# Module Detection Based on Significant Shortest Paths for the Characterization of Gene Expression Data

Daniele Pepe(✉)

I-BioStat, Hasselt University, Campus Diepenbeek, Hasselt, Belgium
daniele.pepe@uhasselt.be

**Abstract.** The characterization of diseases in terms of perturbed gene modules was recently introduced for the analysis of gene expression data. Some approaches were proposed in literature, but most of them are inductive approaches. This means that they try to infer key gene networks directly from data, ignoring the biological information available. Here a unique method for the detection of perturbed gene modules, based on the combination of data and hypothesis-driven approaches, is described. It relies upon biological metabolic pathways and significant shortest paths evaluated by structural equation modeling (SEM). The procedure was tested on a microarray experiment concerning tuberculosis (TB) disease. The validation of the final disease module was principally done by the Wang similarity semantic index and the Disease Ontology enrichment analysis. Finally, a topological analysis of the module via centrality measures and the identification of the cut vertices allowed to unveil important nodes in the disease module network. The results obtained were promising, as shown by the detection of key genes for the characterization of the studied disease.

**Keywords:** Disease module · Structural equation modeling · Gene expression data · Significant shortest paths

## 1 Introduction

The reductionist approach in medicine, based on the principle of "divide and conquer", although useful, present limits when it is necessary to explain the onset and progression of complex diseases. In fact, the approach is rooted in the assumption that if a complex problem is divided into more understandable and smaller units, then by their reconstruction, it is possible to unveil the studied complex problem.

For this reason, there are lists of genes associated with diseases. OMIM, a free database [1], for example, offers a catalogue of genes with the relative description of their role in the associated phenotypes. Conversely to this point of view, in the 1972, Anderson, in the article "More is complex" [2] affirms that the behaviour of large and complex aggregate of elementary particles cannot be understood in terms of a simple extrapolation of the properties of a few particle. At each level of complexity entirely new properties appear. For this reason, we are assisting to the passage from the reductionist approach to the systemic approach [3]. Most of biological networks are


A. Bracciali et al. (Eds.): CIBB 2016, LNBI 10477, pp. 1–11, 2017.
DOI: 10.1007/978-3-319-67834-4_1

subjected to specific laws [4], as the small world phenomena, which affirms that there are relatively short paths between any pair of nodes, the scale-free principle, with the consequence that there are few highly connected nodes; the local hypothesis i.e. the presence of modules, highly interlinked local regions in the network in which the components are involved in same biological processes.

The last property, modularity, is a general design principle in biological systems and has been observed also in transcriptional regulation networks [5]. In biology, modularity refers to a group of physically or functionally linked molecules (nodes) that work together to achieve a (relatively) distinct function.

Applying module level analysis should help to study biological systems at different levels and to understand which properties characterize the level of complexity considered. Many approaches exist that use a gene-module view as the basic building blocks of the analysis [6]. In general, it is used to divide the module identification in three main approaches: (1) network-based approach; (2) expression-based approach; (3) pathway-based approach [7]. The first approach is based on the topology of network, and modules are defined as subsets of vertices with high connectivity between them and less with external nodes. The second approach uses gene expression data for inferring modules of genes exhibiting similar expression by, for example, clustering methods. The third approach detects expression changes in biological pathways, group of genes that accomplishes specific biological functions.

The approach proposed in this paper, it is a mix and more general approach that takes advantage of the three approaches previously described. In fact, the pathway-based approach was used to detect perturbed KEGG pathways, then the network-approach was employed to identify the shortest paths between the differentially expressed genes (DEGs) and finally significant shortest paths (SSPs) were found using the expression data and structural equation modeling (SEM) [8].

The idea to consider shortest paths between DEGs to understand how they are connected is not new [9–11]. However, differently from the methods previously proposed, the key elements to test are constituted by shortest paths got from the network generated by the fusion of the relevant pathways. In this way, it is possible to consider the inter-pathway connectivity of DEGs by significant shortest paths tested by multiple group SEM. All the SSPs were joined to have the final perturbed disease module.

