

# A Leader Genes Approach-based Tool for Molecular Genomics: From Gene-ranking to Gene-network Systems Biology and Biotargets Predictions

Nicola Luigi Bragazzi<sup>1,2</sup> and Claudio Nicolini<sup>1,2,3\*</sup>

<sup>1</sup>Laboratories of Biophysics and Nanobiotechnology, Department of Experimental Medicine (DIMES), University of Genova, Via Antonio Pastore 3, 16132 Genova, Italy

<sup>2</sup>Nanoworld Institute, Fondazione EL.B.A. Nicolini, Largo Redaelli 7, Pradalunga, Bergamo, Italy

<sup>3</sup>Biodesign Institute, Arizona State University, Tempe, USA

## Abstract

Nanogenomics, being the interplay of nanobiotechnologies and bioinformatics, is emerging as an intriguing approach in the field of nowadays biomedicine. Microarrays can produce a wealth of data and details, but they need an algorithm for data reduction to be clearly understood and exploited. The Leader Genes approach, integrating the different available databases and genomics tools, enables the user to search for genes linked to a disease or a cellular process, and to visualize the class of the most important genes, that is to say those having the highest number of interconnections. In this manuscript, we will review the algorithm which has been validated with both experimental and clinical studies. We will describe the different steps that lead to its final version, and we will discuss future perspectives and developments.

**Keywords:** Bioinformatics; Gene prioritization; Gene-ranking; Leader gene approach; Molecular genomics; Nanogenomics; Systems biology

## Introduction

Nanogenomics [1-3] is emerging as a promising discipline in the field of biomedicine, being the interplay of nanobiotechnologies and bioinformatics. Gene microarrays and the other biomolecular high-throughput tools of the omics sciences (such as linkage and linkage disequilibrium, GWAS or genome-wide association study, QTL or quantitative trait loci and its derivative, such as eQTL, expression QTL, phenome profile, transcriptome analysis for genome mapping, and so on) [4], can produce an incredible wealth of data. If on one hand, these “big data” [5,6] are necessary for a deep and profound underpinning of the entire molecular picture, which is extremely complex, on the other hand, researchers risk to be overwhelmed by these data, without a proper data reduction.

For this reason, gene prioritization or gene-ranking has been advocated as a crucial task in bioinformatics, leading to a list of few strong candidate genes that can be further validated with ad hoc experiments. Elucidating genetic traits is particularly useful when studying complex and multi-factorial diseases [7,8], such as Alzheimer’s dementia and other neurological pathologies [9], diabetes and other metabolic disorders [10], cancer [11], and cardiovascular disorders [12]. Usually these diseases are chronic, present a relatively mild phenotype, are slowly progressive, and have a tremendous burden for the society and impact on the population in term of quality of life (QOL). The physio-pathology of complex pathologies is characterized by various biologic pathways and networks, dependent upon the contribution of a large number of genes and gene products, and therefore, the knowledge of molecular mechanisms of complex multi-factorial diseases must include and deal with a large array of genes. These genes form complex networks of interactions, which may be direct (that is to say, physical interactions between the proteins, confirmed by experimental techniques, such as NMR or crystallography), or indirect (involvement in the same metabolic pathway or co-expression under different conditions).

There are a lot of gene-ranking methods, which for example,

exploit the Pearson correlation with clinical information (such as Kaplan-Meier survival time) ([4] and references therein, [13]), or are based upon gene expression time series or gene co-expression series [14], regulatory networks (protein-protein interaction, PPI [15-18], gene-protein interaction, or gene-gene interaction maps [19]) (for further explanations and references, the reader is referred [20,21], ortholog mapping, biomedical literature mining (ab initio approach), or integrating and combining the above mentioned approaches [4]. The techniques that are usually employed are: support vector machine, neural networks, fuzzy algorithms, kernelized score functions [22], and other learning machine procedures, biomedical annotation [23] (such as LSI, Latent Semantic Indexing [24], PosMed or Positional PubMed [25]), and clustering [26,27].

The main purposes of gene-ranking [28] are to provide robust molecular signatures of the studied events, to discover reliable biomarkers, as well as to predict gene function (GFP), or further biological targets of potential pharmaceutical interest [29].

In this manuscript, we will review our algorithm, which has been validated with both experimental and clinical studies. We will describe the different steps that lead to its final version, and we will discuss future perspectives and developments.

## Materials and Methods

The Leader Gene approach exploits a search-dependent statistical

**\*Corresponding author:** Claudio Nicolini, <sup>1</sup>Laboratories of Biophysics and Nanobiotechnology, Department of Experimental Medicine (DIMES), University of Genova, Via Antonio Pastore 3, 16132 Genova, Italy, Tel: +39 010 353 38217; Fax: +39 010 353 38215; E-mail: [info@fondazioneelba-nicolini.org](mailto:info@fondazioneelba-nicolini.org)

Received June 30, 2013; Accepted July 22, 2013; Published July 25, 2013

**Citation:** Bragazzi NL, Nicolini C (2013) A Leader Genes Approach-based Tool for Molecular Genomics: From Gene-ranking to Gene-network Systems Biology and Biotargets Predictions. J Comput Sci Syst Biol 6: 165-176. doi:10.4172/jcsb.1000113

**Copyright:** © 2013 Bragazzi NL, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

algorithm, which is based on the systematic, exhaustive and recursive search for the genes and their products involved in a given process or event (biological, such as a biochemical or biophysical reaction, or pathological, like a disorder), on the computation of their interaction map, and on their ranking according to the number of all experimentally established interactions, as derived from free and open source Web-available databases, such as STRING (Search Tool for the Retrieval of Interacting Genes, Heidelberg, Germany) [30].

The systematic search is usually carried out mining different databases (like PubMed using National Library of Medicine vocabulary of terms or MESH terms, NCBI OMIM, GeneBank, GeneCards, GeneAtlas, using the standardized nomenclature HUGO), and repositories containing DNA microarrays data (normalized with any normalization technique, such as LOWESS, or locally weighted scatterplot smoothing, LOESS, or local regression), integrating the obtained hits.

The calculation of an interaction map of the obtained list of genes is computed using the weighted number of links (WNLs). This measure is calculated for each gene using the program STRING [30], and this value is derived from the weighed sum of three types of interactions:

1. literature co-occurrence of the names of genes, and/or their products in abstracts of papers available on the Internet. The scores assigned are derived from benchmarked and validated scoring system based on the frequencies and distributions of gene/gene products names in the aforementioned abstracts. The benchmarks themselves are set from manual evaluation of predictions of gene and protein interactions by experts, and are typically below 0.5;
2. scores derived from databases dedicated to gene networks, containing data on induction and expression of a particular genes by other genes derived from microarray experiments, or other high-throughput omics techniques. The score of 1 is assigned if the link is already present in the database, while putative links have lower values (typically 0.6–0.8);
3. the same range of scores is assigned to gene interactions *via* physically observed interactions between proteins. The software used does not discriminate between *in vivo* or *in vitro* experiment derived data. Generally the scores are close to those of interaction type 2, but links of this type occur much rarely than of type 2.

The combined association scores  $S_{ij}$  were summed for each gene  $i$  over its neighbors ( $i,j$ ), giving the final weighted number of links for the gene  $i$ . Further, we applied clustering methods to the weighted number of links in order to identify the group of leader genes. Cluster analysis, also called segmentation analysis or taxonomy analysis, is a way to partition a set of objects into homogeneous and separated groups or clusters, in such a way that the profiles of objects in the same cluster are very similar and the profiles of objects in different clusters are quite distinct. In particular, cluster analysis can be defined as follows:

Given a set  $S$  of  $n$  objects  $\{x_1, x_2, \dots, x_n\}$ , where each object is described by  $m$  attributes  $x_i = (x_{i1}, x_{i2}, \dots, x_{im})$ , determine a classification that is most likely to have generated the observed objects.

Genes belonging to the highest rank are defined as “leader genes” or “hub genes” because they may be assumed to play an important role in the analysed processes. The “Leader Gene approach” can suggest a list of few, but strong candidate genes potentially relevant

within a given cellular process or a pathology, according to the already available experimental data. Moreover, the interaction map among all the genes involved in the same process may be useful in interpreting the experimental and clinical results, and in planning new targeted experimentation. Interestingly, such experimentation may be simpler to be analysed than mass-scale molecular genomics, whose wealth of details may raise problems and complications [1,2]. This computational method gave promising results, when applied to the human T lymphocyte cell cycle, human kidney transplant, oral lichen planus and periodontitis. These results were also integrated with a targeted experimental analysis, to draw an overall picture of these processes, and are reviewed in the following paragraphs.

