2010

A survey of current software for network analysis in molecular biology

Sterling Thomas
Virginia Commonwealth University

Danail Bonchev
Virginia Commonwealth University, dgbonchev@vcu.edu

A survey of current software for network analysis in molecular biology

Sterling Thomas[1] *and Danail Bonchev*[1,2*]

[1]Center for the Study of Biological Complexity, Virginia Commonwealth University, PO Box 842030, Richmond, VA 23284-2030, USA
[2]Department of Mathematics and Applied Mathematics, Virginia Commonwealth University, Richmond, VA 23284, USA
*Correspondence to: Tel: +1 804 827 7375; Fax: +1 804 828 1961; E-mail: dgbonchev@vcu.edu

Date received (in revised form): 22 April 2010

Abstract

Software for network motifs and modules is briefly reviewed, along with programs for network comparison. The three major software packages for network analysis, CYTOSCAPE, INGENUITY and PATHWAY STUDIO, and their associated databases, are compared in detail. A comparative test evaluated how these software packages perform the search for key terms and the creation of network from those terms and from experimental expression data.

Keywords: *software, network analysis, microarray expression analysis*

Introduction to network-related software tools

Post-genomic biology makes extensive use of network analysis at all levels of the hierarchy of life. Networks are basic tools in systems biology for expressing the essence of living things as whole integrated systems.[1–3] The explosive development of the theory of dynamic evolutionary networks during the past decade[4–6] stimulated the creation of numerous algorithms and software programs for constructing, manipulating and analysing networks. Many of those are multi-purpose programs with applications to most of the available types of complex networks: social, transportation, communication, financial, etc. This review focuses on software for the analysis of networks in living cells, the nodes in which represent genes, proteins, metabolites and other cell components. Examples of such networks are protein–protein interaction networks (PPN), gene regulatory networks (GRN), and metabolic and signalling networks and pathways, as well as disease-related or cell function-related networks.

The detailed analysis in this review is devoted to several of the most comprehensive and multi-functional software packages for network analysis in molecular biology. Other essential types of software in this field, which solve more specific network tasks, are also listed. One such kind of software performs a substructure search for identifying over-represented sub-graphs called motifs.[7,8] Viewed as the smallest building blocks of networks, motifs serve as a signature for distinguishing species, or different states of a single species, and are of interest for evolutionary and biomedical studies. The concept of motifs was developed in the Laboratory of Uri Alon from the Weizmann Institute in Israel, where a library of identification (ID) numbers of all motifs having three to eight nodes was created, along with the downloadable MFinder software for motif identification.[9] Other groups followed with freely available software: MAVisto (Schreiber and Schwöbbermeyer)[10] and FANMOD (Wernicke and Rasche).[11] FANMOD is particularly user-friendly and fast software which runs under Linux, MacOS and Windows, and classifies the motifs according to their frequency of

distribution, *p* values and z-scores, in comparison with the generated randomised networks having the same size and the same node degree distribution. Recently, another freely available software, named Kavosh (Kashani *et al.*),[12,13] claimed slightly better performance than FANMOD and added the option of handling motifs having more than eight nodes.

Modularisation of networks is another area of intensive research, aimed at facilitating the analysis of complex networks by partitioning them into modules, presumably related to a certain biological function. A large variety of approaches have been proposed, many of them constructing different network profiles and hierarchical trees, such as those based on node connectivity (the topological overlap method of Ravasz *et al.*)[14] and node distance (the association matrix method of Rives and Galitski).[15] Another group of methods includes extreme pathway analysis (Papin *et al.*)[16] and flux analysis (Burgard *et al.*)[17] in metabolic networks. Considerable network software resources are available online for flux analysis.[18,19] Several modularisation programs enjoy considerable popularity. The simulated annealing algorithm (Guimerá and Nunes Amaral)[20] is a stochastic optimisation technique that enables the discovery of low computational cost modular configurations without getting trapped in 'high-cost' local minima. The optimised modularisation function is based on accounting for the fractions of intra-module and inter-module links. The first algorithm of Newman[21] makes use of a similar type of modularisation function and agglomerative hierarchical clustering procedure. The Newman 2006 algorithm[22] uses a modularity score defined in terms of the eigenvectors and eigenvalues of a specifically defined modularity matrix. The algorithm is very fast: a network with 27,000 nodes runs for 20 minutes on a standard personal computer. Both software programs are available from the author on request.

