Bioinformatics, 31(10), 2015, 1686–1688
doi: 10.1093/bioinformatics/btu864
Advance Access Publication Date: 18 January 2015
Applications Note

OXFORD

Systems biology

Functional Gene Networks: R/Bioc package to generate and analyse gene networks derived from functional enrichment and clustering

Sara Aibar, Celia Fontanillo†, Conrad Droste and Javier De Las Rivas*

Bioinformatics and Functional Genomics Research Group, Cancer Research Center (Consejo Superior de Investigaciones Científicas, Universidad de Salamanca and Instituto de Investigación Biomédica de Salamanca, CSIC/USAL/IBSAL), Salamanca, Spain

*To whom correspondence should be addressed.
†Present address: Celgene Institute for Translational Research Europe (CITRE), Sevilla, Spain
Associate Editor: Jonathan Wren

Received on June 20, 2014; revised on December 16, 2014; accepted on December 29, 2014

Abstract

Summary: Functional Gene Networks (*FGNet*) is an R/Bioconductor package that generates gene networks derived from the results of functional enrichment analysis (FEA) and annotation clustering. The sets of genes enriched with specific biological terms (obtained from a FEA platform) are transformed into a network by establishing links between genes based on common functional annotations and common clusters. The network provides a new view of FEA results revealing gene modules with similar functions and genes that are related to multiple functions. In addition to building the functional network, *FGNet* analyses the similarity between the groups of genes and provides a distance heatmap and a bipartite network of functionally overlapping genes. The application includes an interface to directly perform FEA queries using different external tools: *DAVID*, *GeneTerm Linker*, *TopGO* or *GAGE*; and a graphical interface to facilitate the use.
Availability and implementation: *FGNet* is available in Bioconductor, including a tutorial. URL: http://bioconductor.org/packages/release/bioc/html/FGNet.html
Contact: jrivas@usal.es
Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Due to the increasing number of omic studies, efficient functional analysis of large lists of genes or proteins is essential to understand the biological processes in which they are involved. Functional enrichment analysis (FEA) is the most popular bioinformatic methodology to obtain significant functional information from sets of cooperating genes. FEA methods search in biological databases (such as Gene Ontology and KEGG pathways, among others) and use statistical testing to find biological terms and functional annotations that are significantly enriched in a list of genes. However, in most cases the results of these analyses are very long lists of biological terms associated to genes that are difficult to digest and interpret. Some tools cluster the FEA results, like *DAVID-FAC* (Huang *et al.*, 2009) and *GeneTerm Linker* (Fontanillo *et al.*, 2011), but their output is provided as large tables and there are not many tools to integrate and visualize these results. Here we present Functional Gene Networks (*FGNet*), an R/Bioconductor package that uses FEA results to perform network-based analyses and visualization. The main network is built by establishing links between genes annotated to similar functional terms. In this way, *FGNet* generates and provides a network representing the links and associations between the clusters of genes and enriched terms. The network summarizes and facilitates the interpretation of the biological processes significantly enriched in the initial list of genes, revealing important information such as: distance and overlap

between clusters, identification of modules and hubs. The tool can also help to disclose new associations among genes cooperating in hidden biological processes not annotated yet, which can be revealed by the topology of the functional network.

2 Methods

2.1 Input: functional enrichment and clustering

FGNet builds functional networks based on the groups obtained from clustering *gene-term sets* (*gtsets*, genes and terms associated by an enrichment p-value) returned by a FEA. The package includes an interface to do queries with gene lists using four FEA tools: *DAVID* with *Functional Annotation Clustering* (that returns clustered *gtsets*, Cl); *GAGE* (that also provides clusters) (Luo *et al.*, 2009); *GeneCodis* with *GeneTerm Linker* (that returns metagroups, Mg) and *TopGO* (that only returns *gtsets*) (Alexa *et al.*, 2010). The package can be also applied to the results from other EA tools, as long as the input results are transformed into tables of genes and associated terms.

