Systems biology

**MODifieR: an Ensemble R Package for Inference of Disease Modules from Transcriptomics Networks**

Hendrik A. de Weerd 1,2,†, Tejaswi V. S. Badam 1,2,†, David Martínez-Enguita 2, Julia Åkesson 1,2, Daniel Muthas 3, Mika Gustafsson 2,†,* and Zelmina Lubovac-Pilav 1,†,*

1School of Bioscience, Systems Biology Research Center, Skövde 541 45, Sweden, 2Department of Physics, Chemistry and Biology, Linköping University, Linköping 581 83, Sweden and 3Translational Science and Experimental Medicine, Early Respiratory, Inflammation and Autoimmunity, BioPharmaceuticals R&D, AstraZeneca, Mölndal 43183, Sweden

*To whom correspondence should be addressed.
†The authors wish it to be known that, in their opinion, the first two authors and last two authors should be regarded as Joint Authors.
Associate Editor: Lenore Cowen

Received on December 4, 2019; revised on March 27, 2020; editorial decision on April 1, 2020; accepted on April 2, 2020

Abstract

**Motivation:** Complex diseases are due to the dense interactions of many disease-associated factors that dysregulate genes that in turn form the so-called disease modules, which have shown to be a powerful concept for understanding pathological mechanisms. There exist many disease module inference methods that rely on somewhat different assumptions, but there is still no gold standard or best-performing method. Hence, there is a need for combining these methods to generate robust disease modules.

**Results:** We developed MODule IdentiFIER (MODifieR), an ensemble R package of nine disease module inference methods from transcriptomics networks. MODifieR uses standardized input and output allowing the possibility to combine individual modules generated from these methods into more robust disease-specific modules, contributing to a better understanding of complex diseases.

**Availability and implementation:** MODifieR is available under the GNU GPL license and can be freely downloaded from https://gitlab.com/Gustafsson-lab/MODifieR and as a Docker image from https://hub.docker.com/r/ddeweerd/modifier.

**Contact:** zelmina.lubovac@his.se or mika.gustafsson@liu.se

**Supplementary information:** Supplementary data are available at Bioinformatics online.

1 Introduction

Various algorithms have been proposed to infer groups of disease-associated genes using networks, i.e. disease modules. Recently, the DREAM community compared different module inference methods that were based on network topologies to identify disease-associated genes (Choobdar et al., 2019) which led to the tool MONET (Tomasoni et al., 2019). However, no similar tools exist for extracting disease-specific modules generated from these methods into more robust disease-specific modules, contributing to a better understanding of complex diseases.

2 Software implementation

The main concept of MODifieR is illustrated in Figure 1. The general workflow is centered around two main objects: MODifieR input objects, that contain preprocessed data from...
microarray or RNA-Seq and MODifieR_module objects, containing inferred modules and related data.

2.1 Input
Input objects are created from transcriptomic data from microarrays or RNA-Seq data. Differential expression analysis of microarray data is performed on probe level using LIMMA (Ritchie et al., 2015). In case of multiple probes mapped to the same gene, the lowest P-value is selected. For co-expression analysis, probes are collapsed into genes using WGCNA collapseRows function (Miller et al., 2011). RNA-Seq differential expression analysis follows the standard edgeR workflow (Robinson et al., 2010). The preprocessing for co-expression-based analysis consists of variance-stabilizing transformation from the DESeq2 package (Love et al., 2014) and an optional quantile normalization. The objects contain DEGs, an expression matrix, grouping of samples and a list of parameters used when creating the object.

To work with a pre-computed list of DEGs or a co-expression matrix, the data can be enclosed in an input object. If a method requires DEGs, it also requires a protein–protein interaction network to overlay the DEGs on, provided separately when calling the module inference function. For details, see sub-chapter Inference methods in Vignette.

2.2 Module inference
All module inference methods take the input object described in Section 2.1 as their primary input. The same input object can be used for all inference methods.

The output is always a MODifieR_module object. All MODifieR_module objects contain at least the disease module genes and a list of parameters used to generate the object. Several inference methods offer post-processing functions, such as splitting the module into sub-modules.

2.3 Consensus modules
Two approaches for generating a consensus module by integrating individual modules are provided in MODifieR. The first is a simple count-based method where a consensus module is derived from a set of MODifieR_module objects by counting the frequency of each gene in the individual methods. Supplementary Material S1 provides an example of the inference of an asthma count-based consensus module. The second method is a network-based approach called S2B (double specific betweenness), relying on betweenness centrality, originally developed to find the overlap between gene sets associated with two diseases (Garcia-Vaquero et al., 2018). The integration methods also return a MODifieR_module object.

2.4 Exporting objects
All objects can be exported to either a series of csv files or an Excel workbook with multiple sheets.

3 Conclusion
MODifieR is a novel R package that provides a set of common disease module inference methods using standardized input and output. It helps to derive more robust disease-specific modules by combining results from these methods. MODifieR provides a valuable complement to existing tools for module-based biomarker discovery and thus contributes to understanding complex diseases.

Funding
This work was supported by the Knowledge Foundation, Swedish Research Council and Swedish foundation for strategic research.

Conflict of Interest: none declared.

References

Fig. 1. Workflow of MODifieR.Printini (A) Creating input objects. (B) Disease modules are inferred by individual methods. (C) Consensus module is derived by integrating modules from individual methods.