Gene expression

**oneChannelGUI: a graphical interface to Bioconductor tools, designed for life scientists who are not familiar with R language**

Remo Sanges\(^1\), Francesca Cordero\(^2\) and Raffaele A. Calogero\(^3\)*

\(^1\)CBM (Cluster in Biomedicine) S.c.r.l., AREA Science Park, Basovizza SS 14, km 163,5 34012 Trieste, \(^2\)Department of Informatics, University of Torino, Corso Svizzera 185, 10149 Torino and \(^3\)Department of Clinical and Biological Sciences, University of Torino, Regione Gonzole 10, 10043 Orbassano, Torino, Italy

Received on May 26, 2007; revised on August 27, 2007; accepted on September 10, 2007

Advance Access publication September 17, 2007

Associate Editor: David Rocke

**ABSTRACT**

Summary: OneChannelGUI is an add-on Bioconductor package providing a new set of functions extending the capability of the affylmGUI package. This library provides a graphical interface (GUI) for Bioconductor libraries to be used for quality control, normalization, filtering, statistical validation and data mining for single channel microarrays. Affymetrix 3’ expression (IVT) arrays as well as the new whole transcript expression arrays, i.e. gene/exon 1.0 ST, are actually implemented. oneChannelGUI is available for most platforms on which R runs, i.e. Windows and Unix-like machines.

**Availability:** http://www.bioconductor.org/packages/2.0/bioc/html/oneChannelGUI.html

**Contact:** raffaele.calogero@unito.it

**Supplementary information:** http://www.bioinformatica.unito.it/oneChannelGUI/

1 INTRODUCTION

The Bioconductor project (Gentleman et al., 2004) is an enormous repository of academic software for the analysis of genomic data and is centered around the analysis of microarray data. The software packages, available in Bioconductor, have sophisticated command-line driven interfaces tuned to users from mathematical or computing backgrounds. The command-line approach is very powerful since it gives complete control of the statistical tool in use. However, learning its use can be very frustrating for users without consolidated programming experience. oneChannelGUI is designed specifically for life scientists who are not familiar with R language but do wish to capitalize on the vast analysis opportunities of Bioconductor.

oneChannelGUI offers a comprehensive microarray analysis for Affymetrix (www.affymetrix.com) Affymetrix 3’ (IVT) expression arrays as well as for the new generation of whole transcript arrays: human/mouse/rat exon 1.0 ST and human gene 1.0 ST arrays. It extends the capability of affylmGUI (Wettenhall et al., 2006). affylmGUI was designed as an interface to the affy, germa (Wu et al., 2004), affyPLM and limma (Smyth, 2004) package. oneChannelGUI inherits the core affylmGUI functionalities and permits a wider range of analysis allowing biologists to choose among different criteria and algorithms in order to analyze their data. The actual design of the package adds the following functionalities to the core affylmGUI:

- **Data set loading**
  - Affymetrix exon 1.0 ST gene/exon level expression data can be loaded as well as single channel array expression data derived from GEO (www.ncbi.nih.gov/geo) or from tab delimited files containing only expression data.

- **Quality controls**
  - The visualization of source of variation present in a microarray study can be visualized by principal component analysis (PCA) and hierarchical clustering.

- **Study design**
  - New functionalities are added to the core affylmGUI to investigate the statistical quality of a microarray study.

- **Probe sets summary and normalization**
  - Expresso function (affy package) is implemented.
  - A module that allows quantile/loess/qspline normalizations is available in case no normalized data are loaded as tab delimited files.

- **Filtering**
  - Modules for filtering microarray data set on the basis of signal and annotation features are added to the core affylmGUI.

- **Statistical analysis**
  - SAM (Tusher et al., 2001) and rank product (Breitling et al., 2004) algorithms for two-class unpaired samples are implemented, as well as an interface to maSigPro time-course analysis package (Conesa et al., 2006).

- **Classification**
  - Basic classification analyses can be performed using the nearest shrunked centroids (Tibshirani et al., 2002) and the penalized discriminant analysis (Hastie, 1994) algorithms.

---

*To whom correspondence should be addressed.
*The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.
• Biological interpretation
  • A graphical interface to GOstats (Falcon and Gentleman, 2007) was added to the core affylmGUI functionalities.
  • IPA template files (www.ingenuity.com) can be generated starting from tables of differentially expressed probe sets.
• Exon analysis
  • Modules for basic alternative splicing detection and filtering are implemented.

2 SOFTWARE OVERVIEW

2.1 Data loading
oneChannelGUI inherits all affylmGUI modules and functions to analyze Affymetrix IVT arrays. In addition expression flat data files from GEO (i.e. Series Matrix Files) can be loaded as well as tab delimited files containing expression data only, e.g. derived from ArrayExpress processed flat files.