2.2 Construction of the functional network

The functional network is built based on the analysis of all the *gtsets* provided by the FEA tool. These sets allow to generate a boolean matrix M of *genes* by *gtsets*, in which each element $m_{g,s} = 1$ if gene g is in set s. This membership matrix is then transformed into an adjacency matrix A $n \times n$; being n the total number of genes and $a_{i,j}$ the number of *gtsets* s in which a gene-pair is included: $a_{i,j} = \sum_s (m_{i,s} \times m_{j,s})(1 - \delta_{i,j})$, where δ is a *Kronecker* delta ($\delta_{i,j} = 1$ if $i = j$, $\delta_{i,j} = 0$ if $i \neq j$). This adjacency matrix is used to generate the functional network by establishing a weighted link between each pair of genes (g_i, g_j) in which $a_{i,j} \neq 0$. Finally, the clustering of *gtsets* provided by the FEA tool is used to generate a second genes' adjacency matrix with the number of common clusters/metagroups (Fig. 1A), that is used to define and allocate gene groups. The network produced is provided as an *igraph* object for further analysis, and can be exported to other network-based tools like *Cytoscape*.

2.3 Visualization and plots of the functional network

The main plot of the network presents the functionally associated genes (Fig. 1B). Edges link the genes that are in the same *gtsets*. Nodes within the same Cl/Mg are placed together using a force-directed *Fruchterman–Reingold* layout, within a common background colour. Genes in only one Cl/Mg are plotted with the colour of such group, while genes that are included in more than one Cl/Mg are left white.

2.4 Analysis of functional modules in the network

To facilitate the analysis and quantification of the modules and the overlap between groups, *FGNet* also provides a distance matrix and a heatmap (Fig. 1C), plus an intersection network (Fig. 1D). The distance matrix is calculated based on the pairwise binary distance in the adjacency matrix of common Cls/Mgs. These distances are analysed by hierarchical average linkage and plotted as a heatmap that reveals the proximity and similarity between the groups of genes (Cls/Mgs). The intersection network is a bipartite network which includes only the genes associated to several Cls/Mgs (white nodes in Fig. 1B,D), showing their connectivity to such Cls/Mgs. This intersection network facilitates the identification of *multifunctional* genes. (For more details see *FGNet* documentation in Bioconductor).