A third, more recent, group of network-related software provides tools for direct network comparison. Software of this type aims to prove that comparative interactomics can reproduce the results provided by comparative genomics and, in addition, can identify conserved functional modules, predict network module functions and query such modules. Three modes of network comparison are implemented: network alignment, integration and querying.[23] The basic software is the Network Comparison Toolkit (NCT), a Java 1.5 library developed to be modular, easily extended and freely downloadable.[24] The project was initiated in the Ideker Laboratory at the University of California, San Diego (UCSD).[25] The toolkit provides options for predicting protein–protein interactions and protein functions. Stanford University developed Graemlin, another software package for network comparison, which provides fast network alignment that scales linearly with the number of networks compared, and supports efficient querying of modules.[26] The software scores separate species into equivalence classes and reconstructs the most parsimonious ancestral history of an equivalence class using dynamic programming based on five types of evolutionary events. Graemlin 2.0 is freely available[27] and the source code is available under the GNU Public Licence. ('GNU' is an acronym for 'GNU's Not UnixI' and is a UNIX-like computer operating system, composed of free software.)

Basic software for network analysis: Pathway Studio, Ingenuity Pathway Analysis (IPA) and Cytoscape

Technical specifications

Pathway Studio (Ariadne Genomics, Rockville, MD) is a software which builds networks and pathways from relationships between biological molecules and processes extracted from the literature, PubMed, databases, expression and proteomics data. Pathway Studio is offered with the ResNet Mammalian and ResNet Plant databas**e**s, and supports KEGG, Science Signaling and Prolexys HyNet protein–protein interaction databases. The ResNet 7.0 database contains over 1.5 million relationships for 110,435 proteins, 814 cellular processes, 2,410 diseases and 248 curated pathways. The ResNet Plant database includes over 90,000 relationships for 71,501 proteins, 915 cellular

processes, 97 plant diseases, and 315 AraCyc and 17 plant signalling pathways. The ResNet databases can be kept updated with Ariadne MedScan technology and quarterly updates. The software also calculates the node degrees in the network built, and compares them with the node degrees in the ResNet database. The software is available as Desktop and Enterprise editions. Pathway Studio Enterprise includes server-side applications with Windows, Linux and Solaris, compared with the client-side Windows (XP, Vista and 7) applications of the Desktop version. The Desktop version of Pathway Studio 7.0 requires a minimum of 2 GB RAM, while 4 GB are recommended. The server makes use of two or more quad-core Intel/AMD processors with 4–10 GB RAM and 32-/64-bit operating system (OS) and Java virtual machine (JVM) — a crucial component of the Java platform.

IPA software, licensed by Ingenuity Systems (Ingenuity Systems Inc., Redwood City, CA), is a service model that requires internet access. IPA operates under Java runtime environment 1.5.x and 1.6.x. A Java-based start-up application downloads to your machine and initiates the connection to Ingenuity Systems back-end server infrastructure. This allows IPA to be run on any machine with a web interface running most versions of Windows, from XP to 7, and Max OS 10.4.2 to 10.6.2. The most recent IPA 8.0 version requires a minimum of 512 MB of RAM but recommends 1 GB for Windows XP and Mac OS X, and 2 GB under Vista and Windows 7. IPA uses a database for human genes/proteins, created from manually curated literature searches (Ingenuity® ExpertAssist Findings). Specific data on the number of interactions and the number of molecules is not reported. IPA also includes interaction data from third party databases, such as IntAct, BIND, DIP, MINT, MIPS, BIOGRID and COGNIA. IPA offers several major functional blocks. The Core Analysis is the basic block for analysis of protein–protein interaction networks. IPA-Metabolomics Analysis analyses metabolite data. A couple of blocks provide tools for the analysis of IPA applications for toxicity and biomarker identification. The Comparative Analysis option provides tools for analysing changes in biological states across experimental conditions.

