Plant Systems Biology
Edited by Sacha Baginsky and Alisdair R. Fernie

Network visualization and network analysis

Victoria J. Nikiforova and Lothar Willmitzer

Max-Planck-Institut für Molekulare Pflanzenphysiologie, Am Mühlenberg 1, 14476 Potsdam-Golm, Germany

Abstract

Network analysis of living systems is an essential component of contemporary systems biology. It is targeted at assemblance of mutual dependences between interacting systems elements into an integrated view of whole-system functioning. In the following chapter we describe the existing classification of what is referred to as biological networks and show how complex interdependencies in biological systems can be represented in a simpler form of network graphs. Further structural analysis of the assembled biological network allows getting knowledge on the functioning of the entire biological system. Such aspects of network structure as connectivity of network elements and connectivity degree distribution, degree of node centralities, clustering coefficient, network diameter and average path length are touched. Networks are analyzed as static entities, or the dynamical behavior of underlying biological systems may be considered. The description of mathematical and computational approaches for determining the dynamics of regulatory networks is provided. Causality as another characteristic feature of a dynamically functioning biosystem can be also accessed in the reconstruction of biological networks; we give the examples of how this integration is accomplished. Further questions about network dynamics and evolution can be approached by means of network comparison. Network analysis gives rise to new global hypotheses on systems functionality and reductionist findings of novel molecular interactions, based on the reliability of network reconstructions, which has to be tested in the subsequent experiments. We provide a collection of useful links to be used for the analysis of biological networks.

Introduction

A living organism consists of a lot of elements (e.g., genes, proteins, metabolites, etc.) organized in a functional structure capable simultaneously to maintain its homeostasis and to develop. In addition, this structure must be able to react to the changes in both external and internal environment. This reaction itself constitutes a chain of consecutive events starting from signal perception through signal transduction and various subsequent transformations towards an endpoint response reaction. These events need to be integrated in a proper spatial and temporal context. The events in such chains are changes in a state of elements, and information concerning these changes propagates along the chain. From this explanation, the answer to the

biological questions why and how a particular response to a given signal develops seems to be relatively straightforward. However, the complexity of living systems is so high, that to date hardly any such chains of reactions have been elucidated. Actually, for the vast majority of reactions our knowledge is at a rudimentary 'black box' stage: we know the initial signal (the exciter) and a response endpoint, but how spatio-temporal aspects of responses are executed remains largely unknown. A further complexity is introduced by the fact that a single exciter generally influences more than one physiological reaction. For the above-described simplified concept of information exchange this suggests that the chains of consecutive events occurring in response to the exciter must branch, and change in a state of each element within a chain can result in multiple downstream effects. This response plurality can nowadays be easily illustrated with the use of transcript profiles, which are rapidly accumulating in public repositories and hence available for the research community. In most underlying physiological experiments a single environmental parameter is altered, and in response expression of a large number of genes is changed. For example, in experiments in which sulfur was depleted from the *Arabidopsis* growth medium, up to 5% of all genes and 11.5% of measured metabolites exhibited significantly different levels [1]. These multiple changes in response to a single initial exciter have to be extrapolated to the whole system of response development. Each new change in a chain (being in turn an exciter for the downstream changes) is also potentially able to cause multiple changes downstream in the network. Thus, information on the initial exciter spreads in multiple downstream directions, forming a dense causally directed network of interactions. Studying the network of interacting elements within living systems is facilitating efforts to fill the 'response black box' – a task that represents a major challenge for network analysis as a component of contemporary systems biology.

Types of recognized biological networks

According to the Webster's dictionary, a network is an intricately connected system of things or people. A type of a biological network is defined by what these 'things' are (nodes, vertices, etc.), what the nature of their connections (edges) is, and ideally why these things are connected. Below we give the examples of the most common types of biological (often termed also cellular or molecular) networks, with comments on what knowledge is usually gained from networks of these types. It is worth mentioning that in this relatively new research area the terminology is not yet well-established. Table 1 illustrates the frequency of different terms used for biological networks in the related literature as of January 2006.

In what are currently termed metabolic networks, or biochemical reaction networks, vertices are represented by metabolites (substances), and metabolic reactions are represented by directed edges, which interconnect substrates and products of these reactions. Metabolic networks describe the potential pathways that may be used by a cell to accomplish metabolic processes. These are probably the first cellular networks, which biologists started to reconstruct as schematic representations