This interactive, automatic and user-friendly stand-alone tool has been written in house in Java, Javascript, PHP and HTML. The completely automated pipeline is performed *via* NCBI e-utilities (e-search, e-fetch, for further information the author is referred to the NCBI site), and other similar facilities.

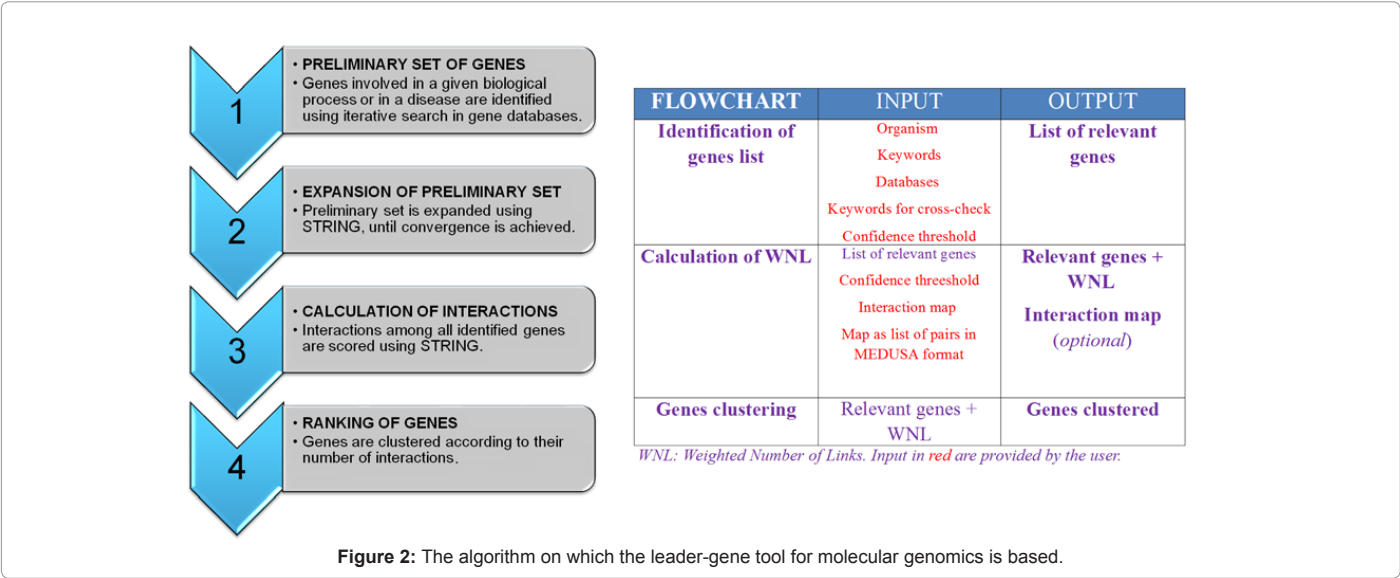
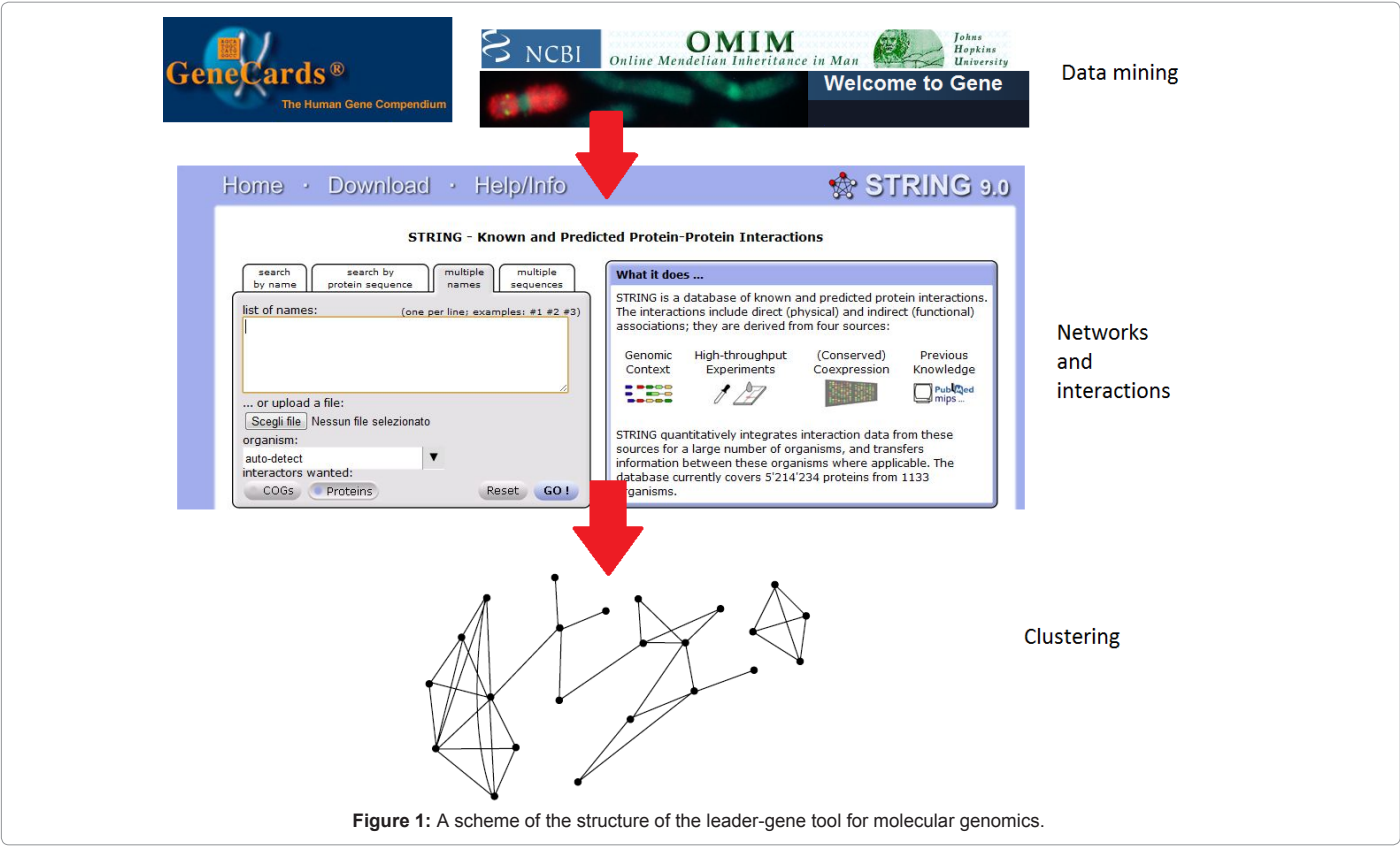
A pictorial scheme of this tool is given in Figure 1, while the algorithm is shown in Figure 2.

The clustering techniques the user can choose are: Clustering K-means and Chinese whispers (which has been thought specifically for graph clustering, [31]); as far as the number of clusters is concerned, the user can choose from heuristic number or provided by the user himself. A screen-shot of the output is given in Figure 3.

## Results

### Human T-lymphocytes cell cycle

Quiescent non-activated peripheral blood T cells, when stimulated by antigen-presenting cells (APCs), enter the cell cycle *via* the CD3/CD28 pathway. T cells cycle is a key biological event leading to proliferation, increase in size and production of molecules, such as cell surface receptors and cytokines/interleukins. Moreover, impairment and dysregulation of T cells cycle lead to several important diseases. The key genes involved in the cell cycle of human T lymphocytes stimulated by the mitogen compound phytohemagglutinin or PHA [32–34], were identified by iterative and recursive searches of different genomic databases, as well as derived also from repositories containing DNA microarray experiments. This systematic search resulted in the identification of 238 related genes. Later, after this step, software and tools predicting interactions among those genes were employed, assigning scores to each of the genes, according to the number of interactions for each gene weighted by the significance of each interaction (WNLs), and finally applying several types of clustering algorithms to the genes. All the clustering algorithms and techniques applied, either hierarchical or K-means, invariably selected the same six “leader” genes involved in controlling the cell cycle of human T lymphocytes (namely, CDK2, CDK4, MYC, CDC2, CDKN1A, CDKN1B). CDK2 (Cyclin-Dependent Kinase 2) is involved at the transitions from G1 to S binding Cyclin A and Cyclin E, while CDK4 at the progression in G1 phase. MYC is a gene involved in the early phase of T cells proliferation, controlling G1 phase. CDC2 (Cell Division Control 2, also known as CDK1) is a phosphoprotein belonging to the Ser/Thr kinases, and is also a catalytic subunit of the Maturation-Promoting Factor (MPF) complex. It is responsible of the transition from G1 to S and G2 to M phases, binding both Cyclin A and Cyclin B. CDKN1A (Cyclin-Dependent Kinase inhibitor 1A or CDK-interacting protein 1A, also known as p21Cip1/WAF) and CDKN1B (Cyclin-Dependent Kinase inhibitor 1B, also known as p27Kip1) are inhibitors



of cyclin-CDK2 and -CDK4 dependent complexes, blocking Cyclin D and Cyclin E, and thereby, contribute to the control of G1/S transition and of G1 progression

Moreover, human lymphocytes gene expression was monitored before and after PHA stimulation after 48 hours in an experiment, and over 72 hours in an experiment, using a commercial Human Starter DNA microarray technology, and an in-house instrumentation (DNASER). Results were then compared with our previous bioinformatics predictions, and we found that experimental data were comparable with our bioinformatics findings. We studied and investigated in details the specific roles of PCNA (Proliferating Cell Nuclear Antigen) and CCNA2 (Cyclin A2) at all phases of the cell cycle, and of CHEK1 (Checkpoint Kinase 1) in regulating DNA repair and preservation, and their interactions with the leader genes.