Cytoscape,[28,29] similar to Pathway Studio, is an installable program that resides on your computer. It is a collaborative effort of the Institute for Systems Biology (ISB; Seattle, WA), UCSD, Sloan–Kettering Cancer Center, Institut Pasteur, Agilent Technologies and the University of California, San Francisco. Cytoscape is available as a platform-independent open-source Java application and can be installed in Windows, Mac OSX or Linux environments with at least 256 MB RAM. By contrast with Pathway Studio, Cytoscape has the ability to connect to external data sources, either directly or using plug-ins. The latter are uniquely characteristic of Cytoscape. They are small programs developed by the Cytoscape team and third-party developers. Most are free, but some cost a small amount of money. From these plug-ins, Cytoscape users can connect to IntAct, KEGG, Pathway Commons and several other interaction databases. Users can also calculate many network analytics like centrality, eccentricity and node degree. The Cytoscape application programming interface (API) is available publicly, so with the proper training, many scientists can create plug-ins unique to their projects.

Analysis

To compare the functionality of Cytoscape, Pathway Studio and IPA, two tests were performed, based on the expectations of what a typical scientists would do. The two tests were: (i) searching for key terms associated with disease and treatment to build a network and (ii) importing expression data from an experiment associated with a disease state, and using that data to create a network. Each of the software packages had functionality around each of the tasks. Two machines were used to test the three software packages: an HP Pavilion Tower Desktop with Windows Vista with an Intel Core 2 Quad processor and 8 GB systems memory to test Pathway Studio 7.0, and an Apple MacBook running OS 10.5 with an Intel Core 2 Duo processor and 2 GB of RAM to test Cytoscape 2.6.3 and IPA 8.0 versions.

Method of analysis

Search is a key function of network discovery. Common usage of these packages requires the ability to search specific genes and keywords associated with a gene, protein or molecular function. Each software solution allows for searching of gene ontology indexes, but a more important function is the ability to identify genes, proteins or molecules associated with key terms. In the present analysis, two key terms were used: (i) 'resistance', for finding networks associated with drug resistance; and (ii) 'migration', with the intention of identifying networks associated with metastasis. Another key function that has become available in the past two years is importing data from mRNA-based expression platforms and building a network associated with a specific experiment. The algorithms range from simple searches of the gene symbols to *de novo* network discovery based on expression patterns. This analysis used data from 27 patients with adenocarcinoma of the lung, with the data generated by an Affymetrix Human 133A chip. The dataset was limited to 61 key probes identified in previous studies.[30]

Analysis of results: Searches

Pathway Studio produced 1,378 results with the search term 'resistance'. The network built included 5,064 interactions. The entities were of two categories: (i) genes/proteins and other molecules, and (ii) functions or groups. The top ten most frequent results from both types are given in Table 1.

The search for 'migration' using Pathway Studio produced 2,293 entities and 7,218 interactions: 5,852 regulations, 1,011 expressions, as well as some protein modifications, molecular transport, promoter binding, and molecular synthesis. The top ten most frequent results from both types are given in Table 2.

The Pathway Studio 'Gene Sets Enrichment Analysis' (GSEA) function offers additional analysis by which the user can identify commonly occurring gene ontology keywords in their network. The available options are: Ariadne Metabolic Pathways, Ariadne Signaling Pathways, Users Pathways, Ariadne Ontologies, Gene Ontology Cellular Components, Gene Ontology Molecular Functions, Gene Ontology Biological Processes and Users Groups. In the authors' analysis, all the

Table 1. The top ten proteins and top ten terms for function or group with their frequencies, as produced by Pathway Studio software 7.0, using the keyword 'resistance'.*

Top 10 proteins	Top 10 terms for function or group
INS (110)	Apoptosis[a] (169)
AKT1 (54)	Neoplasm[b] (68)
MAPK1 (43)	Cell proliferation[a] (62)
TNFSF10 (38)	Insulin resistance[b] (60)
ADIPOQ (35)	NF-kappaB[c] (59)
TNF (34)	Cytokine[c] (52)
ABCB1 (33)	Cell death[a] (44)
TP53 (32)	Drug resistance[a] (27)
MAPK8 (31)	Cell differentiation[a] (23)
LEP (31)	TNF family[c] (22)

NF, nuclear factor; TNF, tumour necrosis factor
*Numbers in parentheses are the numbers of protein in each group. [a]Cell process; [b]disease; [c]functional class.