## 2 Materials and Methods

The classical differential gene expression analysis allows to identify DEGs. The differential analyses at gene level were performed by Significance Analysis of Microarray (SAM) [12], but any other procedure can be used. The next step is to find the network context where the DEGs act. The classical way is by pathway analysis also if it is not always able to detect the required information. In this situation, a solution could be take all the pathways containing at least one DEG. In the tuberculosis case the “Signaling Impact Pathway Analysis” (SPIA) was applied [13]. The corresponding perturbed KEGG pathways can be represented as mixed graphs, where the nodes represent genes and the edges represent multiple functional relationships between genes as activation, inhibition, binding etc. The core idea for building a disease module is to understand

how the DEGs, in the perturbed pathways, are connected between them. The first step for reaching this goal is to merge all the relevant pathways in a unique graph. The second step is to find the significant shortest paths that put in communication every couple of DEGs.

Each shortest path could be represented as a list of nodes P = ($p_i$, $p_{i+1}$,.., $p_{j-1}$, $p_j$) and a list of the corresponding edges E = ($e_{i(i+1)}$, …, $e_{(j-1)j}$) where ($p_i$, $p_j$) are DEGs and ($p_{i+1}$,..., $p_{j-1}$) can be DEGs or other microarray genes. In this analysis, every edge is directed. A shortest path can be codified as a structural equation (SE) model, in the following way:

$$P_j = \beta_{ji} P_i + E_j \tag{1}$$

where $P_j$ represents every gene in the path that is influenced directly by the gene $P_i$; $\beta_{ji}$ is the strength of relationship between node $P_i$ and $P_j$; $E_j$ is a term that represent external causes that have an effect on $P_j$ but not explicated in the model. Considering that the shortest paths selected are induced paths, every shortest path can be represented by j − 1 simple linear equations, where j is the number of nodes in the path.

For the estimation of the parameters $\beta_{ij}$, the Maximum Likelihood estimation (MLE) is used, assuming that all observed variables have a multinormal distribution. For finding the SSPs, the following omnibus test was performed:

$$H_0 : \sum\nolimits_1 (\theta) = \sum\nolimits_2 (\theta) \, vs. \, H_1 : \sum\nolimits_1 (\theta) \neq \sum\nolimits_2 (\theta) \tag{2}$$

where $\sum_1(\theta)$ and $\sum_2(\theta)$ are the model-implied covariance matrices of the groups one and two, and θ represents the parameter of the model. The test verifies if the difference between the model-implied covariance matrices of each group are statistically significant ($H_1$) or not ($H_0$) The statistical significance is determined by comparison of likelihood ratio test (LRT) chi-square ($\chi^2$diff) values at a given degree of freedom (d.f. ($\chi^2$diff)). If there is a significant difference ($P < 0.05$), after the Benjamin-Hochberg correction in the chi-squared goodness-of-fit index, the shortest path is considered statistically significant. All the SSPs were merged to obtain the final disease module. The final module is a weighted graph, where the weights correspond to the parameters estimated by SEM. The weights were used for the topological analysis, subsequently described. To validate the procedure two different approaches were employed: (1) enrichment analysis based on Disease Ontology (DO), to verify if the module genes are associated to the family of diseases connected with the analysed disease; (2) semantic similarity index, based on DO terms, between the list of genes associated "a priori" to the disease and the list of genes present in the module. DO creates a single structure for the classification of disease and permits to represent them in a relational ontology [14]. For the semantic similarity, the graph-based strategy proposed by Wang et al. [15] was applied. The similarity goes from 0, when the lists of genes are not associated, to 1, when the lists of genes contribute to the same DO terms. For finding the a priori genes, a search on Entrez Gene [16] was done. For each set of genes, an enrichment analysis on DO was done and the list of enriched terms was compared with those obtained by the genes in the module. Finally, a basic network analysis was performed based on

measures of centralities (betweenness and the Bonacich power centrality score) and connectivity (detection of articulation nodes). The betweenness centrality for the gene $v$ is defined as:

