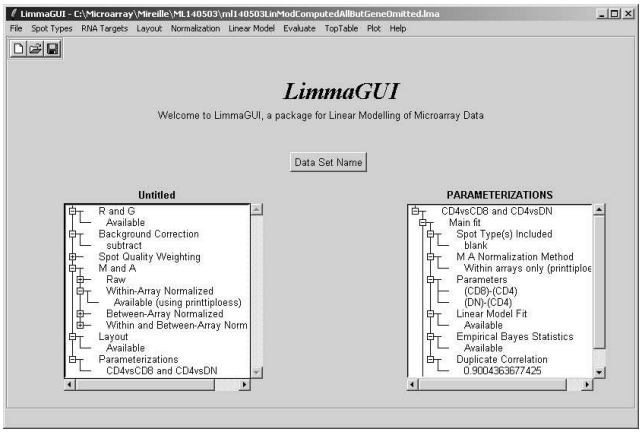

Fig. 1. The main window of *limmaGUI*.

A particular area of need that *limmaGUI* has addressed is to facilitate the construction of appropriate design matrices for linear modeling and appropriate contrast matrices to extract comparisons of interest from those linear models. While the linear modeling approach allows for testing of differential expression in very complex experimental designs, users who are not professional statisticians may be discouraged by the difficulty of specifying appropriate design and contrast matrices using a command-line interface. *limmaGUI* automates the construction of these matrices, asking users only to specify the RNA sources they would like to compare.

DESCRIPTION

The analysis session is controlled through a main window (Figure 1). At each stage of the analysis, available options are indicated through drop-down menu items at the top of the window. A graphical tree display is used to record the data structures which have been created so far during the session and which are available for further analysis.

LimmaGUI reads text files containing raw intensity data exported by a variety of image analysis programs includ-

SlideNumber	FileName	Cy3	Cy5
81	swirl.1.spot	swirl	wild type
82	swirl.2.spot	wild type	swirl
93	swirl.3.spot	swirl	wild type
94	swirl.4.spot	wild type	swirl

Table 1. The RNA Targets input file listing the hybridizations.

ing GenePix, SPOT, ImaGene, QuantArray and ArrayVision. Normally a gene list file should also be provided in tab-delimited text format, for example a GenePix Array List, although in some cases this information can be read from the image analysis files. It should be noted that *LimmaGUI* can only accommodate one array layout for each data set. *LimmaGUI* also expects an RNA Targets file (Table 1) describing the RNA sources hybridized to the arrays and the names of the files containing the intensity data, and a Spot Types file which provides information to identify control spots such as blanks or spike-in controls using a simplified regular expression notation.

The key steps in an analysis are: (1) read the image analysis files listed in the RNA Targets file; (2) perform diagnostic plots; (3) if desired, compare the effects of different normalization options; (4) normalize the data; (5) specify the comparisons of interest (e.g., mutant vs wild type), enabling *limmaGUI* to fit a linear model and to estimate some empirical Bayes statistics; (6) view the differentially expressed genes and confidence scores; and (7) if desired, display summary results using Venn diagrams.

Diagnostic plots include MA-plots and spatial image plots. A selection of standard and non-standard background correction and normalization options are provided, including those described by Smyth and Speed (2003). Within-array normalization methods included print-tip loess and composite loess normalization. Multi-array or between-array methods include simple scale normalization and quantile normalization. Choices are indicated by radio buttons in the respective dialog boxes.

To create a design matrix for the Swirl Zebrafish experiment described in Table 1, the user only needs to select the RNA sources they want to compare from a pair of drop-down combo boxes. The design matrix describes the linear relationship assumed between the observed red/green log-ratios and the average log fold change(s) to be estimated. In this case, the design matrix automatically accommodates the two dye swaps used in the experiment. Table 2 gives the top ten selected genes for this experiment with columns for the log2-fold change (M), moderated-t, adjusted p-value and B-statistic. The

ID	Name	M	t	P-Value	B
control	BMP2	−2.2	−21	0.00087	8
control	BMP2	−2.3	−20	0.0011	7.8
control	Dlx3	−2.2	−20	0.0013	7.7
control	Dlx3	−2.2	−20	0.0014	7.6
fb94h06	20−L12	1.3	14	0.015	5.8
fb40h07	7−D14	1.3	14	0.019	5.5
fc22a09	27−E17	1.3	13	0.021	5.5
fb85f09	18−G18	1.3	13	0.021	5.5
fc10h09	24−H18	1.2	13	0.023	5.4
fb85a01	18−E1	−1.3	−13	0.025	5.3

Table 2. A top table of genes ranked n order of evidence for differential expression. The BMP2 gene which causes the Swirl mutation in the zebrafish is correctly ranked at the top of the table. The negative M-value shows it is down-regulated in the Swirl mutant.

B-statistic is an estimate of the log-odds of differential expression for that gene. The table of genes can be viewed in a table widget in *limmaGUI* or exported to an external spreadsheet program.

The *limmaGUI* menus are intended to provide capabilities suitable for most users, but advanced users can use a command-line interface within *limmaGUI* to issue commands directly to the R interpreter. New plots or data structures created in this way can be added to the pull-down menus for easy access in the future.

RELATED WORK

A sister package, *affylmGUI*, has been developed to provide linear modeling and data analysis for data from single channel arrays, including high-density oligonucleotide arrays such as Affymetrix.

ACKNOWLEDGEMENTS

This work was funded by an NHMRC transitional institute grant.

REFERENCES

Dalgaard, P. (2001). The R-Tcl/Tk interface. In K. Hornik and F. Leisch (Eds.), *Proceedings of the 2nd International Workshop on Distributed Statistical Computing*, Vienna, Austria.

Smyth, G. K. (2004). Linear models and empirical bayes for assessing differential expression in microarray experiments. *Statistical Applications in Genetics and Molecular Biology 3*(1), Article 1.

Smyth, G. K. and T. Speed (2003). Normalization of cDNA microarray data. *Methods 31*(4), 265–73.