Table 2. The top ten proteins and top ten terms for function or group with their frequencies, as produced by Pathway Studio software, using the keyword 'migration'.*

Top10 proteins	Top 10 terms for function or group
MAPK1 (178)	Cell migration[a] (506)
VEGFA (167)	Cell proliferation[a] (406)
AKT1 (156)	Neoplasm[b] (268)
ITG (118)	Angiogenesis[a] (169)
MMP9 (117)	Cell differentiation[a] (117)
TGFB1 (111)	Apoptosis[a] (154)
PTK2 (97)	Neoplasm metastasis[b] (125)
MMP2 (95)	Cytokine[c] (115)
SRC (92)	NF-kappaB[c] (98)
PI3K (89)	Chemokine[c] (91)

NF, nuclear factor; TNF, tumour necrosis factor
*Numbers in parentheses are the numbers of proteins in each group. [a]Cell process; [b]disease; [c]functional class.

Gene Ontology categories were selected. The software provides a comparison of the number of entities found in the user's network and in the entire category. It also provides the number and names of nodes that are found in more than one category, a p value and a ranking. The 'resistance' search resulted in a list of 331 categories; the top two ranked were 'ATPase activity', with p values of $2.13\,e^{-34}$ and $2.67\,e^{-22}$. The 'migration' search produced a list of 234 categories; the top two were 'inflammatory response' and 'cytosol', with p values of $4.86\ e^{-5}$ and $1.11\ e^{-4}$, respectively.

IPA has a search function that is very simple to use. It allows the user to search for three categories of information: 'Genes and Chemicals', 'Functions and Diseases' and 'Pathways and Tox Lists'. The authors used the 'Functions and Diseases' category. After the initial search, the results can be further refined by selecting specific groups of the genes, proteins and molecules found. This is similar to the GSEA analysis of Pathway Studio. In addition, however, the results in IPA are in graphical form for easy analysis, showing the results as bar graphs and the significance overlaid as a line graph. In the authors' analysis, they selected the top two categories, 'Cancer' and 'Drug Resistance Based Keywords' for the resistance search, and 'Cell Migration Keywords' for the migration search. The 'resistance' search produced a network of 91 nodes and 533 links associated with the search term and the two keywords. As with Pathway Studio, IPA allows the user to adjust the views of the network based on cellular compartment and functional groups, as well as to view more information by selecting nodes and edges. When the authors attempted to visualise the network for migration, they ran into the 500-node limit on visualisation within IPA's viewer. To get around this limitation, they had to refine their search further by selecting two more keywords ('Cancer' and 'Drug Resistance'). This narrowed their network to 59 nodes, with 508 links between the nodes.

Cytoscape also allows keyword searching through 'Import' from the 'Network from Web Services' function. The user needs first to select the database and then to execute the search. Depending upon the database, the time and size of the resulting set can be controlled. The authors chose to search the Pathway Commons database for their analysis. When searching this database, it is possible to specify the species to search and to import curated pathways or all interactions associated with the search result. To do the latter, one selects the protein ID or gene name reported as a search result, selects the tab labelled 'Interaction Networks' and then selects 'Retrieve Interactions'. The protein or gene names associated with each search are reported in Table 3. The authors used this process for both of their search terms. For 'Resistance', the resulting network had 411 nodes, with 10,561 edges. One can adjust how these results are viewed by selecting a specific layout from the Layout menu (as with IPA) or import a third-party plug-in to create layouts based on node characteristics. In addition to layouts, one can calculate network topology statistics to identify the most connected nodes, most central nodes, most eccentric nodes and many more.