$$\sum_{i \neq j,\, i \neq v,\, j \neq v} g_{ivj} \backslash g_{ij} \tag{3}$$

where $g_{ivj}$ are the shortest paths between the node $i$ and $j$ in which the node $v$ is present, while $g_{ij}$ is the number of all shortest paths between the nodes $i$ and $j$. The normalized betweenness centrality is obtained:

$$B_n = 2B/\left(n^2 - 3n + 2\right) \tag{4}$$

where B is the raw betweenness and n the number of nodes. The Bonacich power measure corresponds to the notion that the power of a vertex is recursively defined by the sum of the power of its alters. The formula is the following:

$$C(\alpha, \beta) = \alpha (\boldsymbol{I} - \beta R)^{-1} R \boldsymbol{1} \tag{5}$$

where α is a scaling vector, β is an attenuation factor to weight the centrality of the nodes, R is the adjacency matrix, **I** is the identity matrix, **1** is a matrix of all ones. The articulation nodes are the minimum set of nodes which removal increases the number of connected components. They represent the nodes that allow to have a connected module.

The procedure was tested on the dataset GSE54992, where the group of samples of active tuberculosis (TB) (9 samples) were compared to healthy control (6 samples).

## 3 Results

SAM revealed 2152 significant genes using as delta value a value of 1.178 corresponding to a FDR value less than 0.05 and a minimum fold change of 2. On the 2152 DEGs, SPIA revealed important perturbed pathways (see Table 1), most of them associated to inflammation and infection as the cytokine-cytokine receptor interaction, the chemokine signaling pathway, the NF-κB signaling pathway, Legionellosis, Malaria. The first two pathways for example are induced in the lung in the response to TB infection to accumulate and mediate formation of granulomas, bacterial control and protection against the infection [17]. The pathways were transformed in graph and subsequently merged.

The next step was to find the shortest paths between every couple of DEGs on the total graph. The total number of shortest paths resulted of 1493 with 316 genes and 745 connections involved. For each shortest path a structural model was generated and tested to detect those significant. 260 out of 1493 were found relevant involving 206 genes and 330 connections. Figure 1 shows the graph obtained by the fusion of the 260 significant paths.

**Table 1.** Significant perturbed pathways for tuberculosis data by SPIA analysis.

| Name | pSize | NDE | pNDE | pPERT | pGFdr |
|---|---|---|---|---|---|
| Cytokine-cytokine receptor interaction | 241 | 66 | 0,000 | 0,000 | 0,000 |
| Chemokine signaling pathway | 177 | 42 | 0,000 | 0,000 | 0,000 |
| NF-kappa B signaling pathway | 75 | 23 | 0,000 | 0,021 | 0,000 |
| Osteoclast differentiation | 120 | 27 | 0,000 | 0,001 | 0,000 |
| Legionellosis | 39 | 16 | 0,000 | 0,287 | 0,000 |
| Complement and coagulation cascades | 65 | 22 | 0,000 | 0,882 | 0,000 |
| Staphylococcus aureus infection | 26 | 11 | 0,000 | 0,035 | 0,000 |
| Proteoglycans in cancer | 198 | 36 | 0,000 | 0,007 | 0,001 |
| Rheumatoid arthritis | 17 | 7 | 0,001 | 0,014 | 0,003 |
| Pathways in cancer | 308 | 49 | 0,001 | 0,011 | 0,003 |
| Inflammatory mediator regulation of TRP channels | 87 | 18 | 0,003 | 0,009 | 0,004 |
| Toxoplasmosis | 91 | 22 | 0,000 | 0,510 | 0,007 |
| Focal adhesion | 206 | 32 | 0,011 | 0,006 | 0,008 |
| MAPK signaling pathway | 248 | 40 | 0,003 | 0,051 | 0,014 |
| Viral myocarditis | 26 | 9 | 0,001 | 0,219 | 0,016 |
| Mineral absorption | 8 | 5 | 0,000 | 0,344 | 0,016 |
| Pertussis | 49 | 14 | 0,000 | 0,630 | 0,016 |
| Systemic lupus erythematosus | 12 | 4 | 0,028 | 0,007 | 0,016 |
| ECM-receptor interaction | 86 | 14 | 0,054 | 0,004 | 0,017 |
| Intestinal immune network for IgA production | 25 | 9 | 0,001 | 0,441 | 0,017 |
| Leishmaniasis | 47 | 12 | 0,002 | 0,115 | 0,018 |
| Influenza A | 105 | 21 | 0,002 | 0,143 | 0,019 |
| Amoebiasis | 45 | 13 | 0,000 | 0,740 | 0,019 |
| Rap1 signaling pathway | 204 | 35 | 0,002 | 0,228 | 0,021 |
| Toll-like receptor signaling pathway | 97 | 21 | 0,001 | 0,900 | 0,032 |
| Melanogenesis | 99 | 15 | 0,079 | 0,008 | 0,032 |
| Regulation of actin cytoskeleton | 182 | 30 | 0,006 | 0,121 | 0,034 |
| Malaria | 11 | 5 | 0,003 | 0,266 | 0,037 |
| Sphingolipid signaling pathway | 98 | 20 | 0,002 | 0,483 | 0,041 |

