FunNet: an integrative tool for exploring transcriptional interactions

Edi Prifti¹,²,³, Jean-Daniel Zucker¹,³, Karine Clement¹,²,⁴ and Corneliu Henegar¹,∗

¹INSERM, UMR-S 872, Les Cordeliers, Eq. 7 Nutriomique, Paris F-75006, ²Pierre et Marie Curie-Paris 6 University, Cordeliers Research Center, UMR-S 872, Paris F-75006, ³IRD, UR Géodes, Centre IRD de l’Ile de France, Bondy F-93143 and ⁴Assistance Publique-Hôpitaux de Paris, Pitié Salpêtrière Hospital, Nutrition and Endocrinology department, Paris F-75013, France

Received on June 17, 2008; revised on July 23, 2008; accepted on September 12, 2008

ABSTRACT

Summary: We describe here an exploratory tool, called FunNet, which implements an original systems biology approach, aiming to improve the biological relevance of the modular interaction patterns identified in transcriptional co-expression networks. A suitable analytical model, involving two abstraction layers, has been devised to relate expression profiles to the knowledge on transcripts’ biological roles, extracted from genomic databases, into a comprehensive exploratory framework. This approach has been implemented into a user-friendly web tool to promote its open use by the community.

Availability: http://www.funnet.info
Contact: edi.prifti@crc.jussieu.fr

1 MOTIVATION

Interactions analysis has emerged as an increasingly popular framework for exploring the complex system of relations that characterize the functional organization of cellular environments. In this context, network abstractions are used to represent the interdependence between various interacting entities (i.e. molecules, cellular and extracellular structures, processes and regulatory pathways), and support the formal assessment of their contextual relevance. Among many other applications, the exploratory analysis of the complex interactions that condition transcriptional expression patterns has brought significant insights into the robustness of the cells’ functional organization, its evolution and the adaptive mechanisms that are continuously shaping transcriptional regulation. Built by relating together transcripts that share similar expression profiles, the gene co-expression networks display many topological properties commonly observed in other complex interaction networks. Among them, the hierarchical organization of their modular architectures, spread over multiple scales, is thought to carry remarkable biological significance, illustrating the strong relationship which connects transcripts’ regulatory patterns to the functional organization of the cell (Allocco et al., 2004; Barabasi and Oltvai, 2004).

The assessment of gene co-expression relies on various measures, including correlation coefficients between transcriptional profiles, as well as entropic informational criteria. Despite their intuitive character, the biological relevance of these measures is limited by their sensitivity to experimental noise that affects high-throughput gene expression measurements. To increase the robustness to noise several topological overlap criteria have been recently suggested (Yip and Horvath, 2007; Zhang and Horvath, 2005).

Here, we propose a conceptually different strategy, aiming to improve the biological relevance of the modular interaction patterns identified in co-expression networks, in a manner that is complementary to the aforementioned techniques. To this purpose, an original systems biology approach has been devised to extract the available knowledge of transcripts’ roles from genomic databases, and use it to direct the analysis of the transcriptional interactions. A user-friendly web tool, named FunNet, has been built to promote the open use of this integrative approach by the community.

2 THE ANALYTICAL MODEL

An original two layer analytical model has been imagined to support the integration of the available information on transcripts’ biological roles into the co-expression network analysis (Fig. 1). A detailed description of the integrative approach implemented in FunNet is presented elsewhere (Henegar et al., 2008; Henegar et al., manuscript in preparation). The analytical model is organized around an abstraction of the cellular processes, represented formally by the biological themes that annotate transcripts’ roles in genomic databases [Gene Ontology (GO), KEGG, etc.]. A non-linear dynamical system allows quantifying the contextual proximity of annotating themes based on the similarity of their annotated transcriptional profiles. Relevant themes are then related into a theme proximity network, overlaying the gene co-expression network, to capture and illustrate the interdependence in between the transcriptional regulation of associated cellular processes. A spectral clustering technique is applied to identify co-expression modules by grouping together themes sharing closely related transcriptional patterns. In the end, interactional centrality measures, such as intramodular connectivity or network betweenness (Yu et al., 2007), are computed to assess the contextual significance of the analyzed elements (i.e. transcripts and biological themes). This approach has been applied to explore the transcriptomic signature of the adipose tissue, to identify the pathological mechanisms underlying...
FunNet: functional analysis of transcriptional interactions