Table 3. Top ten terms found by Cytoscape software searching the Pathway Commons database for key words 'migration' and 'resistance'.*

Search term 'resistance'	Search term 'migration'
BCAR1[a] (248)*	MIF[a] (248)
BCAR3[a] (59)	NUDC[a] (32)
GBF1[a] (15)	DCX[a] (14)
CROP[a] (13)	Migration-inducing protein (19)
MDR1[a] (7)	RAC1[a] (321)
Arsenite resistance protein (2)	Microphage migration inhibitory factor (1)
MRP1[a] (2)	RL36A[a] (8)
Breast cancer anti-oestrogen resistance (3)	SIOA8[a] (71)
Multi-drug resistance protein (3)	PGK1[a] (16)
Multi-drug resistance protein (1)	SMAD3[a] (578)

*The number of proteins in each group is shown in parenthesis. [a]Human

Due to the availability of more than one database, the authors chose to search IntAct, as well as Pathway Commons, using Cytoscape. The IntAct search produced a network with 1,316 nodes and 3,884 interactions. The IntAct search function can be accessed from the same menu and allows the search to be limited by time and size of result set. The search for 'migration' using Pathway Commons produced a network of 1,166 nodes and 13,755 links. The same search of the IntAct database produced a network of 3,250 nodes, with 8,799 interactions. Thus, Cytoscape produced networks containing more interactions than Pathway Studio- and Ingenuity-based networks.

Analysis of results: Building and analysing a network from experimental data

Another key function that has become available with all three software packages over the past couple of years is the facility to import data from mRNA-based expression platforms and building a network associated with a specific experiment. The algorithms range from simple searches of gene symbols to *de novo* network discovery based on expression patterns.

Pathway Studio offers a function for performing a comparative analysis of imported multiple gene expression, proteomics and metabolomics experiments. A user can select the platform for generating the data and the column with the matching probe IDs. Pathway Studio requires the entire dataset to be uploaded as one table, but allows the user to select normal and abnormal columns. The results from analysis provide a heat map of the diseased and normal expression and the *p* value associated with the uploaded data. The data can then be correlated and used to generate a network, or be overlaid onto a network from the ResNet database or from a MedScan v. 3.0 search. The correlation analysis is limited to pairwise correlations and can not be used to infer interaction, but may provide additional insight into the data.

IPA has a function for Dataset Search and Analysis, which can import files. Most commonly used formats are available, except the newer XML-based Excel format. After uploading the data, the user is prompted to select the commercial mRNA platform (human U133a or similar) to generate the data. This step annotates the probe IDs and provides a real-time matching percentage. IPA limits the number of experimental data columns to 20, so some of the authors' 27 columns could not be uploaded. Additional menus allow for modifications to the search and analysis options. Modifying these options would limit the scope of the search results, with the hope of increasing the accuracy, while reducing the size of the set of results. For the authors' study set, IPA generated 20 unique networks that showed little overlap. Each network could then be expanded and adjusted as described earlier. The expression data are overlaid onto each node, and each network is created based on the highest-fold change in the data. IPA does not have the ability to infer interactions based on data.

Cytoscape functions for experimental data analysis are available as third-party plug-ins. At the time of writing, there are two types of such plug-ins: expression overlay (similar to IPA) and network inference. The plug-in Genoscape[31] has been developed in a collaboration between scientists at several leading European institutions. Genoscape allows users to import gene expression information from GenoScript and KEGG pathways. Additionally, a user can create a tab-delineated file of original gene expression data to import into Genoscape. The plug-in visualises gene expression changes for each node and provides statistical analysis of the significance of these changes.

Cytoscape also allows users to import gene expression, proteomics or metabolomics data through the Network Attribute import function. After selecting an attribute file to import, Cytoscape allows the user to select a column to map the expression to nodes, and to identify the columns with expression data. After the gene expression data are imported, the Vismapper tool can be used to visualise the expression by colour on the nodes.