**pSize** = number of genes in the pathway; **NDE** = number of DEGs in the pathway; **pNDE** = p-value for the enrichment analysis; **pPERT** = p-value for the accumulated perturbation; **pGFdr** = combined p-value from the two previuos two p-value adjusted for the Fdr.

For the validation of the module two different analyses were performed: (1) one looks for diseases enrichment analysis; (2) the other measures the semantic similarity between the genes associated “a priori” with TB and those in the module. The results of the enrichment analysis on DO were very interesting considering that in the list there of the enriched diseases there was TB and other diseases that share common biological mechanisms as cancer, Salmonella infection and pertussis (see Fig. 2).

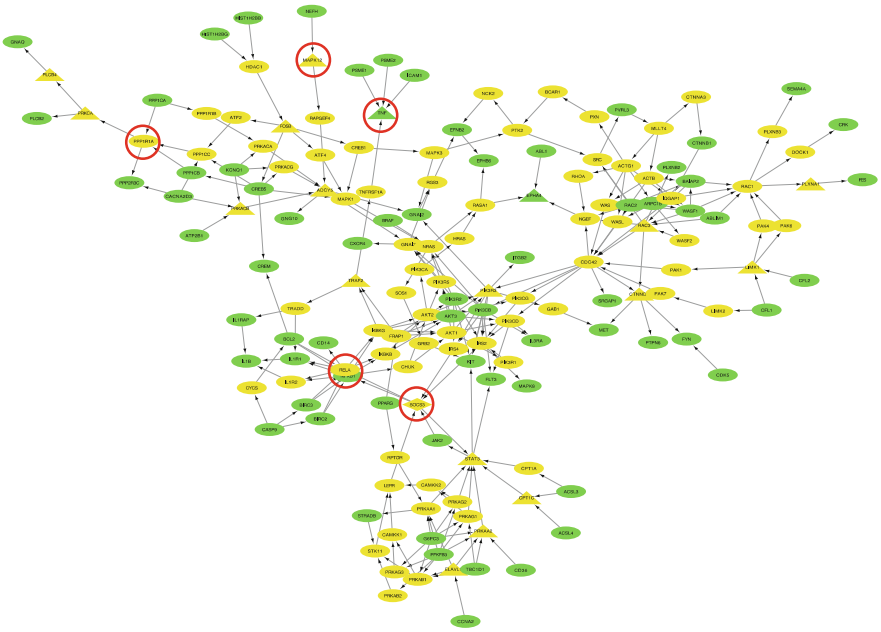

**Fig. 1.** TB disease module where the green nodes are the DEGs, the yellow ones the not DEGs that allows to the perturbation signal to propagate between the DEGs. Triangular nodes are the articulation point and the diamond node represents the gene SOCS3. Some key genes are highlighted with red circles. (Color figure online)