**Periodontitis**

Periodontitis [35] is a complex, multifactorial, chronic and



progressive inflammatory disease affecting the periodontium, that is to say the surrounding and supporting tissues of the teeth. These tissues include the gum (or gingiva), the periodontal ligament, the cementum and the alveolar bone. Because of bacterial infection caused by microbes such as *Porphyromonas gingivalis*, *Prevotella intermedia*, *Bacteroides forsythus*, *Aggregatibacter actinomycetemcomitans*, *Fusobacterium nucleatum*, *Campylobacter species*, *Treponema denticola*, and *Campylobacter rectus*, the periodontal tissues become inflamed, and are slowly destroyed by the action of the inflammatory process. If left untreated, the teeth lose their ligamentous support to the alveolar bone, become mobile, and are eventually lost. A large majority of periodontitis patients respond well to conventional therapies. However, the results of some studies have suggested that a small percentage of patients present a poor response. Those patients are defined as “downhill patients” (loss of 4 up to 9 teeth), or as “extremely downhill patients” (loss of 12 up to 23 teeth).

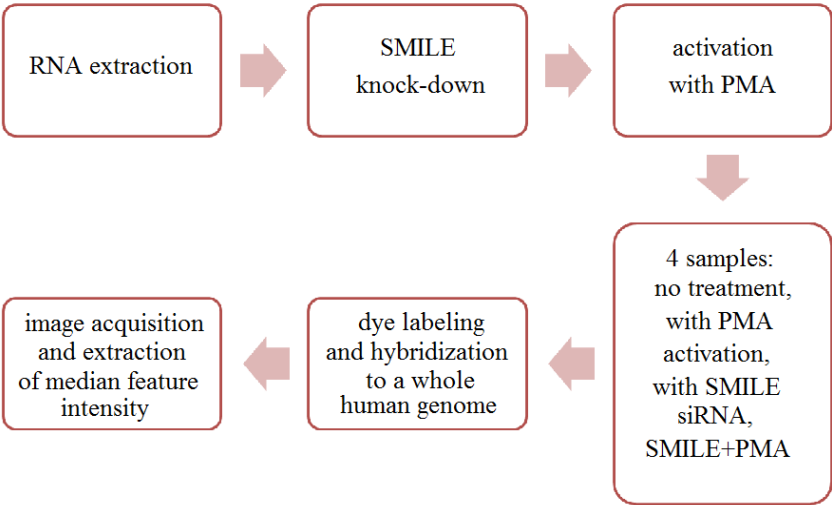
Genes involved in human periodontitis were tentatively identified and ranked according to their number of interactions to obtain a preliminary, broader view of molecular mechanisms of periodontitis and plan targeted experimentation. Genes were identified with interrelated queries of several databases. The interactions among these genes were mapped and given a significance score. The weighted number of links was calculated for each gene. Genes were clustered according to this parameter. The genes in the highest cluster were termed leader genes. Sixty-one genes involved or potentially involved in periodontitis were identified. Only five were identified as leader genes (namely, NFKB1, CBL, GRB2, PIK3R1, RELA), whereas 12 others were ranked in an immediately lower cluster (namely, IKKB, SRC, IL6R, TRAF2, TNF, PDGFRB, IL6ST, IL4R, IL6, TNFRSF1A, CRK, IL1R1). For 10 of 17 genes (58.8%), there is evidence of involvement in periodontitis; seven new genes that are potentially involved in this disease were identified. The involvement in periodontitis has been completely established for only two leader genes.

## Kidney transplant

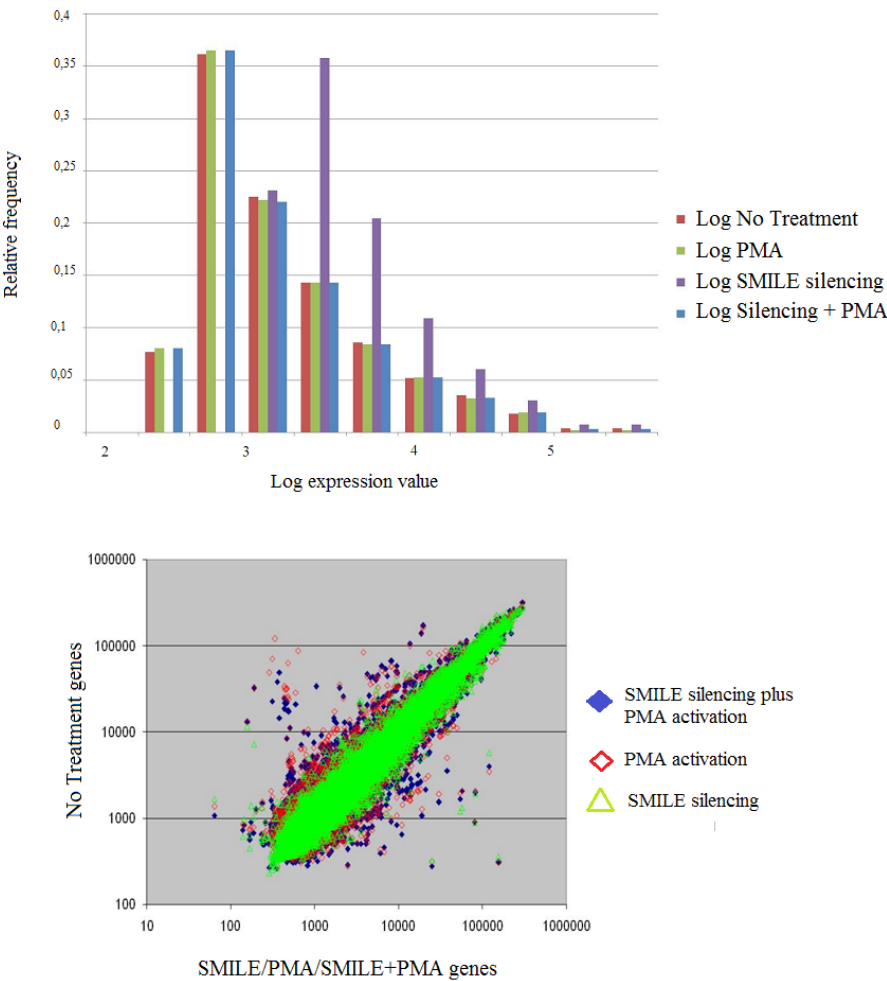
A good rate of survival in patients receiving solid organ transplantation, which is the main objective of the research in the field, depends on life-long immunosuppression (such as corticosteroids and other immunosuppressive drugs), which may result in increased figures of infection and tumors. Induction of tolerance to allograft would represent the optimal solution for controlling both chronic rejection (CR), and side effects of immunosuppression. Although spontaneous “operational tolerance” can occur in human kidney or liver transplantation, the lack of noninvasive peripheral blood biological markers of this rare phenomenon precludes the identification of potentially tolerant patients in whom immunosuppression could be tapered, as well as the development of new tolerance inducing strategies. Here, the potential of high throughput microarray technology to decipher complex pathologies allowed us to study the peripheral blood specific gene expression profile, and corresponding Expression Analysis Systematic Explorer (EASE) molecular pathways associated to operational tolerance in a cohort of human kidney graft recipients. In comparison with patients with CR, tolerant patients displayed a set of 343 differentially expressed genes, mainly immune and defense related genes, in their peripheral blood mononuclear cells (PBMC), of which 223 were also different from healthy volunteers. Using the expression pattern of these 343 genes, we were able to classify correctly >80% of the patients in a cross-validation experiment, and classified correctly all of the samples throughout the years. Collectively, this study identifies a unique PBMC gene signature associated with human operational tolerance in kidney transplantation, using both by a classical statistical microarray analysis [36], and a nonstatistical analysis [37]. The findings obtained [37] were replicated using a murine model [38], which confirmed the importance of AKT pathway in the tolerance network.

