

Meta-dimensional analysis in gene network inference and gene prioritization associated to complex diseases

Carlos Eduardo Marchi, David Corrêa Martins Junior, Fábio Marchi

Federal University of ABC, A.C.Camargo Cancer Center

Gene prioritization approaches are rich tools for personalized medicine and system biology, helping to reveal genes involved in an abnormal condition and being a decisive factor in the diagnosis, the advancement of therapeutic practices and the development of less aggressive and more effective drugs. In this context, we are developing a computational system for gene prioritization which integrates a number of biological data sources in order to reveal genes associated to complex diseases. To pursue this goal, this system implements integrative methodologies based on meta-dimensional data analysis, such as concatenation and transformation to integrate gene expression, miRNA expression and DNA methylation data under case and control conditions. Furthermore, it has involved inference and analysis of complex networks in order to rank candidate genes to be associated to a given disease. The data preprocessing consists of cleaning and removal of outliers. After that, a technique of differential analysis involving the clinical data of the patients are applying to obtain the most promising genes. Next, several gene correlation networks are created for each data type and also by the combination of them, such as mRNA-miRNA, mRNA-DNA methylation and mRNA-miRNA-DNA methylation. Pearson and Spearman correlation methods can be applied in the networks inference process. Finally, the topological analysis is executed to rank genes, including several algorithms and criteria based on complex network theory, such as degree and betweenness centrality; Page Rank, Hyperlink-Induced Topic Search and Katz centrality. This topological analysis results in a scored gene list for each network, which can be analyzed together to create a final list of the top ranked genes. Our method is unique integrative approach that involves the biological data sources aforementioned and that focused on clinical data of patients, enabling future analysis to develop new drugs and treatments for groups of patients. We expect that the evolution of this system could be an important step toward personalized medicine useful to improve the understanding of a myriad of complex diseases, including cancers, neurodegenerative and neurodevelopment disorders.

Funding support: UFABC, CAPES