For the validation with the semantic similarity approach, the genes associated with TB in the database "Gene" of NCBI were detected. They were 312. The intersection of the genes in the module with those associated to TB revealed 24 genes in common. Of these, 13 were DEGs while the remaining genes were not DEGs. This result shows the limits of differential analysis when this stops to the detection of DEGs without understanding how these are connected. The semantic similarities between the two lists, based on DO terms was of 0.913. This means that the lists are highly related in terms of associated diseases. The following analyses were performed: (1) detection of the top 10 genes with the highest betweenness score and with the highest Bonacich's power centrality measure (Table 2); (2) detection of the cut vertices (Table 3). The highest values of normalized betweenness and Bonacich measure, on the weighted disease module, were associated to the gene SOCS3. This is very encouraging as it is known its fundamental role in immune responses to pathogens [18]. In fact, SOCS3 was considered as the most important family member for the association to autoimmunity, oncogenesis, diabetes and pathogenic immune evasion. It regulates both cytokine- and pathogen-induced cascades. Considering its importance, it was proposed as a therapeutic target [19]. Regarding the articulation point analysis, many interesting genes are present as the TNF, whose inhibition increases the risk of infections [20], as well as MAPK12 and some relevant kinases [21].

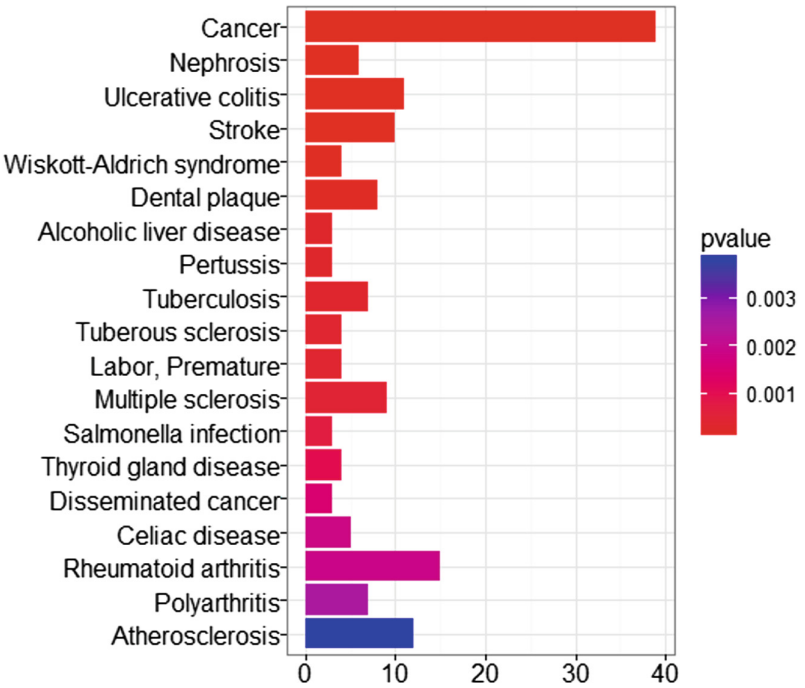


**Fig. 2.** Enriched diseases using the gene in the SSP module. Tuberculosis in one of the most enriched disease.

**Table 2.** Most relevant tuberculosis genes in the disease module based on betweeness (Betw) and Bonacich centrality measures.

| Entrez | Official name | Description | Centrality Measure | Centrality |
|---|---|---|---|---|
| 9021 | socs3 | suppressor of cytokine signaling 3 | 0,06 | Betw |
| 4790 | NFKB1 | nuclear factor kappa light polypeptide gene enhancer in B-cells 1 | 0,05 | Betw |
| 8517 | ikbkg | inhibitor of kappa light polypeptide gene enhancer in B-cells, kinase gamma | 0,05 | Betw |
| 998 | Cdc42 | cell division cycle 42 (GTP binding protein, 25 kDa); | 0,04 | Betw |
| 8660 | irs2 | insulin receptor substrate 2 | 0,03 | Betw |
| 4301 | mllt4 | myeloid/lymphoid or mixed-lineage leukemia | 0,02 | Betw |
| 5294 | PIK3CG | phosphoinositide-3-kinase, catalytic, gamma polypeptide | 0,02 | Betw |
| 8826 | IQGAP1 | IQ motif containing GTPase activating protein 1 | 0,02 | Betw |
| 25945 | PVRL3 | poliovirus receptor-related 3 | 0,02 | Betw |
| 6714 | Src | v-src sarcoma (Schmidt-Ruppin A-2) viral oncogene homolog | 0,02 | Betw |
| 9021 | socs3 | suppressor of cytokine signaling 3 | 20,25 | Bonacich |
| 4301 | mllt4 | myeloid/lymphoid or mixed-lineage leukemia | 14,24 | Bonacich |
| 5209 | pfkfb3 | 6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 3 | 12,24 | Bonacich |
| 10458 | BAIAP2 | BAI1-associated protein 2 | 10,91 | Bonacich |

