Using WGCNA and NERI algorithms for the identification of biological pathways associated to schizophrenia

A.S. Feltrin[1], A.C. Tahira[2], S.N. Simões[3], H. Brentani[2], D.C. Martins-Jr [1]
[1]CMCC - Universidade Federal do ABC, [2]IPq-HCFMUSP-USP, [3] Instituto Federal do Espírito Santo

Introduction: Using two gene expression databases related to schizophrenia (BAHN and KATO), we propose a new approach consisting of combining the results of two network analysis algorithms: Weighted Gene Correlation Network Analysis (WGCNA) [1] and Network-Medicine Relative Importance (NERI) [2]. Considering the differences between the two methods, our hypothesis is that both are capable of producing compelling results related to different aspects of schizophrenia's biological pathways; therefore, are complementary to each other. For that, we used replication and enrichment analysis using public databases.

Materials and Methods: WGCNA uses gene expression from two groups to build co-expression pairwise correlation matrices, using connectivity parameters for evaluation of the network. NERI also uses expression data, but its network construction is based on the integration of PPI (protein-protein interaction) databases, gene expression, and a previously chosen seed genes list. The network analyses are based on shortest ranking path and relative importance calculation. We conducted an enrichment analysis using DAVID 6.8 (Database for Annotation, Visualization and Integrated Discovery) [3] for the identification of partial biological function of each result as well the SuperExactTest (*R* Package) [4] to calculate the intersection of the gene lists of WGCNA and NERI results. The Modular Single Set Enrichment Test (MSET) [5] analysis for GWAS, transcriptome, methylation and *de novo* mutation databases related to schizophrenia was applied to evaluate the replication and accuracy of our new approach when compared with each method in separate.

Results: The WGCNA module represents a final network of 435 and 300 genes on BAHN and KATO expression data. The enrichment analysis (DAVID 6.8) of this group using PPI modules leads to 88 genes across 10 hyper-represented human modules (Bonferroni adj.p<0.05), mostly involving immunological and inflammatory processes. By using NERI, the final gene list was 150 genes for both BAHN and KATO with the enrichment analysis leading to modules related to glutamate receptor signaling, MAP Kinase, apoptotic processes and neurotrophin pathways.

Discussion: Both methods achieved statistical relevant replication results (p<0.05, SuperExactTest), but only with one gene shared between WGCNA and NERI. In the MSET analysis, NERI was capable of achieving meaningful results for the methylation and *de novo* mutation databases; while our proposal of combining both results achieved better results for these two databases and additionally, for transcriptome (also increasing the number of candidate genes for each list).

Conclusion: Our study suggests that using both methods in combination could be a promising approach for establishing a group of modules and pathways related to schizophrenia (or any complex disease). Combining both methods results provided a meaningful outcome, resulting on genes related to different aspects of schizophrenia, such as immunological activity, glutamate and neurotrophin pathways.

References: [1] Langfelder P. et al. BMC Bioinformatics, 9:559, 2008; [2] Simões, SN. et al. BMC Bioinformatics, 16 Suppl. 19:S9, 2015; [3] Huang, DW. et al. Genome Biol. 8(9): R183, 2007; [4]; Wang, M. et al. Sci. Rep. 5: 16923, 2015; [5] Eisinger, BE. et al. BMC Neurosci. 14: 147, 2013.