Network discovery is a new function that is emerging in biological network analysis. Currently, most networks are created by searching databases of curated literature-sourced interactions like ResNet and IntAct. Network Builder is a new plug-in that

allows the user to infer interactions from gene expression or mass spectrometry data. An example of how this type of network creation can be used in the search for lung cancer biomarkers is presented in Kuznetsov *et al.*[30]

Conclusion

The three basic software packages for network analysis discussed here offer similar functions and tools. The commercial Pathway Studio and IPA packages produce more visually appealing networks, but limit the number of analytical tools available to the user. Cytoscape, as an open-source software package, has been developed by a community of scientists and programmers from different universities and research institutions, collaborating to create better tools. IPA and Pathway Studio offer less of a development community, but provide a more refined and stable software solution. It is difficult to predict where the future of these software tools lies, but one may expect them to become even more universal by including blocks or plug-ins for substructure analysis (modules and motifs) and calculation of network descriptors (such as based on connectivity, distances, centrality, clustering, etc.). Being a more flexible dynamic structure, the Cytoscape community shows promise as a future front-runner for this type of scientific software; however, IPA and Pathway Studio will continue to be strong and very popular, with their online training videos, webinars and specialised conferences devoted to the software applications. Reasonable advice to researchers interested in network analysis applications is to use at least two of the leading software packages and rely on the results that overlap.

Currently, proprietary databases are the key sources of network generation. With the advancement of the National Institutes of Health- and European Bioinformatics Institute-supported interaction databases, and their rapid weekly update schedule, the commercial databases might be expected to become less relevant. Pathway Studio's MedScan function is an obvious response to this challenge, and offers an excellent way of producing the most up-to-date version of an interaction database. Ingenuity Systems has also provided a similar solution, named ExpertAssist Findings. This database is generated by a text-based search of recent publications, as with MedScan, but these are reviewed manually to verify the validity of the interactions. This is updated weekly to provide the most recent interactions.

Systems biology continues to grow and is quickly moving from academic laboratories to commercial R&D. Network discovery and analysis will become increasingly more important in the study of gene signalling and molecular communication in biology and biomedical research, as well as in the field of drug design. This growth will provide more resources to expand the currently existing software solutions, and will, without doubt, bring better network-based technology and more powerful analytical tools in the very near future.

References

1. Kitano, H. (2002), 'Computational systems biology', *Nature* Vol. 420, pp. 206–210.
2. Barabási, A.-L. and Oltvai, Z. N. (2004), 'Network biology: Understanding the cell's functional organization', *Nat. Rev. Genet*. Vol. 5, pp. 101–114.
3. Alon, U. (2006), *An Introduction to Systems Biology: Design Principles of Biological Circuits*, Chapman and Hall/CRC, Boca Raton, FL.
4. Barabási, A.-L. (2002), *Linked: The New Science of Networks*, Perseus, Cambridge, MA.
5. Newman, M., Barabási, A.-L. and Watts, D. J. (2006), *The Structure and Dynamics of Networks*, Princeton University Press, Princeton, NJ.
6. Watts, D.J. and Strogatz, S.H. (1998), 'Collective dynamics of "small-world" networks', *Nature* Vol. 393, pp. 440–442.
7. Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N. *et al*. (2002), 'Network motifs: Simple building blocks of complex networks', *Science* Vol. 298, pp. 824–827.
8. Shen-Orr, S.S., Milo, R., Mangan, S. and Alon, U. (2002), 'Network motifs in the transcriptional regulation network of *Escherichia coli*', *Nat. Genet*. Vol. 31, pp. 64–68.
9. Uri Alon's Molecular Cell Biology Lab, Weizmann Institute of Science, Israel, http://www.weizmann.ac.il/mcb/UriAlon/ (last accessed 4th April, 2010).
10. Schreiber, F. and Schwöbbermeyer, H. (2005), 'MAVisto: A tool for the exploration of network motifs', *Bioinformatics* Vol. 21, pp. 3572–3574.
11. Wernicke, S. and Rasche, F. (2006), 'FANMOD: A tool for fast network motif detection', *Bioinformatics* Vol. 22, pp. 52–53.
12. Kashani, Z.R., Ahrabian, H., Elahi, E., Nowzari-Dalini, A. *et al*. (2009), 'Kavosh: A new algorithm for finding network motifs', *BMC Bioinformatics* Vol. 10, pp. 318–329.
13. The Kavosh algorithm: http://Lbb.ut.ac.ir/Download/LBBsoft/Kavosh/, University of Tehran, Tehran, Iran (last accessed 11th December, 2009).
14. Ravasz, E., Somera, A.L., Mongru, D.A., Oltvai, Z.N. *et al*. (2002), 'Hierarchical organization of modularity in metabolic networks', *Science* Vol. 297, pp. 1551–1555.