Moreover, recent findings [39], carried out following the protocol shown in Figure 4 indicated that the SMILE/TMTC3 gene may be





**Figure 4:** Protocol of SMILE experiment.



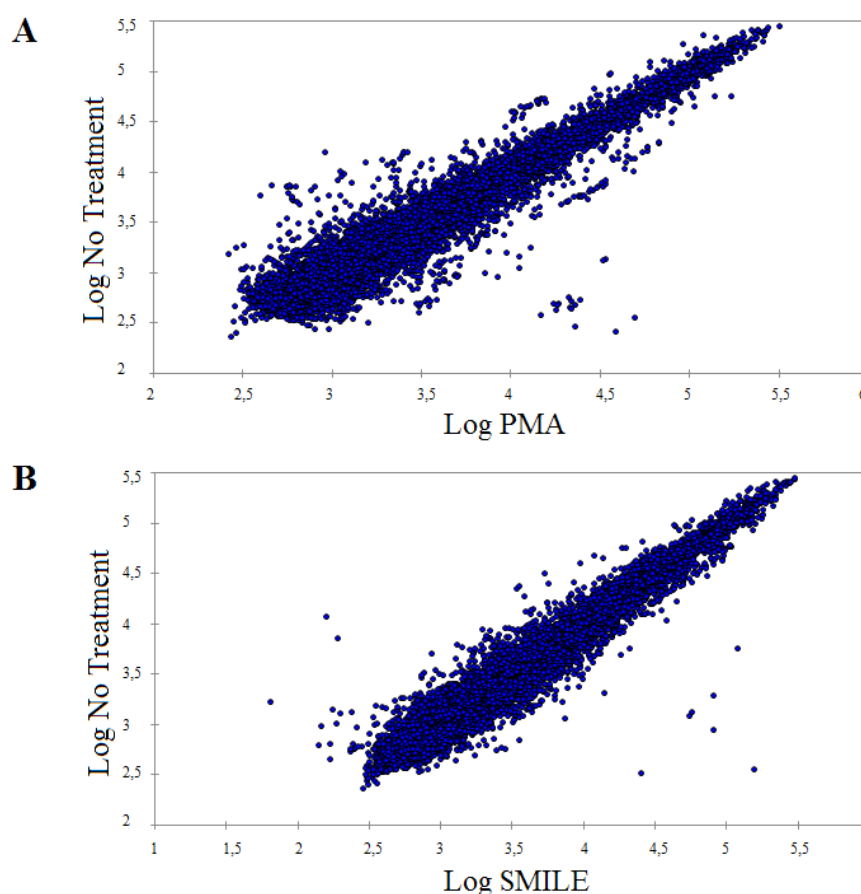
**Figure 5:** LOWESS normalized and log-transformed gene expression values for the SMILE knock-down experiment: no treatment group, PMA activation group, SMILE silencing group and SMILE knock-down plus PMA activation group (up) and scatter plot of PMA activation, SMILE silencing and SMILE knock-down plus PMA activation genes versus no treatment genes (bottom).

involved in kidney graft operational tolerance in human. This gene was found to be up-regulated in blood from patients with a well functioning kidney transplant, in the absence of immunosuppression compared to other transplanted recipients with clinically different status. A microarray study of SMILE knock-down and phorbol 12-myristate 13-acetate (PMA) activation in HeLa cells (Figures 5-7) was herein compared to our earlier analysis based on microarray data of kidney allograft tolerance (Figure 8 and 10) and rejection (Figure 9 and 11) in humans, and in a rat model of allograft transplantation (Figure 12 and 13), to determine possible new genes and gene networks involved in kidney transplantation. The nearest neighbors at the intersection of the SMILE knock-down network with the human tolerance/rejection networks are shown to be NPHS1 and ARRB2, the former (Nephrin) being involved in kidney podocyte function, and the decrease of the latter (Arrestin  $\beta$ 2) being recently shown to be involved in monocyte activation during acute kidney allograft rejection in rat. Moreover, another one of the neighbors at the intersection of SMILE network and tolerance/rejection networks is XBP-1 (X-box binding protein 1), that we report previously to be increased, at a transcript level, after ER stress in SMILE silenced cells (Figure 14). Finally, we were also able to show that topological properties (both local and global) of joint SMILE knock-down network-tolerance/rejection networks and joint PMA activation network-tolerance/rejection networks in rat and human are essentially different, likely due to the inherent nature of the gene SMILE, and the mitogen compound PMA, that do not act the same

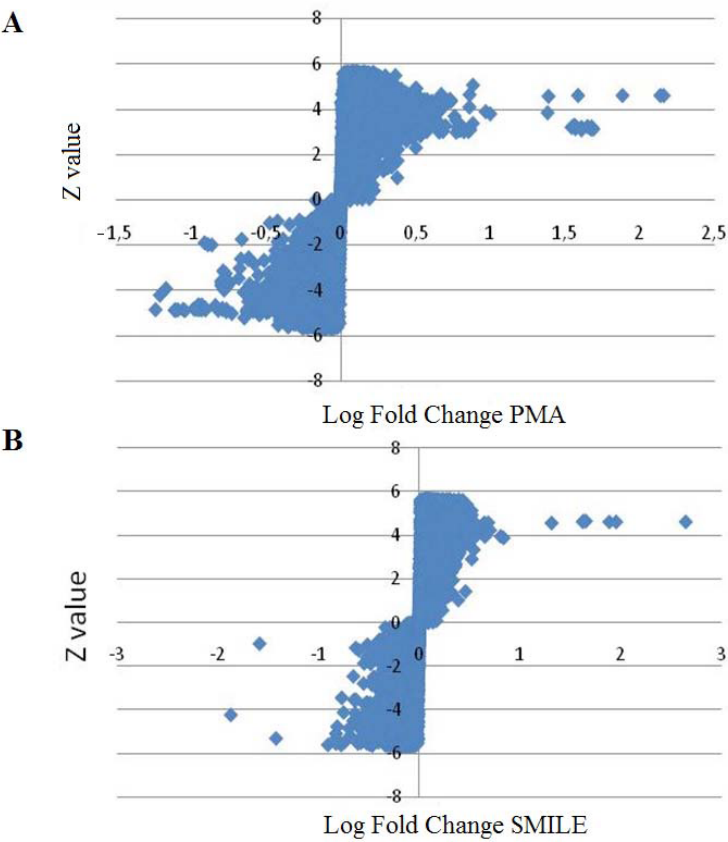
way on genes and do not interfere the same way on networks. We also showed that interestingly SMILE networks contain more feed-forward loop (FFL) motifs, and thus, SMILE calls for a more fine-tuned genetic regulation.

### Oral lichen planus

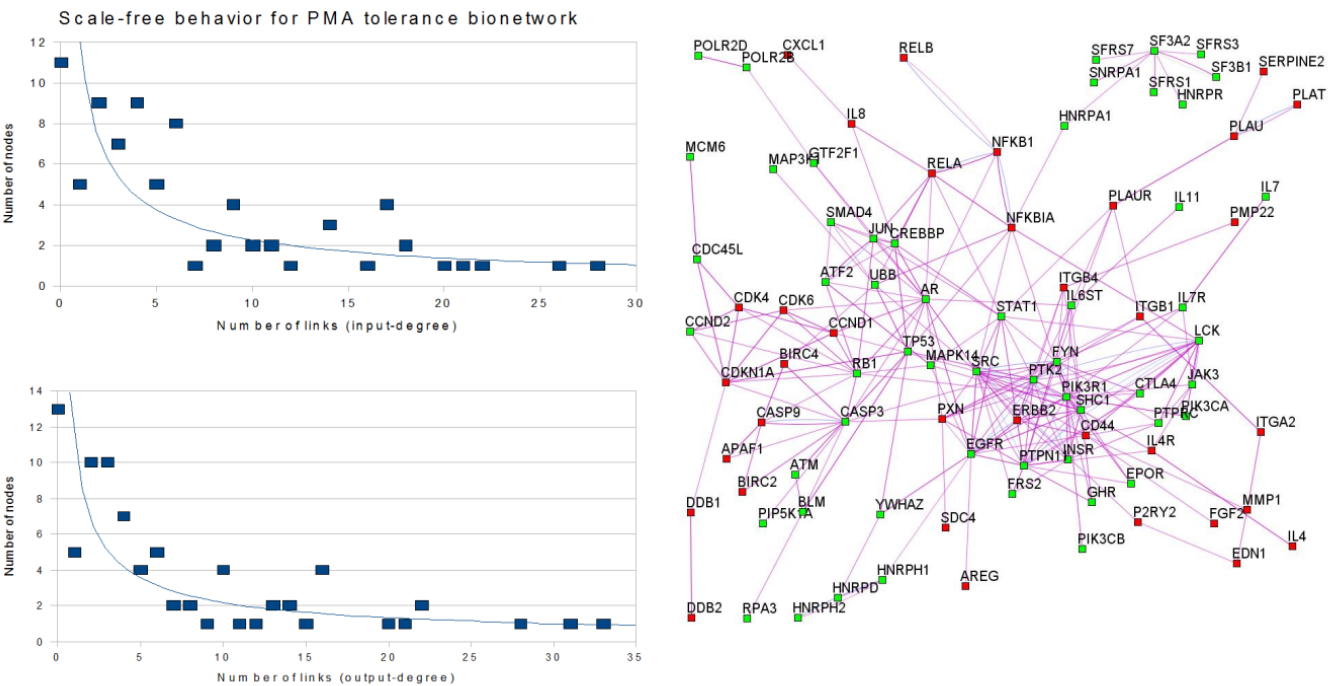
Oral Lichen Planus (OLP) is a chronic inflammatory oral mucosal disease that has an important global burden, both epidemiological and economic, affecting about 0.5-2% of the world population. In the *ab initio* theoretical study, genes involved in human OLP pathogenesis [40] are identified and ranked according to their number of interactions, in order to obtain a broader view of its molecular mechanisms, and to plan targeted experiments. Genes involved or potentially involved in OLP were identified by systematically querying several databases, until the identification of a final set of genes. Interactions among these genes were mapped and given a significance score using STRING database. For each gene, significance scores were summed to obtain a weighted number of links (WNL), and subsequently, genes were clustered according to this parameter. The genes in the highest cluster were termed as leader genes; the other ones were ranked as class B genes, class C genes, and so on. This study was complemented by a topological analysis of the network, carried out using Cytoscape, BinGO and FANMOD software. The interactions in the obtained network showed power law behaviour, in agreement with the scale-free topology theory of the biological graphs.