(*continued*)

**Table 2.** (*continued*)

| Entrez | Official name | Description | Centrality Measure | Centrality |
|---|---|---|---|---|
| 2354 | fosB | FBJ murine osteosarcoma viral oncogene homolog B | 9,88 | Bonacich |
| 8659 | ALDH4A1 | aldehyde dehydrogenase 4 family, member A1 | 8,92 | Bonacich |
| 4790 | NFKB1 | nuclear factor kappa light polypeptide gene enhancer in B-cells 1 | 7,74 | Bonacich |
| 208 | akt2 | v-akt murine thymoma viral oncogene homolog 2 | 7,43 | Bonacich |
| 92579 | G6pc3 | glucose 6 phosphatase, catalytic, 3 | 6,31 | Bonacich |
| 8936 | WASF1 | WAS protein family, member 1 | 6,25 | Bonacich |

**Table 3.** Articulation points from the tuberculosis disease module.

| Entrez | Official name | Description |
|---|---|---|
| 6723 | SRM | spermidine synthase |
| 6774 | Stat3 | signal transducer and activator of transcription 3 |
| 126129 | Cpt1c | carnitine palmitoyltransferase 1C |
| 6300 | MAPK12 | mitogen-activated protein kinase 12 |
| 2582 | gale | UDP-galactose-4-epimerase |
| 5578 | Prkca | protein kinase C, alpha |
| 2354 | fosB | FBJ murine osteosarcoma viral oncogene homolog B |
| 111 | adcy5 | adenylate cyclase 5 |
| 5332 | Plcb4 | phospholipase C, beta 4 |
| 27165 | GLS2 | glutaminase 2 (liver, mitochondrial) |
| 7124 | TNF | tumor necrosis factor (TNF superfamily, member 2) |
| 217 | ALDH2 | aldehyde dehydrogenase 2 family (mitochondrial) |
| 51005 | AMDHD2 | amidohydrolase domain containing 2 |
| 1573 | cyp2j2 | cytochrome P450, family 2, subfamily J, polypeptide 2 |
| 2744 | GLS | glutaminase |
| 3984 | limk1 | LIM domain kinase 1 |
| 7186 | traF2 | TNF receptor-associated factor 2 |
| 7358 | UGDH | UDP-glucose dehydrogenase |
| 1500 | CTNND1 | catenin (cadherin-associated protein), delta 1 |
| 2673 | GFPT1 | glutamine-fructose-6-phosphate transaminase 1 |
| 5563 | prkaa2 | protein kinase, AMP-activated, alpha 2 catalytic subunit |
| 1994 | ELAVL1 | ELAV (embryonic lethal, abnormal vision, Drosophila)-like 1 |
| 4942 | OAT | ornithine aminotransferase (gyrate atrophy) |
| 8503 | PIK3R3 | phosphoinositide-3-kinase, regulatory subunit 3 (gamma) |
| 5881 | rac3 | ras-related C3 botulinum toxin substrate 3 |
| 5743 | PTGS2 | prostaglandin-endoperoxide synthase 2 |
| 2043 | EPHA4 | EPH receptor A4 |
| 5742 | Ptgs1 | prostaglandin-endoperoxide synthase 1 |
| 5361 | PLXNA1 | plexin A1 |

(*continued*)

