INfORM: Inference of NetwOrk Response Modules

Veer Singh Marwah\textsuperscript{1,2}, Pia Anneli Sofia Kinaret\textsuperscript{1,2,3}, Angela Serra\textsuperscript{4}, Giovanni Scala\textsuperscript{1,2}, Antti Lauerma\textsuperscript{5}, Vittorio Fortino\textsuperscript{1,2} and Dario Greco\textsuperscript{1,2,3,*}

\textsuperscript{1}Faculty of Medicine and Life Sciences, University of Tampere, 33014 Tampere, Finland, \textsuperscript{2}Institute of Biosciences and Medical Technologies (BioMediTech), 33520 Tampere, Finland, \textsuperscript{3}Institute of Biotechnology, University of Helsinki, 00014 Helsinki, Finland, \textsuperscript{4}NeuRoNeLab, DISA-MIS, University of Salerno, 84084 Fisciano, Italy and \textsuperscript{5}Department of Dermatology and Allergology, University of Helsinki and Helsinki University Central Hospital, 00029 Helsinki, Finland

\*To whom correspondence should be addressed.

Abstract

Summary: Detecting and interpreting responsive modules from gene expression data by using network-based approaches is a common but laborious task. It often requires the application of several computational methods implemented in different software packages, forcing biologists to compile complex analytical pipelines. Here we introduce INfORM (Inference of NetwOrk Response Modules), an R shiny application that enables non-expert users to detect, evaluate and select gene modules with high statistical and biological significance. INfORM is a comprehensive tool for the identification of biologically meaningful response modules from consensus gene networks inferred by using multiple algorithms. It is accessible through an intuitive graphical user interface allowing for a level of abstraction from the computational steps.

Availability and implementation: INfORM is freely available for academic use at https://github.com/Greco-Lab/INfORM.

Contact: dario.greco@staff.uta.fi

Supplementary information: Supplementary data are available at Bioinformatics online.
Second, the ranked connections from the summarized networks are aggregated by using the Borda method implemented in TopKLists R package (Schimek et al., 2015) and the top ranked edges are progressively included in the final network until all the nodes are incorporated, i.e. all the nodes have degree \( \geq 1 \) (Supplementary File S2). This procedure ensures the robustness of the network inference. Alternatively, the user can decide the top \( n\% \) of the ranked edges to include in the final network.

2.2 Gene module detection and responsive module definition

INfORM provides widely used community clustering algorithms for identifying the relevant modules from the inferred gene network, the ‘Walktrap’ (Pons and Latapy, 2005), ‘Spinglass’ (Newman and Girvan, 2004; Reichardt and Bornholdt, 2006), ‘Louvain’ (Blondel et al., 2008) and the ‘Greedy’ (Clauset et al., 2004) algorithms. Benchmark analysis showed that all the algorithms generate highly similar modules (Supplementary File S2). The relevance of the identified modules is evaluated by scoring them based on characteristics of the member nodes and edges: i) centrality, ii) differential log2(fold-change), iii) differential \( P \)-value, iv) median rank of edge weights and v) number of nodes. These are graphically represented as a radar chart, making it easy to evaluate the modules. Moreover, enrichment analysis is performed to find the GO terms overrepresented in each module and compute the similarity between sets of GO terms from different modules. A heatmap representation of the GO-based module similarity is provided to aid the selection of functionally related modules. INfORM allows the user to merge statistically significant and biologically relevant modules into an optimized response module. The biological functions associated with the response module are visualized by the means of a tile plot (Supek et al., 2011), in which the semantically similar GO terms are grouped together (Yu et al., 2010).

3 Conclusion

Here we present INfORM, a novel tool for robust inference of gene co-expression networks from transcriptomics data. The graphical user interface helps to perform the analysis without any obligatory technical expertise and in-depth knowledge of the implementation. Results from a case study analysis of a publically available data are provided in Supplementary File S3.

Acknowledgements

We are grateful to Dr. Lasse Ruokolainen (University of Helsinki) for benchmarking the INfORM software and providing technical expertise and in-depth knowledge of the implementation. We thank MSc. Juho Väinänen and Dr. Iiris Hovatta (University of Helsinki) for his comments on the methods. We thank MSc. Juho Väinänen and Dr. Iiris Hovatta (University of Helsinki) for his comments on the methods. We thank MSc. Juho Väinänen and Dr. Iiris Hovatta (University of Helsinki) for his comments on the methods.

Funding

This study was supported by the Academy of Finland (grant agreements 273151 and 292307), EU H2020 calIBRAtE project (grant agreement 686239), EU H2020 LIFEPATH (grant agreement 633666) and EU FP7 NANOSOLUTIONS project (grant agreement FP7-309329).

Conflict of Interest: none declared.

References