**Figure 6:** Scatter plot of PMA activation log-transformed genes values (a) and scatter plot of SMILE knock-down log-transformed genes values (b).



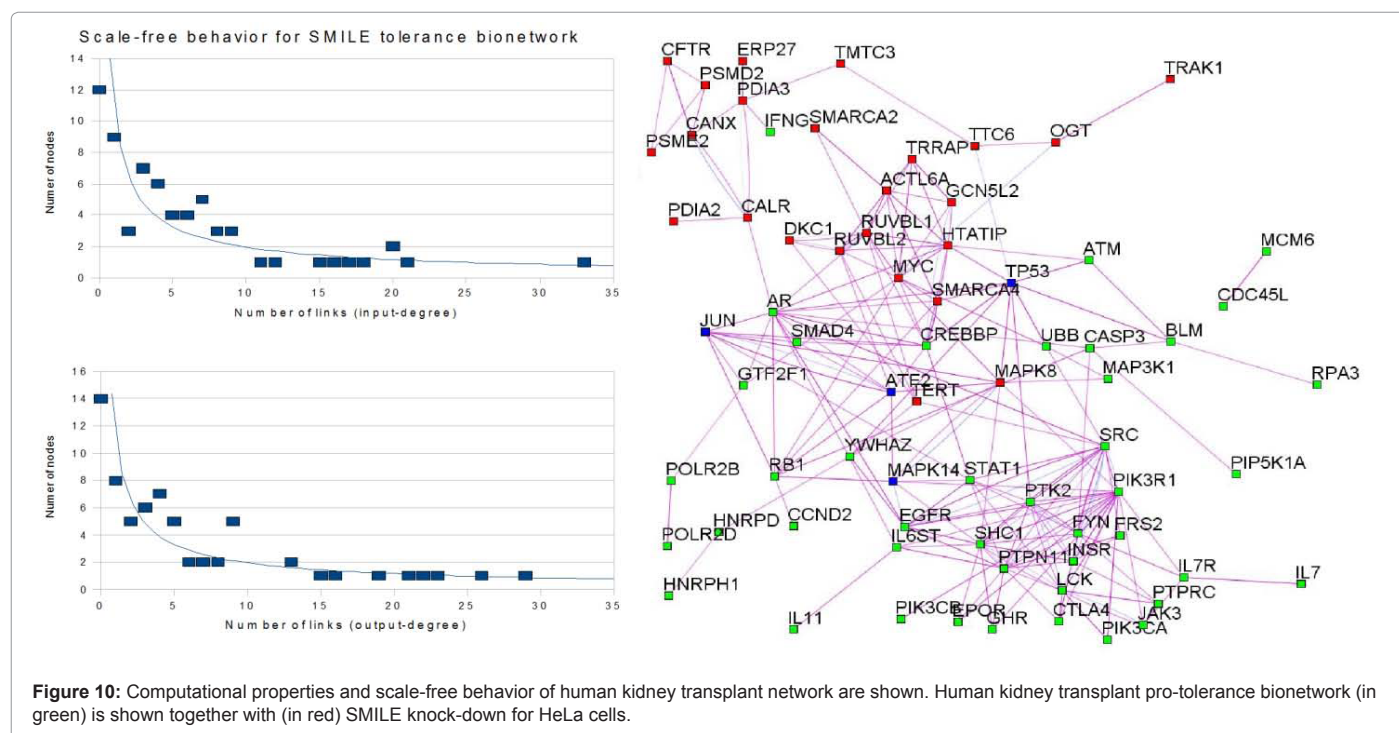
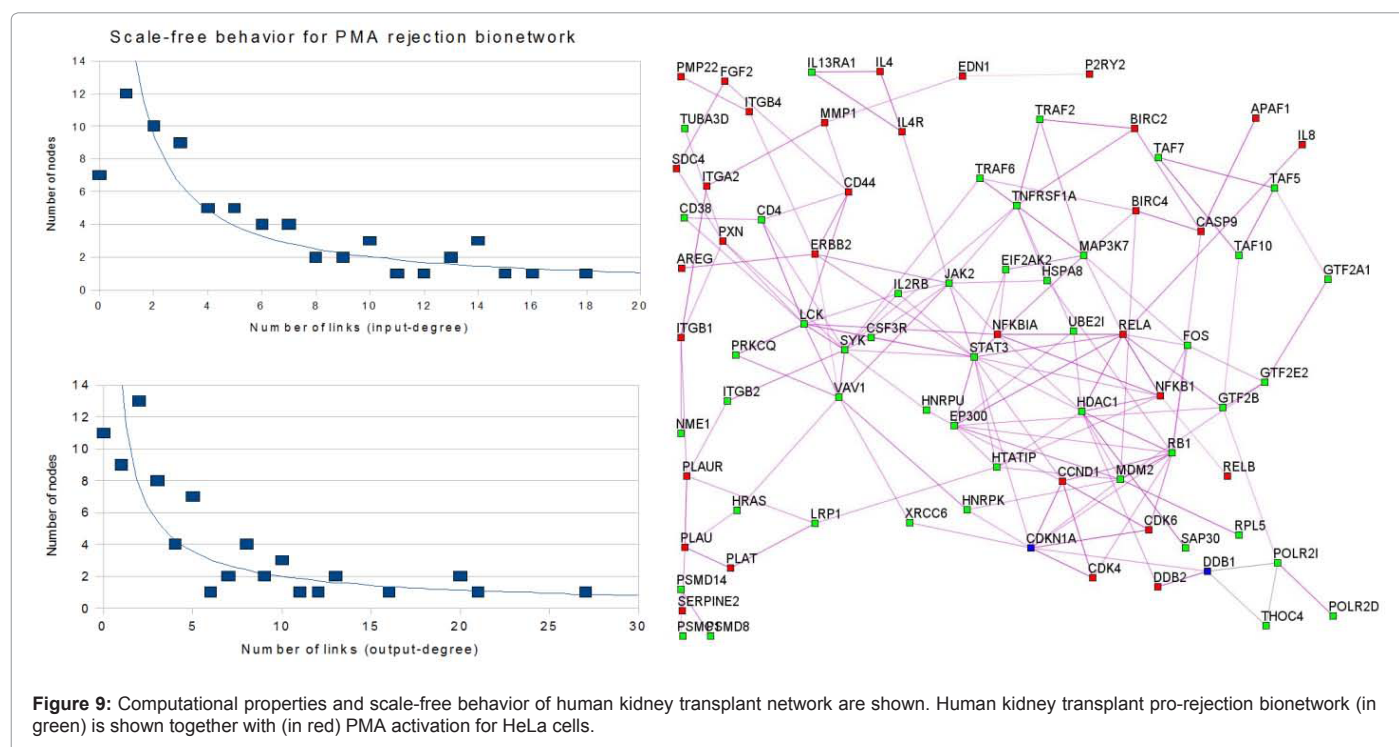
**Figure 7:** Volcano plot for PMA activation (a) and Volcano plot for SMILE knock-down (b). These graphs can help in visualizing differentially expressed genes.



