BioNet: an R-Package for the functional analysis of biological networks

Daniela Beisser1, Gunnar W. Klau2, Thomas Dandekar1, Tobias Müller1,* and Marcus T. Dittrich1,*

1Department of Bioinformatics, Biocenter, University of Würzburg, Am Hubland, 97074 Würzburg, Germany and 2Life Sciences group, CWI, Science Park 123, 1098 XG Amsterdam, The Netherlands

Associate Editor: Limsoon Wong

ABSTRACT

Motivation: Increasing quantity and quality of data in transcriptomics and interactomics create the need for integrative approaches to network analysis. Here, we present a comprehensive R-package for the analysis of biological networks including an exact and a heuristic approach to identify functional modules.

Results: The BioNet package provides an extensive framework for integrated network analysis in R. This includes the statistics for the integration of transcriptomic and functional data with biological networks, the scoring of nodes as well as methods for network search and visualization.

Availability: The BioNet package and a tutorial are available from BioNet
http://bionet.bioapps.biozentrum.uni-wuerzburg.de
Contact: marcus.dittrich@biozentrum.uni-wuerzburg.de; tobias.mueller@biozentrum.uni-wuerzburg.de

Received on December 14, 2009; revised on February 5, 2010; accepted on February 22, 2010

1 INTRODUCTION

Integrated analysis of microarray data in the context of biological networks such as protein–protein interaction (PPI) networks has become a major technique in systems biology. The primary objective is the identification of functional modules (significantly differentially expressed subnetworks) within large networks. This can be achieved by computing a score for each node in the network reflecting its functional relevance. Subsequently, a network search algorithm is required to find the highest scoring subgraph. In fact, this problem has been proven to be NP-hard (Iedeker et al., 2002). Various heuristic approaches have been proposed, most of them inspired by the seminal work from Ideker et al. (2002) with a human PPI network based on human protein reference databases (Gentleman et al., 2004) of the graph packages igraph, RBGL as well as igraph are supported (Carey et al., 2005; Csardi and Nepusz, 2006). Networks can be imported and exported in different formats, allowing a smooth data exchange with standard network analysis tools like Cytoscape (Shannon et al., 2003).

2 DESCRIPTION

The BioNet package provides a comprehensive set of methods for the integrated analysis of gene expression data and biological networks. P-values are distributed uniformly under null hypotheses. Therefore, as a first step, multiple P-values derived from the analysis of different experiments (e.g. t-test or regression models) can be aggregated using a uniform order statistics (aggrPvals) (Dittrich et al., 2008). The resulting distribution of combined P-values can be considered as a mixture of signal and noise, where the signal component is modelled to be Beta(α,1) distributed (Pounds and Morris, 2003). The model fit can be verified by the provided diagnostic plots (plot.bum, hist.bum). By fitting a beta-uniform mixture (BUM) model fitBumModel, the maximum-likelihood estimates for the mixture model can be obtained. These parameters are subsequently used to score the nodes of the network (scoreNodes, scoreFunction).

3 APPLICATION

We apply our package to gene expression data from diffuse large B-cell lymphomas (DLBCL) and survival data (Rosenwald et al., 2002) with a human PPI network based on human protein reference database HPRD; Prasad et al., 2009) as described in Dittrich et al. (2008). The data consist of 112 tumors with the germinal center B-like phenotype (GCB) and 82 tumors with the activated B-like phenotype (ABC) and includes information on patient survival. All data are available in the BioNet and DLBCL package. We use standard microarray analysis and Cox regression to obtain gene-wise P-values for differential expression and risk association.

*To whom correspondence should be addressed.
respective. Then we aggregate both \( p \)-values by the second-order statistics using BioNet.

\[
\begin{align*}
\text{data(dataLym)} \\
pvals <- \text{cbind}(t=dataLym$t.pval, s=dataLym$s.pval) \\
pval <- \text{aggrPrvals}(pvals, order=2, plot=FALSE)
\end{align*}
\]

We now fit a BUM model to the distribution of aggregated \( p \)-values and score the nodes using an FDR threshold of 0.001.

\[
\begin{align*}
\text{fb <- fitBumModel(pval, plot=FALSE)} \\
scores <- \text{scoreNodes(network, fb=fb, fdr=0.001)} \\
\text{writeHeinz(network, file="lym_001", node.scores=scores)}
\end{align*}
\]

The exact search algorithm can be started from R by \text{runHeinz} if the CPLEX library is installed (Dittrich \textit{et al}., 2008). Alternatively, the fast heuristic search algorithm (\text{runFastHeinz}) often delivers a close approximation. Finally, the resulting modules can be visualized in 2D or 3D.

\[
\begin{align*}
\text{module <- readHeinzGraph(node.file=}
\text{ "lym_001_n.txt.0.hnz", network)} \\
\text{plot3dModule(module)}
\end{align*}
\]

BioNet captures an interaction module that has been described to play major biological roles in the GCB and ABC DLBCL subtypes (Fig. 1C). The combination of biological and clinical data with PPI networks generates a meaningful biological context in terms of functional association for differentially expressed, survival-relevant genes.

\textbf{Funding:} BMBF (Funcrypta); DFG (Da208/10-1); DFG (TR34/A5).

\textbf{Conflict of Interest:} none declared.

\textbf{REFERENCES}