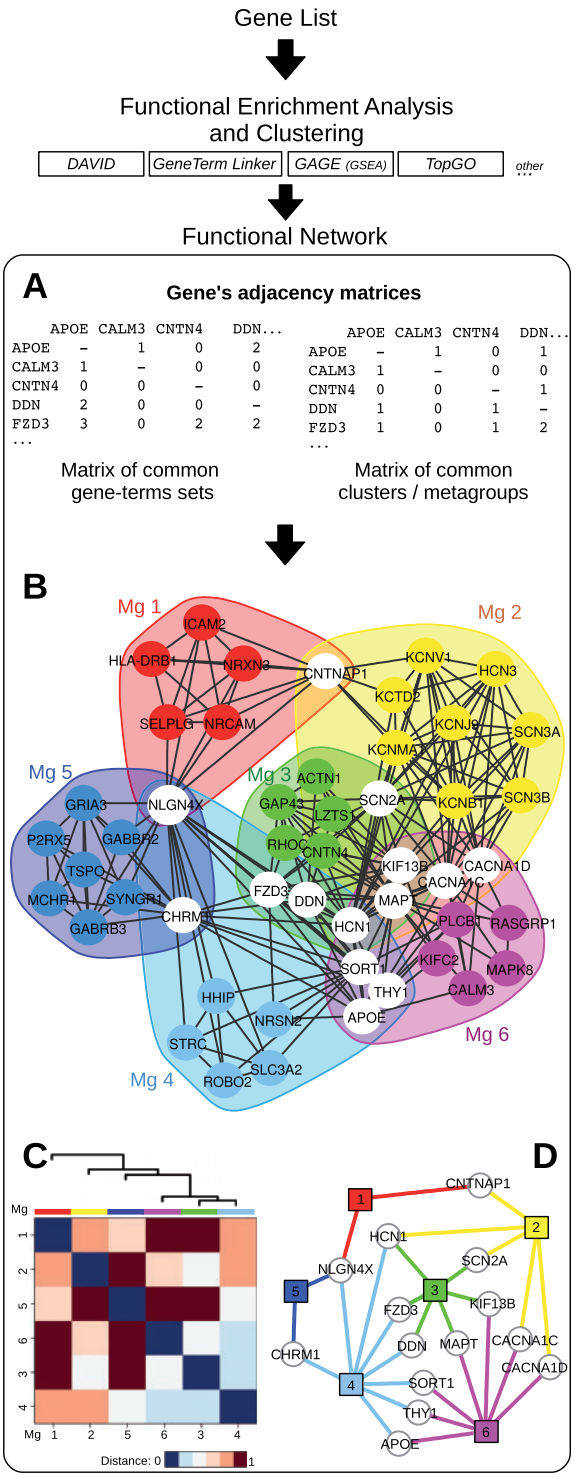

Fig. 1. Schematic workflow. A query gene list is analysed through a FEA tool and the generated *gene-term sets* are used to build: (**A**) gene's adjacency matrices; (**B**) a functional network (general view); (**C**) a distance heatmap and (**D**) an intersection network (to highlight multifunctional genes)

3 Example of use

We applied the method to several datasets and confirmed that the functional network greatly facilitates the analysis of enrichment results. Figure 1 shows the results of *FGNet* for a list of 175 genes

differentially expressed in human samples of entorhinal cortex neurons from Alzheimer's disease (AD) patients (obtained from Gene Expression Omnibus database, GEO: dataset GSE4757). Performing a FEA through *GeneTerm Linker*, we obtained six meta-groups that we labelled according to their main annotations: (Mg1) cell adhesion; (Mg2) voltage-gated ion/potassium channels; (Mg3) axon and cell projection; (Mg4) dendrite and neuronal cell body; (Mg5) synaptic neuroactive ligand-receptor interaction and (Mg6) MAPK signaling and Alzheimer. The network of these six Mgs (Fig. 1B) provides a global overview of the functionally overlapping genes and allows to identify hub genes that interconnect groups. For example, CNTNAP1 and NLGN4X appear as hubs in Mg1. CNTNAP1 (that regulates distribution of K+ channels) links Mg1 and 2; and NLGN4X (that facilitates synaptic neurotransmission) links Mg1 with 4 and 5. NLGN4X is the gene with highest betweenness centrality in this network. Another important hub is APOE, recently associated to Alzheimer. The distance matrix (Fig. 1C) allows to quantify the similarity between gene groups, showing that the closest Mgs are 3, 4 and 6, sharing eight nodes. This is also presented in the intersection network (Fig. 1D). Finally, the functional network can reveal further information about some Mgs. For example, if a Mg shares many genes with several other Mgs, it will indicate that such Mg brings the most common features that define the studied biological state. This is the case for Mg6, which, in fact, is annotated to Alzheimer's Disease.

Funding

This work was supported by the "Accion Estrategica en Salud" (AES) of the "Instituto de Salud Carlos III" (ISCiii) from the Spanish Government (projects granted to J.D.L.R.: PS09/00843 and PI12/00624); and by the "Consejeria de Educación" of the "Junta Castilla y Leon" (JCyL) and the European Social Fund (ESF) with grants given to S.A. and C.D.

Conflict of Interest: none declared.

References

Alexa,A. and Rahnenfuhrer,J. (2010). topGO: Enrichment analysis for Gene Ontology. R package version 2.18.0.

Fontanillo,C. *et al.* (2011). Functional analysis beyond enrichment: non-redundant reciprocal linkage of genes and biological terms. *PLoS One,* **6**, e24289.

Huang,D. *et al.* (2009). Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. *Nat. Protoc.*, **4**, 44–57.

Luo,W. *et al.* (2009). GAGE: generally applicable gene set enrichment for pathway analysis. *BMC Bioinformatics,* **10**, 161.