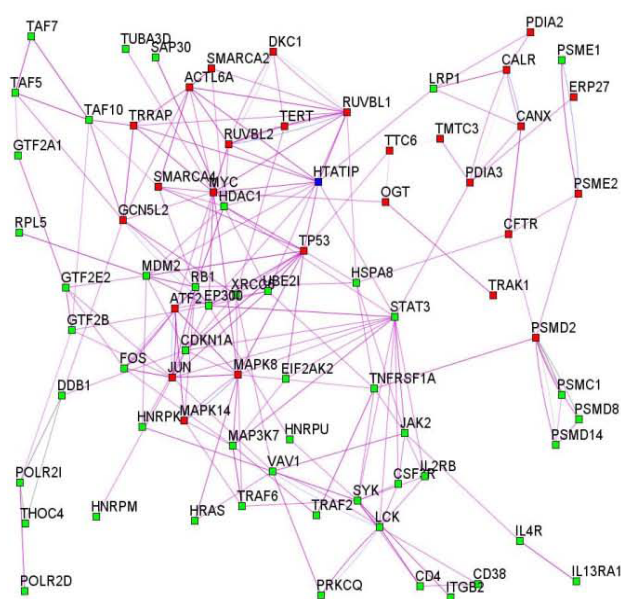
**Figure 8:** Computational properties and scale-free behavior of human kidney transplant network are shown. Human kidney transplant pro-tolerance bionetwork (in green) is shown together with (in red) PMA activation for HeLa cells.

132 genes were identified and five of them (namely, JUN, EGFR, FOS, IL2, ITGB4) were classified as leaders. JUN encodes a protein which interacts directly with specific target DNA sequences to regulate gene expression. This gene is mapped to 1p32-p31, a chromosomal region involved in both translocations and deletions in human malignancies. EGFR is a transmembrane glycoprotein that is a member of the protein kinase superfamily, and is a cell surface protein

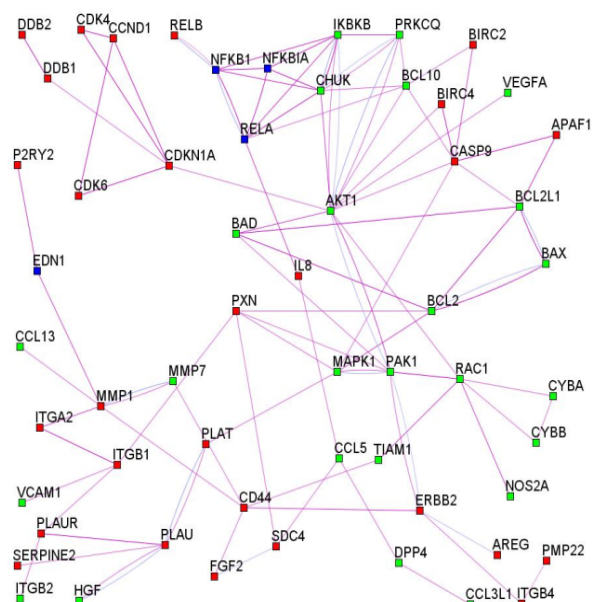
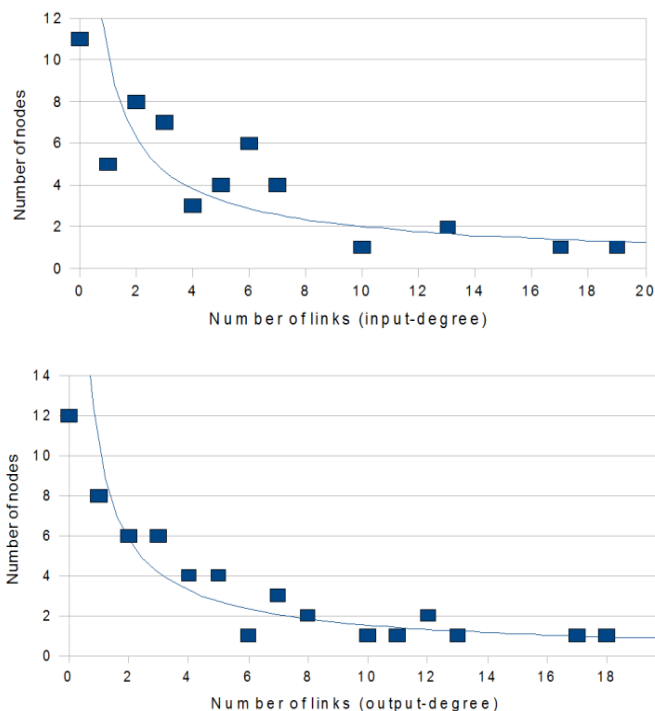
that binds to epidermal growth factor. Binding of the protein to a ligand leads to cell proliferation. FOS encodes leucine zipper proteins that can dimerize with proteins of the JUN family, thereby forming the transcription factor complex AP-1. FOS has been implicated as regulators of cell proliferation, differentiation and transformation. In some cases, expression of the FOS gene has also been associated with apoptotic cell death. IL2 is a secreted cytokine important for the



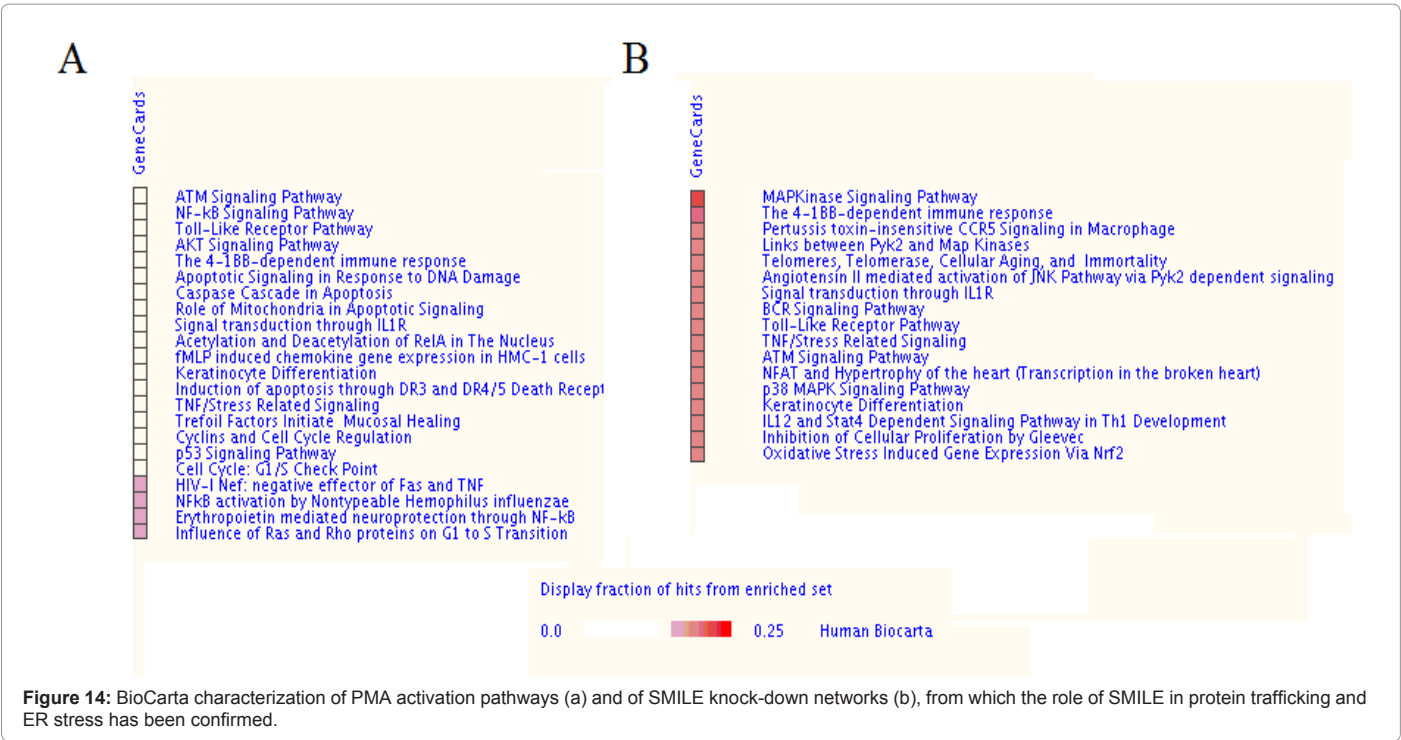
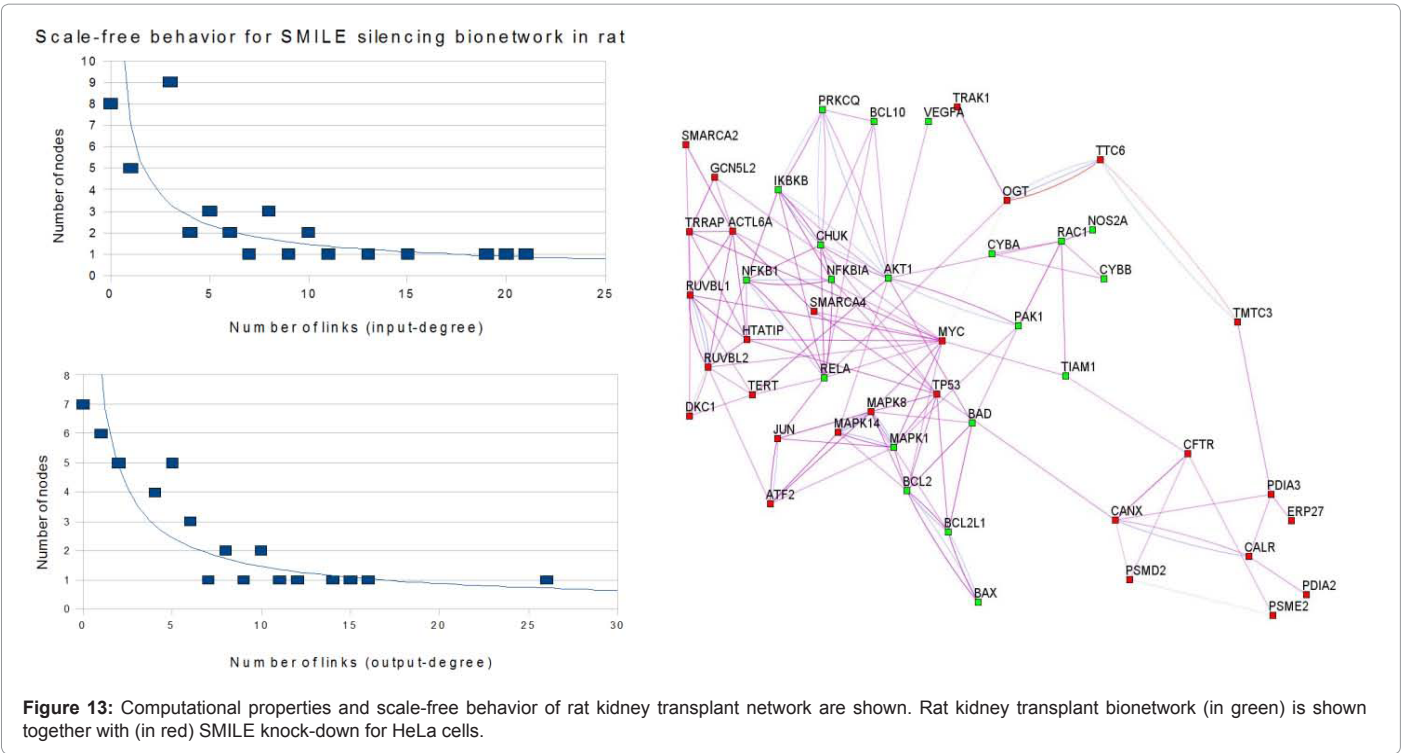




