Nooshin Omranian[1,2]
Bernd Mueller–Roeber[1,2]
Zoran Nikoloski[2]
1-University of Potsdam, Institute of Biochemistry and Biology, Karl-Liebknecht-Str. 24-25, 14476 Potsdam,Germany
2-Systems Biology and Mathematical Modeling, Max-Planck-Institute for Molecular Plant Physiology, Am Mühlenberg 1, 14476 Potsdam, Germany
e-mail: omranian@mpimp-golm.mpg.de

PageRank-based identification of signaling crosstalk from transcriptomics data in Arabidopsis thaliana

The levels of cellular organization, from gene transcription to translation to protein-protein interaction and metabolism, operate via tightly regulated mutual interactions facilitating organismal adaptability and various stress responses. Characterizing the mutual interactions between genes, transcription factors, and proteins involved in signalling, termed crosstalk, is therefore crucial for understanding and controlling cell's functionality. Based on the type of data used in the analysis, the existing methods for identifying crosstalk can be divided into two groups: (1) proteomics-based, relying on integration of protein-protein interaction data with existing pathway information and (2) transcriptomics-based, employing high-throughput transcriptomics data sets from different conditions.

Here we propose and analyze a novel method for crosstalk identification which relies on transcriptomics data and overcomes the lack of available information for the signalling pathways in Arabidopsis thaliana. Our method employs a network-based transformation of the results from the statistical analysis of differential gene expression in carefully constructed groups of experiments (conditions). Modification of the PageRank algorithm is then used on the network constructed in the previous step to determine the putative transcripts interrelating different signalling pathways. With the help of the proposed method, we analyze a transcriptomics data set incorporating experiments on four different stresses/signals: nitrate, sulfur, iron, and hormone and identified a promising gene candidates involved in crosstalk.

In addition, we conduct a comparative analysis with the state-of-the art methods in this field which used a biclustering-based approach [1]. Unlike approaches based biclustering, our approach does not rely on any hidden parameters. To compare the two approaches, we use transcriptomics data sets from Arabidopsis thaliana under 31 different experimental conditions: 5 nitrate, 4 sulfur, 2 iron and 20 hormone experiments. Surprisingly, the biclustering-based approach fails to identify any candidate genes involved in the crosstalk of the analyzed signals. On the other hand, with the proposed method, we find a small set of interesting genes putatively involved in crosstalk (verified by literature search). The small number of genes involved in crosstalk of these signals could be attributed to: (1) the heterogeneity of the analyzed data and (2) the lack of raw data for all experiments, resulting in a non-uniform normalization. Consequently, we demonstrate that our proposed method is more efficient for species for which large transcriptomics data sets, normalized with same techniques, are available.

References

[1] D. Nero, G. Krouk, D. Tranchina, GM. Coruzzi: *A system biology approach highlights a hormonal enhancer effect on regulation of genes in a nitrate responsive "biomodule"* BMC Syst Biol **3** 59.