Identification of differentially expressed genes in SARS-Cov-2 infected cells using Bayesian network models

Nevena Ćirić†
Department of Computer science and Informatics
Faculty of Mathematics
Belgrade, Serbia
nevena.ciric@matf.bg.ac.rs

Aleksandar Veljković
Department of Computer science and Informatics
Faculty of Mathematics
Belgrade, Serbia
aleksandar@matf.bg.ac.rs

dr Gordana Pavlović-Lažetić
Department of Computer science and Informatics
Faculty of Mathematics
Belgrade, Serbia
gpavloviclazetic@gmail.com

ABSTRACT
The current outbreak of infectious disease caused by a novel coronavirus, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has been affecting millions of people and has caused devastating mortality worldwide. In this regard, development of drugs, vaccines and treatments addressing the SARS-CoV-2 infection have become a major focus. The identification of differentially expressed genes due to SARS-CoV-2 infection may provide valuable information about the underlying biology of the disease. It can give an insight into molecular mechanisms of disease by indicating the signaling pathways altered during the infection and finding key molecular players that can be targeted.

Network analysis is the most convenient method for representation of a functionally related set of genes and detection of changes in their expression. This study builds upon the Bayesian network model and coexpression network analysis applied to identification of differentially expressed genes in SARS-CoV-2 infected cells. Weighted gene coexpression network analysis (WGCNA) is used to group related genes into gene modules based on their coexpression patterns in non-infected cells. For each gene module, WGCNA computes one eigengene – weighted average of the expression of all the genes in that module, whereby weights are determined so that loss in the biological information is minimized. These eigengenes are used to train a Bayesian network in which nodes (random variables) represent gene modules and directed edges represent the conditional dependencies between corresponding gene modules. Besides random variables that model the expression value of each eigengene, the network has one additional binary variable which models type of sample – infected or non-infected. According to the Markov property of the Bayesian networks, the parents of that node are the modules most related to, thus they should be enriched with genes that are associated with the disease.

This task has been performed elsewhere to specific types of cells and using different network analysis methods - gene ontology, pathway enrichment analysis and functional protein network construction. The results obtained may serve as a doubled evidence of an important finding.

KEYWORDS
SARS-CoV-2, gene expression, Bayesian networks, coexpression network

REFERENCES


†First author