**Figure 11:** Computational properties and scale-free behavior of human kidney transplant network are shown. Human kidney transplant pro-rejection bionetwork (in green) is shown together with (in red) SMILE knock-down for HeLa cells.



**Figure 12:** Computational properties and scale-free behavior of rat kidney transplant network are shown. Rat kidney transplant bionetwork (in green) is shown together with (in red) PMA activation for HeLa cells.

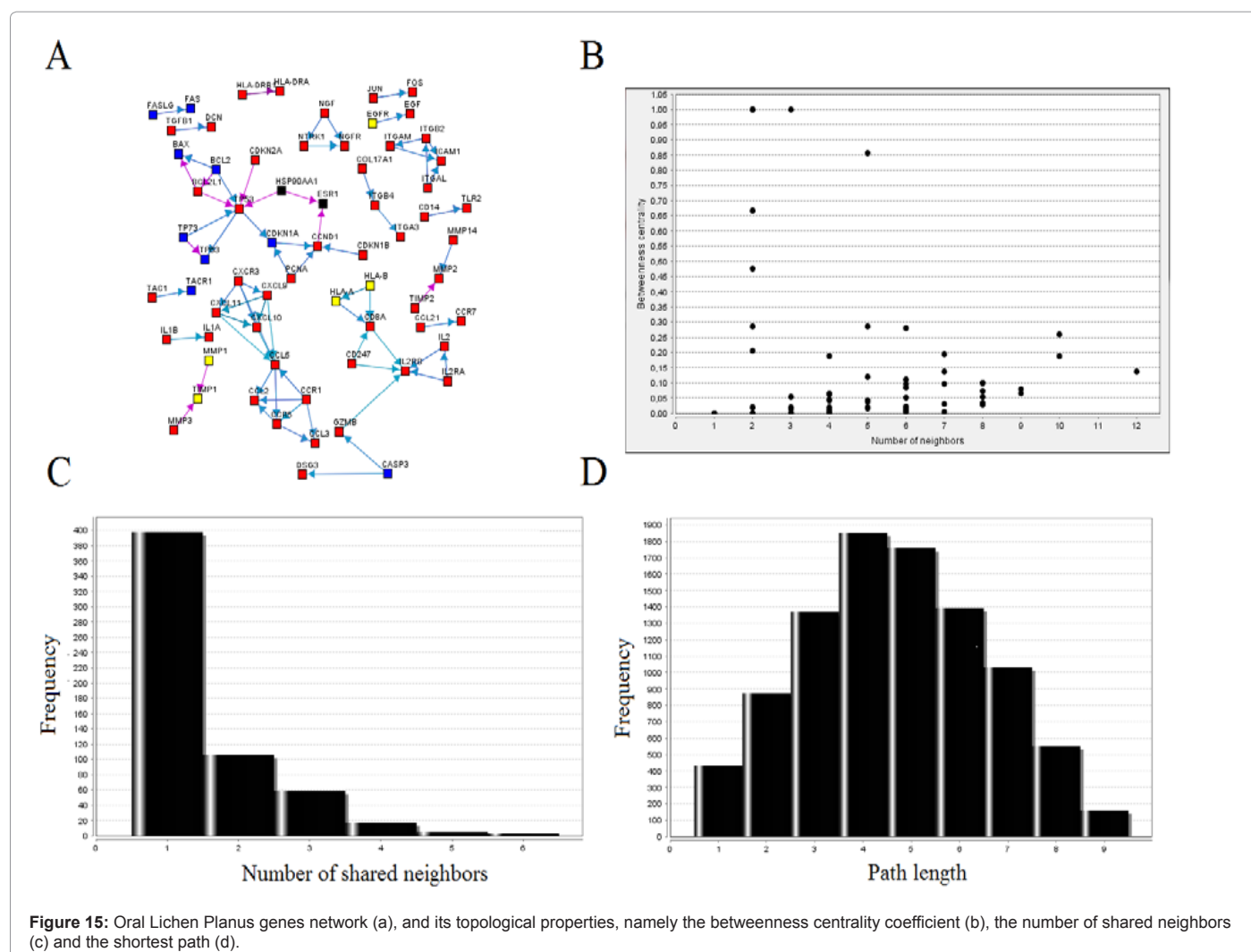


proliferation of both T and B lymphocytes. ITGB4, a receptor for the laminins, mediates cell-matrix or cell-cell adhesion, and transduced signals that regulate gene expression and cell growth. It is likely to play a pivotal role in the biology of invasive carcinoma. Interestingly, all of them but EGFR were up-regulated, and were widely distributed in the network (in term of topological parameters such as stress, eccentricity and radiality), and showed higher topological coefficients than the

other genes (Figure 15).

### Conclusions

In conclusion, our Leader Genes approach-based tool provides a general executive framework for identifying and ranking genes associated with complex and multifactorial diseases. However, the approaches and techniques that were systematically implemented



**Figure 15:** Oral Lichen Planus genes network (a), and its topological properties, namely the betweenness centrality coefficient (b), the number of shared neighbors (c) and the shortest path (d).

in the present study are general, and not confined by specific trait or species, and therefore, can be applied to various biological events, as well as diseases in different organisms. For this reason, the tool we presented is a general-purpose tool.

