NetworkTrail—a web service for identifying and visualizing deregulated subnetworks

Daniel Stöckel[1,*,†], Oliver Müller[1,†], Tim Kehl[1], Andreas Gerasch[2], Christina Backes[3], Alexander Rurainski[4], Andreas Keller[3,5], Michael Kaufmann[2] and Hans-Peter Lenhof[1]

[1]Center for Bioinformatics, Saarland University, 66041 Saarbrücken, [2]Wilhelm Schickard Institute for Computer Science, Eberhard Karls University, 72076 Tübingen, [3]Department of Human Genetics, Saarland University, 66421 Homburg, [4]Scientific Consilience, Starterzentrum, Saarland University, 66123 Saarbrücken and [5]Siemens Healthcare, Strategy, 91052 Erlangen, Germany

Associate Editor: Janet Kelso

ABSTRACT

Summary: The deregulation of biochemical pathways plays a central role in many diseases like cancer or Parkinsons's disease. *In silico* tools for calculating these deregulated pathways may help to gain new insights into pathogenic mechanisms and may open novel avenues for therapy stratification in the sense of personalized medicine. Here, we present NetworkTrail, a web service for the detection of deregulated pathways and subgraphs in biological networks. NetworkTrail uses a state-of-the-art integer linear programming-based approach for this task and offers interfaces to the Biological Network Analyzer (BiNA) and Cytoscape Web for visualizing the resulting subnetworks. By providing an accessible interface to otherwise hard-to-use command line tools, the new web service enables non-experts to quickly and reliably carry out this type of network analyses.

Availability and implementation: NetworkTrail is a JavaServer Pages-based web service. The algorithm for finding deregulated subnetworks has been implemented in C++. NetworkTrail is available at http://networktrail.bioinf.uni-sb.de/.

Contact: dstoeckel@bioinf.uni-sb.de

1 INTRODUCTION

The deregulation of cell signaling pathways plays an important role during tumor development. To identify such pathways a number of approaches have been developed during the past decade. The method of Ideker *et al.* (2002) for the detection of active subgraphs is available as a Cytoscape (Shannon *et al.*, 2003) plug-in. The BioNet software developed by Dittrich *et al.* (2008) can be downloaded as an R package (http://www.r-project.org). Vandin *et al.* (2011) have developed HotNet, which is available as MATLAB source or through a web interface. The algorithms developed by Dao *et al.* (2011) and Keller *et al.* (2009) are provided as C++ source code. We have recently published an efficient approach based on integer linear programming (ILP) (Backes *et al.*, 2011), which is available as executable.

Although a multitude of approaches for the analysis of expression data is exposed through web services (e.g. Martín-Requena *et al.*, 2010; Rainer *et al.*, 2006), the tools for the search of deregulated subnetworks aforementioned (except HotNet) have to be downloaded and installed before usage. Furthermore, the majority of the tools do neither provide a graphical user interface nor an option to visualize the resulting subgraphs directly. Instead, most of the programs have to be executed from the command line, and their outputs need to be processed using third party packages. Thus, besides installing the required software properly, users are required to possess intricate technical knowledge of the complete tool chain. According to our experiences, both tasks can be challenging for non-expert users. An easy-to-use web interface hiding such technical details, and thereby allowing the user to focus on the analysis, would overcome these problems. Therefore, we have developed a web interface called NetworkTrail that enables the user to detect deregulated pathways and subgraphs in biological networks using our recently published ILP-based approach (Backes *et al.*, 2011). Based on a score for each node that measures the degree of deregulation of the corresponding gene or protein, the heaviest connected subnetwork of a specified size k is computed. Here, connectivity means that the subnetwork contains a designated root node from which all other nodes in the subnetwork are reachable. This node can be interpreted as a potential molecular key player responsible for the observed deregulation. By computing all such subnetworks in a specified size range, it is possible to directly assess the stability of the computed result.

As input, the user can specify a Gene Expression Omnibus (GEO) record. If the user desires to analyze raw expression data, she/he currently is required to normalize the data and provide per-gene expression scores. After the computations are completed, the resulting subnetwork can then be visualized using Cytoscape Web (Lopes *et al.*, 2010) or the Biological Network Analyzer (BiNA, http://bina.sourceforge.net/).

2 FEATURES AND METHODS

2.1 The web interface

A wizard-like web interface guides the user through the steps depicted in Figure 1: (1a) upload of pre-computed gene scores provided in a tabular format. For this purpose, various external
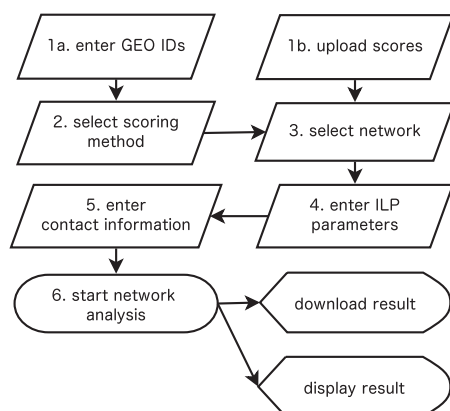
---

*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.