**Table 3.** (*continued*)

| Entrez | Official name | Description |
|---|---|---|
| 5567 | PRKACB | protein kinase, cAMP-dependent, catalytic, beta |
| 8660 | irs2 | insulin receptor substrate 2 |
| 55577 | nagK | N-acetylglucosamine kinase |
| 2773 | GNAI3 | guanine nucleotide binding protein (G protein) |
| 26 | ABP1 | amiloride binding protein 1 |
| 5365 | PLXNB3 | plexin B3 |
| 1793 | DOCK1 | dedicator of cytokinesis 1 |
| 7132 | TNFRSF1A | tumor necrosis factor receptor superfamily, member 1A |
| 2805 | GOT1 | glutamic-oxaloacetic transaminase 1 |
| 8659 | ALDH4A1 | aldehyde dehydrogenase 4 family, member A1 |
| 4893 | NRAS | neuroblastoma RAS viral (v-ras) oncogene homolog |
| 5502 | ppp1r1a | protein phosphatase 1, regulatory (inhibitor) subunit 1A |
| 3065 | Hdac1 | histone deacetylase 1 |
| 2534 | FYN | FYN oncogene related to SRC, FGR, YES |
| 11069 | Rapgef4 | Rap guanine nucleotide exchange factor (GEF) 4 |
| 4790 | NFKB1 | nuclear factor of kappa light polypeptide gene enhancer in B-cells 1 |
| 5879 | rac1 | ras-related C3 botulinum toxin substrate 1 |
| 998 | Cdc42 | cell division cycle 42 |

## 4 Conclusion

In this paper, it was proposed a module level analysis for gene expression data that could take over the present methods for the identification of modules. It is used to divide the detection approaches in network-based, expression-based and pathway based. The approach here described is a mixed approach that starting from relevant pathways in which the DEGs are involved, detects the significant shortest paths by network, gene expression information and statistical analysis. The new concept is surely connected to the use of SEM for testing the significance of each shortest path model and the possibility to consider more pathways together, allowing to overcome the limiting idea of the pathway independence. Briefly, the pipeline consists in the following points: (1) discovering of DEGs associated to the disease; (2) understanding on which pathways the DEGs act; (3) joining in a unique graph all the relevant pathways; (4) performing the significant shortest path analysis for finding the disease module. The procedure was tested on a gene expression microarray concerning TB, but it can be applied to any gene expression experiment where the two-groups comparison is requested. The differential analysis of the shortest paths revealed significant shortest paths that characterize the experimental group on the control. The module obtained merging all the SSPs allowed to detect the key molecular network that could explain the disease. Very important genes were found as the SOCS3, TNF and MAPK2. The validation of the module by DO enrichment and similarity analysis has highlighted that

the genes in the modules are strictly associated to the a priori genes connected with the disease. In conclusion, the approach, is surely notable as new approach for downstream analysis of gene expression data. Future developments could be the application of the procedure to data from the integration of different NGS experiments.

**Funding acknowledgement.** This research was funded by the MIMOmics grant of the European Union's Seventh Framework Programme (FP7-Health-F5-2012) under the grant agreement number 305280.