15. Rives, A.W. and Galitski, T. (2003), 'Modular organization of cellular networks', *Proc. Natl. Acad. Sci. USA* Vol. 100, pp. 1128–1133.
16. Papin, J.A., Reed, J.L. and Palsson, B.O. (2004), 'Hierarchical thinking in network biology: The unbiased modularization of biochemical networks', *Trends Biochem. Sci.* Vol. 29, pp. 641–647.
17. Burgard, A.P., Nikolaev, E.V., Schilling, C.H. and Maranas, C.D. (2004), 'Flux coupling analysis of genome-scale metabolic network reconstructions', *Genome Res.* Vol. 14, 301–312.
18. Network flux analysis software: http://systemsbiology.ucsd.edu/Downloads (last accessed 5th October, 2009).
19. Zamboni, N., Fischer, E. and Sauer, U. (2005), 'FiatFlux — a software for metabolic flux analysis from ^{13}C-glucose experiments', *BMC Bioinformatics* Vol. 6, pp. 209–216.
20. Guimerá, R. and Nunes Amaral, L.A. (2005), 'Functional cartography of complex metabolic networks', *Nature* Vol. 433, pp. 895–900.
21. Newman, M.E.J. (2004), 'Detecting community structure in networks', *Eur. Phys. J. B.* Vol. 38, pp. 321–330.
22. Newman, M.E.J. (2006), 'Finding community structure in networks using the eigenvectors of matrices', available at: *arXiv:physics*/0605087 v3 23 Jul.
23. Sharan, R. and Ideker, T. (2006), 'Modeling cellular machinery through biological network comparison', *Nat. Technol.* Vol. 24, pp. 427–433.
24. NCT Software for network comparison, http://chianti.ucsd.edu/nct/, UCSD, San Diego, CA (last accessed 4th April, 2010).
25. Ideker laboratory, http://chianti.ucsd.edu/idekerlab/, UCSD, San Diego, CA (last accessed 4th April, 2010).
26. Flannick, J., Novak, A., Srinivasan, B.S., McAdams, H.H. *et al.* (2006), 'Græmlin: General and robust alignment of multiple networks', *Genome Res.* Vol. 16, pp. 1169–1181.
27. The Graemlin software, http://graemlin.stanford.edu/download.php, Stanford University, Stanford, CA (last accessed 4th April, 2010).
28. Ideker, T., Thorsson, V., Ranish, J.A., Christmas, R. *et al.* (2001), 'Integrated genomic and proteomic analyses of a systematically perturbed metabolic network', *Science* Vol. 292, pp. 929–934.
29. The Cytoscape software, http://cytosca-pe.org/download.php, UCSD, San Diego, CA (last accessed 4th April, 2010).
30. Kuznetsov, V., Thomas, S. and Bonchev, D. (2008), 'Data-driven networking reveals 5-genes signature for early detection of lung cancer', Proceedings of the International Conference on Biomedical Engineering and Informatics (BMEI), 27–30 May, Sanya, Hainan, China, Peng, Y. and Zhang, Y. (eds), *IEEE*, Vol. 1, pp. 413–417.
31. Clément-Ziza, M., Malabat, C., Weber, C., Moszer, I. *et al.* (2009), 'Genoscape: A Cytoscape plug-in to automate the retrieval and integration of gene expression data and molecular networks', *Bioinformatics* Vol. 25, pp. 2617–2618.