Fig. 1. The user is guided through the workflow from uploading data to visualizing the results. For each step, a detailed tutorial is available

database identifiers are supported. If not specified explicitly, NetworkTrail attempts to infer the used identifiers. (1b) Alternatively, either one GEO series (GSE) file or two GEO dataset (GDS) files (one file represents the sample group to be compared with the reference group contained in the second file) can be uploaded. In case of GSE and GDS files, it is sufficient to enter valid GEO identifiers; the corresponding files are then downloaded from GEO and stored on our server. As GSE files usually contain samples from multiple groups, the user can freely divide them into a sample and a reference set. (2) If no pre-computed scores have been uploaded in the previous step, the user can select among different scoring methods, such as *Student's t-test*, *log-mean fold quotients* or the *Z-score*, if one group consists of only one sample. (3) NetworkTrail detects the organism from which the scores were derived. The user is given the option to correct this choice if necessary and can choose a network for further analysis. Currently, regulatory networks for mouse, rat, human and *Arabidopsis* derived from the KEGG regulatory pathways database (Kanehisa *et al.*, 2010) are supported. (4) As parameters, our ILP-based approach (Backes *et al.*, 2011) requires a range of subnetwork sizes that should be analyzed. As the KEGG pathways contain nodes that correspond to families or complexes of genes/proteins, multiple scores need to be aggregated to obtain a single score for these cases. As aggregation functions we offer min, max and mean. (5) The user can choose to be notified via email on completion of the analysis. (6) The results can then be visualized instantaneously from the web browser using either Cytoscape Web (Lopes *et al.*, 2010) or the Java Webstart-based BiNA. Both tools present the union of all computed subgraphs in a hierarchical layout. The user can individually highlight the subnetwork for specific values of k. The degree of deregulation is indicated by the node color, where white represents no change and shades of red and green represent up- and down-regulation, respectively. Moreover, BiNA allows for a complete customization of the visualization, including layout, colors and node styles.

2.2 Implementation details

NetworkTrail's server implementation is built on top of JavaServer Pages. JavaScript and AJAX are used for client-server communication, with a focus on responsive status and error reporting. The interface has been tested on all major browsers and operating systems. The mapping of scores to network nodes and the ILP approach are implemented in C++ using the CPLEX solver. The results of GEO queries are locally cached for improved performance of repeated computations. No registration is required, and obtained results are stored for a 2 week period. All computed data are provided for download in a single archive containing the computed subnetworks, the scores used for the computation and mappings of nodes to external database identifiers in Cytoscape simple interaction format (.sif) and/or node attribute format (.na). For documentation purposes, a file containing all used parameters is included.

3 CONCLUSION

In this article, we presented NetworkTrail, a web service for detecting deregulated subnetworks in biological network data. The availability of such a web service allows non-experts to carry out network analyses without having to struggle with technical details, such as compiling/installing software or learning cryptic commands. We are convinced that such easy-to-use web services will help to elucidate pathogenic mechanisms, and that they may also prove to be useful for therapy stratification in cancer therapy. In the future, NetworkTrail will be enhanced by including further analysis methods and pre-defined networks.

Conflict of Interest: none declared.

REFERENCES

Backes,C. *et al.* (2011) An integer linear programming approach for finding deregulated subgraphs in regulatory networks. *Nucleic Acids Res.*, **40**, e43.

Dao,P. *et al.* (2011) Optimally discriminative subnetwork markers predict response to chemotherapy. *Bioinformatics*, **27**, i205–i213.

Dittrich,M.T. *et al.* (2008) Identifying functional modules in protein-protein interaction networks: an integrated exact approach. *Bioinformatics*, **24**, i223–i231.

Ideker,T. *et al.* (2002) Discovering regulatory and signalling circuits in molecular interaction networks. *Bioinformatics*, **18** (**Suppl. 1**), S233–S240.

Kanehisa,M. *et al.* (2010) Kegg for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res.*, **38**, D355–D360.

Keller,A. *et al.* (2009) A novel algorithm for detecting differentially regulated paths based on gene set enrichment analysis. *Bioinformatics*, **25**, 2787–2794.

Lopes,C. *et al.* (2010) Cytoscape web: an interactive web-based network browser. *Bioinformatics*, **26**, 2347–2348.

Martín-Requena,V. *et al.* (2010) jORCA: easily integrating bioinformatics web services. *Bioinformatics*, **26**, 553–559.

Rainer,J. *et al.* (2006) CARMAweb: comprehensive R- and bioconductor-based web service for microarray data analysis. *Nucleic Acids Res.*, **34** (**Suppl. 2**), W498–W503.

Shannon,P. *et al.* (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.*, **13**, 2498–2504.

Vandin,F. *et al.* (2011) Algorithms for detecting significantly mutated pathways in cancer. *J. Comput. Biol.*, **18**, 507–522.