## References

1. Hamosh, A., Scott, A.F., Amberger, J.S., Bocchini, C.A., McKusick, V.A.: Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. Nucleic Acids Res. **33**(suppl 1), D514–D517 (2005). doi:10.1093/nar/gki033
2. Anderson, P.W.: More is different. Science **177**(4047), 393–396 (1972). doi:10.1126/science.177.4047.393
3. Ahn, A.C., Tewari, M., Poon, C.S., Phillips, R.S.: The limits of reductionism in medicine: could systems biology offer an alternative? PLoS Med **3**(6), e208 (2006). doi:10.1371/journal.pmed0030208
4. Barabási, A.L., Gulbahce, N., Loscalzo, J.: Network medicine: a network-based approach to human disease. Nat. Rev. Genet. **12**(1), 56–68 (2011). doi:10.1038/nrg2918
5. Girvan, M., Newman, M.: Community structure in social and biological networks. Proc. Natl. Acad. Sci. **99**(12), 7821–7826 (2002). doi:10.1073/pnas.122653799
6. Segal, E., Friedman, N., Kaminski, N., Regev, A., Koller, D.: From signatures to models: understanding cancer using microarrays. Nat. Genet. **37**, S38–S45 (2005). doi:10.1038/ng1561
7. Wang, X., Dalkic, E., Wu, M., Chan, C.: Gene module level analysis: identification to networks and dynamics. Curr. Opin. Biotechnol. **19**(5), 482–491 (2008). doi:10.1016/j.copbio.2008.07.011
8. Kline, R.B.: Principles and Practice of Structural Equation Modeling. Guilford Press (2011). doi:10.1111/insr.12011_25
9. Pepe, D., Grassi, M.: Investigating perturbed pathway modules from gene expression data via structural equation models. BMC Bioinform. **15**(1), 1–15 (2014). doi:10.1186/1471-2105-15-132
10. Pepe, D., Hwan, D.J.: Estimation of dysregulated pathway regions in MPP+ treated human neuroblastoma SH-EP cells with structural equation model. BioChip J. **9**(2), 131–138 (2015). doi:10.1007/s13206-015-9206-3
11. Pepe, D., Hwan, D.J.: Comparison of perturbed pathways in two different cell models for Parkinson's Disease with structural equation model. J. Comput. Biol. **23**(2), 90–101 (2016). doi:10.1089/cmb.2015.0156
12. Tusher, V.G., Tibshirani, R., Chu, G.: Significance analysis of microarrays applied to the ionizing radiation response. Proc. Natl. Acad. Sci. **98**(9), 5116–5121 (2001). doi:10.1073/pnas.091062498
13. Tarca, A.L., Draghici, S., Khatri, P., Hassan, S.S., Mittal, P., Kim, J.S., Kim, C.J., Kusanovic, J.P., Romero, R.: A novel signaling pathway impact analysis. Bioinformatics **25**(1), 75–82 (2009). doi:10.1093/bioinformatics/btn577

14. Schriml, L.M., Arze, C., Nadendla, S., Chang, Y.W.W., Mazaitis, M., Felix, V., Feng, G., Kibbe, W.A.: Disease ontology: a backbone for disease semantic integration. Nucleic Acids Res. **40**(D1), D940–D946 (2012). doi:10.1093/nar/gkr972
15. Wang, J.Z., Du, Z., Payattakool, R., Philip, S.Y., Chen, C.F.: A new method to measure the semantic similarity of GO terms. Bioinformatics **23**(10), 1274–1281 (2007). doi:10.1093/bioinformatics/btm087
16. Maglott, D., Ostell, J., Pruitt, K.D., Tatusova, T.: Entrez Gene: gene-centered information at NCBI. Nucleic Acids Res. **39**(suppl 1), D52–D57 (2011). doi:10.1093/nar/gkq1237
17. Slight, S.R., Khader, S.A.: Chemokines shape the immune responses to tuberculosis. Cytokine Growth Factor Rev. **24**(2), 105–113 (2013). doi:10.1016/j.cytogfr.2012.10.002
18. Carow, B., Reuschl, A.K., Gavier-Widén, D., Jenkins, B.J., Ernst, M., Yoshimura, A., Chambers, B.J., Rottenberg, M.E.: Critical and independent role for SOCS3 in either myeloid or T cells in resistance to Mycobacterium tuberculosis. PLoS Pathog. **9**(7), e1003442 (2013). doi:10.1371/journal.ppat.1003442
19. Mahony, R.A., Diskin, C., Stevenson, N.J.: SOCS3 revisited: a broad regulator of disease, now ready for therapeutic use? Cell. Molecular Life Sci. **1**(1), 1–14 (2016). doi:10.1007/s00018-016-2234-x
20. Sichletidis, L., Settas, L., Spyratos, D., Chloros, D., Patakas, D.: Tuberculosis in patients receiving anti-TNF agents despite chemoprophylaxis. Int. J. Tuberc. Lung Dis. **10**(10), 1127–1132 (2006)
21. Song, C.H., Lee, J.S., Lee, S.H., Lim, K., Kim, H.J., Park, J.K., Paik, T.H., Jo, E.K.: Role of mitogen-activated protein kinase pathways in the production of tumor necrosis factor-α, interleukin-10, and monocyte chemotactic protein-1 by Mycobacterium tuberculosis H37Rv-infected human monocytes. J. Clin. Immunol. **23**(3), 194–201 (2003)