The new Exon 1.0 ST arrays can be also loaded and manipulated in oneChannelGUI. It is important to point out that, due to high memory requirements for the generation of summarized expression values, the loading of even a few exon 1.0 ST .CEL files is unfeasible on 32 bit Windows systems. However, oneChannelGUI offers also the possibility to load Affymetrix exon 1.0 ST expression data using the gene/exon level expression data exported, as tab delimited files, from the Affymetrix Expression Console (www.affymetrix.com/support/technical/software_downloads.affx). Gene/exon level expression data can also be calculated from .CEL files using a oneChannelGUI module that run and upload results produced by Affymetrix APT tools (www.affymetrix.com/support/developer/powertools/index.affx), which embeds a set of cross-platform command line programs generating the summarized expression values for individual or collections of Affymetrix arrays. DABG [detection above background; (Affymetrix-Technical-Note)] P-values calculation is also implemented.

2.2 Quality controls
Probe level QC can be done when .CEL files are loaded, using the affylmGUI QC functions. In addition sample clustering, to evaluate the homogeneity of the experimental groups, can be done by PCA and hierarchical clustering at probe set level.

2.3 Study design
Graphical implementations of the size and sizepower Bioconductor libraries allow the user to determine how many samples are needed to achieve a specified power for a test of whether a gene is differentially expressed or, in reverse, to determine the power of a given sample size.

2.4 Probe sets summary and normalization
The expresso function, which allows the integration of different methods for background correction, normalization, probe specific correction and summary value computation, was added to the affylmGUI inherited probe set summary methods (Irizarry et al., 2003; Wu et al., 2004).

2.5 Filtering
A central problem in microarray data analysis is the high dimensionality of gene expression space, which prohibits a comprehensive statistical analysis without focusing on particular aspects of the joint distribution of the gene expression levels. Possible strategies are to perform data-driven non-specific filtering of genes (von Heydebreck et al., 2004) before the actual statistical analysis or to filter, making use of biologically relevant a priori knowledge. For this purpose IQR/Intensity-based filters (von Heydebreck et al., 2004) as well as filtering based on probe set/Entrez gene annotations methods are implemented. For exon arrays, a filter based on DABG P-values is also available.

2.6 Statistical analysis
SAM (Tusher et al., 2001) and rank product (Breitling et al., 2004) algorithms for two-class unpaired samples are implemented in addition to affylmGUI modules for linear modeling analysis, which are fully inherited. An interface to maSigPro package (Conesa et al., 2006) is also available to allow time-course analyses.

2.7 Classification
A graphical interface in oneChannelGUI provides a link to the pamr and pmdclass packages designed to carry out sample classification from gene expression data, respectively, by the method of nearest shrunken centroids (Tibshirani et al., 2002) and by penalized discriminant methods.

2.8 Biological interpretation
An important point downstream from the identification of a set of differentially expressed genes is the definition of the presence of possible relations between each other, e.g. belonging to the same biological pathway. For this reason we have implemented a graphical interface to the GOstats package (Gentleman et al., 2004), which has extensive facilities for testing the association of gene ontology (GO) terms to genes in a gene list. Graphical modules are also implemented to allow the selection and annotation of differentially expressed probe sets associated to a specific enriched GO term.

Furthermore, a routine to build template files to be loaded on IPA software (www.ingenuity.com) starting from limma/SAM/RankProd outputs has also been implemented.

2.9 Exon analysis
Modules to calculate Splice Index (Affymetrix-Technical-Note) and Microarray Detection of Alternative Splicing [MiDAS; (Affymetrix-White-Paper)] P-value are implemented. A filter based on MiDAS P-values is also available for the selection of putative alternative splicing events.

3407
3 CONCLUSION

oneChannelGUI is a microarray analysis tool intended to enable the interested life scientists to analyze microarray experiments based on the consolidated 3' Affymetrix expression arrays as well as the new whole transcript expression arrays. oneChannelGUI is also a didactical tool since it could be used to introduce young life scientists to the use and interpretation of microarray data. For this purpose various data sets and exercises are available at the oneChannelGUI web site.

4 PERSPECTIVES

We will expand the exon level analysis tools and add new interfaces to allow life scientists to gain maximum advantage from the vast analysis opportunities offered by Bioconductor packages.

ACKNOWLEDGEMENTS

This work was supported by an Oncogenomics grant from the Italian Association for Cancer Research; the Italian Ministero dell'Università e della Ricerca; the University of Torino; the Regione Piemonte: bando regionale sulla ricerca scientifica applicata 2004.

Conflict of Interest: none declared.

REFERENCES