Fig. 1. An illustration of the two layer analytical model implemented in FunNet, applied to explore the transcriptomic signature of the adipose tissue in obese human subjects compared with a set of normally lean subjects: (A) the bi-modal theme proximity network illustrating the interdependence between the regulatory patterns of various KEGG pathways in this condition; continuous lines are indicating the strongest interactions (i.e. superior to the upper quartile of their distribution), while dashed lines are depicting medium strong interactions (i.e. in between the median of the distribution and its upper quartile); the interactional centrality of KEGG themes is illustrated by the size of the corresponding nodes; (B) the underlying transcriptional co-expression network.

3 IMPLEMENTATION

FunNet computational routines were implemented in R and are available as a standalone package from worldwide CRAN mirrors (http://cran.r-project.org/mirrors). The graphical interface was written in PHP and is installed on a public web server, which has been referenced in the microarray tools’ section of the GO website (http://www.geneontology.org). The web interface is divided into four main sections. The first one allows the user to select a reference genome and the type of analytical situation to be considered. In the second, data files are uploaded, while the third provides various analytical options. The last section summarizes the uploaded data and selected options before concluding the submission.

3.1 Input data

FunNet was designed to address two types of situations: the analysis of one set of transcriptional profiles, or the simultaneous analysis of two sets of gene expression profiles distinguished by their overall regulation pattern in the analyzed condition (i.e. up- versus down-regulated transcripts, as those illustrated in Fig. 1). Expression data have to be submitted as specifically formatted text files, in which transcripts are identified through EntrezGene IDs. More details on the format of the data files and the available options are provided online.

3.2 Analytical options

The first set of options relates to the functional profiling of gene expression data, performed to identify a list of significantly overrepresented GO and KEGG categories, among those annotating the biological roles of the analyzed transcripts. Several ways of computing the gene enrichment for ontological themes are provided in relation to their informational specificity (i.e. the degree of precision with which they illustrate the biological information available on transcripts’ roles), or their conceptual granularity (i.e. the degree of generality of their conceptual content, indicated by the terminological level of the GO lattice to which they belong). A decorrelated annotation procedure, inspired by a method proposed by Alexa et al. (2006), is also provided. A false discovery ratio (FDR) correction approach (Storey and Tibshirani, 2003) can be optionally
selected to assess the statistical significance of the annotating themes' overrepresentation.

A second set of options is provided in relation to the analysis of transcriptional interactions. As co-expression measures, FunNet uses either Spearman’s or Pearson’s correlation coefficients, or the Euclidean distance in between the vectors of gene expression measurements. A co-expression significance threshold can be estimated through a separate procedure, which computes the goodness of fit of the theoretical scale-free topological model for various threshold values applied to the analyzed gene expression data. Optionally, a topological overlap criterion can be combined with the aforementioned co-expression measures to improve their robustness to experimental noise (Zhang and Horvath, 2005).

### 3.3 Results

After the completion of the analyses, the results can be downloaded in a compressed archive format. Various files are generated, some of which contain details on the functional profile of the analyzed transcripts, while others are specifically formatted to allow for the visualizing and annotating of the identified transcriptional modules in the Cytoscape® software (Shannon et al., 2003). Besides the overall network architecture, various patterns can be thus illustrated (Fig. 1), including the overall regulation of the analyzed transcripts and annotating themes, the strength of their interactions within the network, and their interactional centrality. The last one is thought to illustrate the contextual relevance of the analyzed elements (Barabasi and Oltvai, 2004), and therefore may provide a valuable insight to the choice of target processes and genes to be further explored in subsequent biological experiments.

*Funding:* European Community 6th framework HEPADIP project; the ‘Association de Langue Française pour l’Étude du Diabète et des Maladies Métaboliques’ (ALFEDIAM).

*Conflict of Interest:* none declared.

### REFERENCES