Table 1. Terminology of biological networks

	Term in Ovid Database Server (http://ovid.gwdg.de/)	Frequency
0	biological network(s)	235
0	cellular network(s) (used also in, e.g., Telecommunication Systems)	1,089
0	molecular network(s)	400
0	biomolecular network(s)	9
0	bioregulatory network(s)	4
1	metabolic network(s)	626
1	biochemical reaction network(s)	45
2	transcription network(s)	47
2	network(s) of transcription interactions	1
2	gene regulation network(s)	26
2	gene-regulatory network(s) (used broader)	234
2	transcriptional regulation network(s)	14
2	regulatory network(s) (used very broad)	1,666
3	protein interaction network(s)	218
3	protein–protein interaction network(s)	62
3	interactome	101
4	correlation network(s) (not only biological networks)	64
4	co-expression network(s)	5
4	coexpression network(s)	9
4	expression network(s)	71
5	signaling network(s)	1,249
	signaling network(s)	1,030
	signaling network(s)	223
6	gene network(s)	552
6	genetic regulatory network(s)	113

of a sum of biosynthetic pathways deduced from biochemical studies. Nowadays the vast biochemical information is compiled in specialized databases, and metabolic networks on top of these data serve as a visualization tool for multiple interconnections between their elements. As an example of such repositories, BioCyc [2] is a collection of 205 (as of January 2006) Pathway/Genome Databases, each of which describes the genome and metabolic pathways of a single organism. Among these organisms plant biologists will find a comprehensive *Arabidopsis* Pathway/Genome Database called AraCyc [3]. Connected to the BioCyc repository is the MetaCyc database, which, in distinction to the organism-specific databases, is a reference source on metabolic pathways from many organisms [4]. Another example is the KEGG PATHWAY [5], a collection of manually drawn pathway maps representing our up-to-date knowledge on the molecular interaction and reaction networks. Although very rich, this database may be less recommended for plant biologists, as the reference metabolic networks represent non-plant metabolism. The enzymes known for plants can be mapped on these networks, but the reactions

which are known not to occur in plants will still stay in the networks as connecting links. However, keeping in mind that, contrary to conventional wisdom, our current knowledge of the structure of plant cellular metabolism is far from complete [6], expansion and integration of the knowledge of metabolism in well characterized 'post-genome' organisms into plant biology will facilitate faster progress in plant systems.

In transcription networks (termed also: networks of transcription interactions, gene regulation networks, gene-regulatory networks, transcriptional regulation networks or simply regulatory networks) directed edges reflect interactions between transcription factors and the genes they regulate or the DNA sites to which they bind, with the direction from the transcription factor to the regulated gene. These networks describe potential pathways cells can use to regulate global gene expression programs. This is a newer type of cellular network which started to develop with the accumulating knowledge on protein factors regulating transcription of target genes by means of binding to the regulatory elements contained in their promoters. As with biochemical repositories, the information on experimentally verified interactions is also collected in major electronically accessible data bases. Here analysis at the network level is essential, because each transcription factor generally regulates the expression of more than one gene, the expression of each gene is often regulated by more than one transcription factor, and furthermore, the expression of transcription factors themselves can be regulated by the other transcription factors in a cascade-like manner. Thus, this type of information exchange also forms a dense network of interactions.

For many model systems the complete arrays of transcription factors and their target genes have been deciphered and compiled into electronic repositories. The major data repository for gene regulation in *Escherichia coli* is stored in RegulonDB [7], while the GRID database compiles information on physical interactions for three organisms whose genomes have been deciphered: yeast *Saccharomyces cerevisiae*, fly *Drosophila melanogaster* and worm *Caenorhabditis elegans*. Among plant-specific databases, the major ones which collect information on transcription factors and cis-regulatory elements are AGRIS, DATF, PlantCare and Place. Data on identified molecular interactions are also collected within the more general databases (such as BIND [8]), which are organism- and interaction-type unspecific. The analysis of genome-scale transcription networks is exemplified by the papers [9] for *E. coli* and [10] for yeast, but no comprehensive survey of this type exists yet for plants.

In the other type of cellular graphs – protein interaction networks – the nodes are proteins, and two nodes are connected by a non-directed edge if the two proteins bind to each other. In parallel with the rapid development of modern molecular techniques for determining protein–protein interactions, such as high-throughput yeast two-hybrid strategies [11], proteome-scale reconstructions of global protein interaction networks have been carried out for some model organisms. An organism's total set of protein–protein interactions is often termed as its interactome [12, 13]. Similarly to the data on metabolic and transcriptional interactions, that concerning protein–protein interactions is stored in electronic repositories and often utilized to construct interactome networks of model organisms, such as yeast [14],

Drosophila [15], *Bacillus subtilis* [16], *Caenorhabditis elegans* [17], the malaria parasite *Plasmodium falciparum* [18] and even humans [19, 20]. Among plants, the interactome of *Arabidopsis* will most probably be the first described. To date, the first *Arabidopsis* interactome fragments have been recently reconstructed, e.g., de Folter and colleagues [21] presented a plant interactome map of proteins from the *Arabidopsis thaliana* MADS box transcription factor family. This network fragment adds data on plants to a growing collection of available interaction maps for a number of different organisms.

Besides organism-specific databases on protein–protein interactions, several large repositories collect information on protein interactions in different organisms, or even more general, on all known biomolecular interactions of different types. One such major collection for data on experimentally verified protein interactions is the Database of Interacting Proteins (DIP [22]), which stores the information on more than 55,000 protein interactions in 110 different organisms (as of January 2006). The above-mentioned BIND compiles published information on more than 200,000 biomolecular interactions in 1,528 different organisms, including 1,537 interactions described for *Arabidopsis thaliana* (as of January 2006). Although the plant-related part of the BIND database remains relatively small (in BIND only 0.76% of all interaction records refer to plants) cataloguing and networking protein interactions is a rapidly expanding area with high gene function discovery potential. The success of such approaches depends on combined efforts of large scientific consortia and mapping of the *Arabidopsis* interactome has been included as an integrated component of the 2010 Project, aimed at determining the function of all genes in *Arabidopsis thaliana*.