Moreover, in the last studies-specifically in the investigations on the role of SMILE/TMTC3 gene in kidney transplant operational tolerance, and on the etiology of OLP-we recently complemented the leader gene approach with a systems biology and topological analysis of the obtained graphs and networks. This is preliminary for further bioinformatics analysis and disease simulations using ad-hoc software. Topological analysis, in fact, can shed light on how molecular pathways work, and how a disease develops and evolves. Our analysis showed that our network exhibits a power law behaviour in agreement with the Scale-free theory of bio-networks, and has more FFL (feed-forward loops) than one would expect to find in a random graph (the so-called Erdős-Rényi random graph). The topological properties of leader genes and their role in controlling each pathway emerged from ontological analysis confirm our results.

For these reasons, we believe this tool can foster further research in the genomics field. Further developments are currently under study.

## Acknowledgments

This project was supported by grants to FEN (Fondazione Elba Nicolini) by the FIRB Italianonet (RBPR05JH2P) from MIUR (Ministero dell'Istruzione, Università e Ricerca; Italian Ministry for Research and University) to Claudio Nicolini of the University of Genova.

## References

- Nicolini C (2006) Nanogenomics for medicine. *Nanomedicine (Lond)* 1: 147-152.
- Nicolini C (2010) Nanogenomics in medicine. *Wiley Interdiscip Rev Nanomed Nanobiotechnol* 2: 59-76.
- Nicolini C, Pechkova E (2010) An overview of nanotechnology-based functional proteomics for cancer and cell cycle progression. *Anticancer Res* 30: 2073-2080.
- Jiang L, Sørensen P, Thomsen B, Edwards SM, Skarman A, et al. (2012) Gene prioritization for livestock diseases by data integration. *Physiol Genomics* 44: 305-317.
- Marx V (2013) Biology: The big challenges of big data. *Nature* 498: 255-260.
- Vivar JC, Pemu P, McPherson R, Ghosh S (2013) Redundancy Control in Pathway Databases (ReCiPa): An application for improving gene-set enrichment analysis in Omics studies and "Big Data" biology. *OMICS*.
- van Nas A, Pan C, Ingram-Drake LA, Ghazalpour A, Drake TA, et al. (2013) The systems genetics resource: a web application to mine global data for complex disease traits. *Front Genet* 4: 84.

8. Cohen JC, Hobbs HH (2013) Genetics. Simple genetics for a complex disease. *Science* 340: 689-690.
9. Hallock P, Thomas MA (2012) Integrating the Alzheimer's disease proteome and transcriptome: A comprehensive network model of a complex disease. *OMICS* 16: 37-49.
10. Jain P, Vig S, Datta M, Jindel D, Mathur AK, et al. (2013) Systems biology approach reveals genome to phenome correlation in type 2 diabetes. *PLoS One* 8: e53522.
11. Liu FJ, Hua XF, Wang WJ (2012) A new bioinformatics insight into human cancer-associated proteins. *Oncol Rep* 27: 1932-1936.
12. King JY, Ferrara R, Tabibiazar R, Spin JM, Chen MM, et al. (2005) Pathway analysis of coronary atherosclerosis. *Physiol Genomics* 23: 103-118.
13. Goh L, Kasabov N (2005) An integrated feature selection and classification method to select minimum number of variables on the case study of gene expression data. *J Bioinform Comput Biol* 3: 1107-1136.
14. Odibat O, Reddy CK (2012) Ranking differential hubs in gene co-expression networks. *J Bioinform Comput Biol* 10: 1240002.
15. Re M, Mesiti M, Valentini G (2012) A fast ranking algorithm for predicting gene functions in biomolecular networks. *IEEE/ACM Trans Comput Biol Bioinform* 9: 1812-1818.
16. Gonçalves JP, Francisco AP, Moreau Y, Madeira SC (2012) Interactogeneous: Disease gene prioritization using heterogeneous networks and full topology scores. *PLoS One* 7: e49634.
17. Guney E, Oliva B (2012) Exploiting protein-protein interaction networks for genome-wide disease-gene prioritization. *PLoS One* 7: e43557.
18. Zheng S, Zhao Z (2012) GenRev: Exploring functional relevance of genes in molecular networks. *Genomics* 99: 183-188.
19. Ni S, Vingron M (2012) R2KS: A novel measure for comparing gene expression based on ranked gene lists. *J Comput Biol* 19: 766-775.
20. Winter C, Kristiansen G, Kersting S, Roy J, Aust D, et al. (2012) Google goes cancer: improving outcome prediction for cancer patients by network-based ranking of marker genes. *PLoS Comput Biol* 8: e1002511.
21. Re M, Valentini G (2012) Cancer module genes ranking using kernelized score functions. *BMC Bioinformatics* 13: S3.
22. Shin M, Lee H, Hong M (2012) A hybrid approach to gene ranking using gene relation networks derived from literature for the identification of disease gene markers. *Int J Data Min Bioinform* 6: 239-254.
23. Feng J, Meyer CA, Wang Q, Liu JS, Shirley Liu X, et al. (2012) GFOLD: a generalized fold change for ranking differentially expressed genes from RNA-seq data. *Bioinformatics* 28: 2782-2788.
24. Roy S, Heinrich K, Phan V, Berry MW, Homayouni R (2011) Latent Semantic Indexing of PubMed abstracts for identification of transcription factor candidates from microarray derived gene sets. *BMC Bioinformatics* 12: S19.
25. Makita Y, Kobayashi N, Yoshida Y, Doi K, Mochizuki Y, et al. (2013) PosMed: ranking genes and bioresources based on Semantic Web Association Study. *Nucleic Acids Res* 41: W109-W114.
26. Jay JJ, Eblen JD, Zhang Y, Benson M, Perkins AD, et al. (2012) A systematic comparison of genome-scale clustering algorithms. *BMC Bioinformatics* 13: S7.
27. Bragazzi NL, Sivozhelezov V, Nicolini C (2011) Leader Gene: A fast data-mining tool for molecular genomics. *J Proteomics Bioinform* 4: 83-86.
28. Masoudi-Nejad A, Meshkin A, Haji-Eghrari B, Bidkhori G (2012) Candidate gene prioritization. *Mol Genet Genomics* 287: 679-698.
29. Xu R, Wang Q (2013) An iterative searching and ranking algorithm for prioritising pharmacogenomics genes. *Int J Comput Biol Drug Des* 6: 18-31.
30. von Mering C, Jensen LJ, Snel B, Hooper SD, Krupp M, et al. (2005) STRING: Known and predicted protein-protein associations, integrated and transferred across organisms. *Nucleic Acids Res* 33: D433-D437.
31. Biemann C (2006) Chinese Whispers-An efficient graph clustering algorithm and its application to natural language processing problems. *Proceedings of the first workshop on graph based methods for natural language processing*.
32. Giacomelli L, Nicolini C (2006) Gene expression of human T lymphocytes cell cycle: Experimental and bioinformatic analysis. *J Cell Biochem* 99: 1326-1333.
33. Nicolini C, Spera R, Stura E, Fiordoro S, Giacomelli L (2006) Gene expression in the cell cycle of human T-lymphocytes: II. Experimental determination by DNASER technology. *J Cell Biochem* 97: 1151-1159.
34. Sivozhelezov V, Giacomelli L, Tripathi S, Nicolini C (2006) Gene expression in the cell cycle of human T lymphocytes: I. Predicted gene and protein networks. *J Cell Biochem* 97: 1137-1150.
35. Covani U, Marconcini S, Giacomelli L, Sivozhelevov V, Barone A, et al. (2008) Bioinformatic prediction of leader genes in human periodontitis. *J Periodontol* 79: 1974-1983.
36. Braud C, Baeten D, Giral M, Pallier A, Ashton-Chess J, et al. (2008) Immunosuppressive drug-free operational immune tolerance in human kidney transplant recipients: Part I. Blood gene expression statistical analysis. *J Cell Biochem* 103: 1681-1692.
37. Sivozhelezov V, Braud C, Giacomelli L, Pechkova E, Giral M, et al. (2008) Immunosuppressive drug-free operational immune tolerance in human kidney transplants recipients. Part II. Non-statistical gene microarray analysis. *J Cell Biochem* 103: 1693-1706.
38. Jovanovic V, Giacomelli L, Sivozhelezov V, Degauque N, Lair D, et al. (2010) AKT1 leader gene and downstream targets are involved in a rat model of kidney allograft tolerance. *J Cell Biochem* 111: 709-719.
39. Racapé M, Bragazzi N, Sivozhelezov V, Danger R, Pechkova E, et al. (2012) SMILE silencing and PMA activation gene networks in HeLa cells: Comparison with kidney transplantation gene networks. *J Cell Biochem* 113: 1820-1832.
40. Orlando B, Bragazzi N, Nicolini C (2013) Bioinformatics and systems biology analysis of genes network involved in OLP (Oral Lichen Planus) pathogenesis. *Arch Oral Biol* 58: 664-673.