In correlation networks nodes are genes (these networks are often termed also as gene coexpression networks, or just expression networks) or/and metabolites; two nodes are connected with non-directed edges, if patterns of changes in their expression/concentration correlate significantly to each other. Unlike in the previously described types of cellular networks, in correlation networks connections do not directly represent a physical interaction between nodes, but coexpression or co-behavior, under applied conditions. The items with similar patterns of co-behavior are usually considered to be more likely functionally associated, due to a variety of different biological reasons. These functional associations imply an exchange of information between items. The whole correlation network represents a sum of such associations, with the branching paths, along which the information is processed in order to finally accomplish endpoint biological reactions. Building of such correlation networks attempts to reconstruct real dynamic interacting networks of genes in the genetic regulatory circuitry. The approach seems to be adequate, as these real networks result *in vivo* in complex gene expression and metabolite concentration patterns.

The initial datasets for reconstruction of correlation networks are ‘omics’-scale profiles of gene expression and metabolite concentrations (what is often termed as transcriptome and metabolome, correspondingly). Current approaches to attain transcript and metabolic profiles are described in the previous chapters. Available collections of transcript profiles are already large and continue to grow rapidly and the necessity of such repositories for metabolic profiles is widely recognized. Major

repositories of genome-scale transcript profiles are compiled in Table 2. Some of these, for example M-CHiPS, NASCArrays or Genevestigator, provide convenient tools for data mining, acting as data warehouses rather than mere repositories. In several of these databases there exists the possibility for pair-wise correlation analysis. For example, utilizing NASCArrays one can build two gene scatter plots to compare expression patterns of two genes, or with another tool, Gene Correlator of Genevestigator repository, coexpression of two genes over a set of array chips can be visualized.

The potential for the analysis of coexpression for functional genetics has been already recognized in pre-genomic era [29, 30], tested experimentally and proved to be useful for decisions on functions of examined genes (e.g., [31, 32]). Later, when 'omics'-scale gene expression/metabolic concentration profiles became available, global analysis of pattern similarities began to be applied [33–35]. Approximately at the same time the first studies on functional genomics based on transcriptional correlations were carried out [36]. Since these pioneering studies systematic approaches for identifying the biological functions of novel genes have been widely applied, signifying an era of genome-wide functional analysis. Finally, matrices of pair-wise correlations across genome-scale arrays have been computed and global correlation networks were built from these correlation matrices. For example, Kim and co-workers assembled data from *Caenorhabditis elegans* DNA microarray experiments [37] involving multiple growth conditions, developmental stages, and varieties of mutants. In this study co-regulated genes were grouped together and visualized in an expression map that displayed correlations of gene expression profiles. Already in this early study of one of the first correlation networks their high potential in gene discovery was visualized demonstrating that it is possible to assign functions through identification of genes that are co-regulated with known sets of genes or even to uncover previously unknown genetic functions. Correlation network analysis has subsequently been applied to yeast, worm, fly and human, and combined analysis of all four allowed identification of global coexpression relationships and their evolutionary conservation [38]. Subsequent demonstrations of the high level of co-regulation conservation in the evolution of prokaryotes and eukaryotes [39] implies that functional relationships predicted from coexpression network analysis in one species can be transferred to another species.

As the next cognitive step alterations in coexpression relationships in two distinct coexpression networks have been studied [40]. With this approach it was possible to show, that functional changes such as alteration in energy metabolism, promotion of cell growth and enhanced immune activity were accompanied with coexpression changes. We shall discuss this approach in more detail below in a chapter devoted to network comparison.

Metabolite correlation networks can be exemplified by the studies of Weckwerth, Fiehn and colleagues [41, 42]. Unlike gene expression correlation networks, most metabolite correlation networks concern plant systems. Recently, given the availability of both metabolite and gene expression profiles, the use of cross-correlation analysis in search for functional gene-metabolite associations became possible. It has been demonstrated by fungal and plant biologists, that the integration of transcript

Table 2. Major repositories for genome-scale transcript profiles

Repository name [Reference #]	Web link	No. of Experiments (sample series)[a]	No. of hybridizations (samples, arrays, slides)[a]	Organisms	Comments
GEO – Gene Expression Omnibus [23]	http://www.ncbi.nlm.nih.gov/geo/	2,967	67,837	variable	367 experiments on *Arabidopsis*
ArrayExpress [24]	http://www.ebi.ac.uk/arrayexpress/	1,226	34,486	variable	no Arabidopsis data
SMD – Stanford Microarray Database	http://genome-www5.stanford.edu//		10,516	variable	630 hybridizations for *Arabidopsis*
M-CHiPS [25]	http://www.mchips.org/	40	316	variable	no Arabidopsis data
yMGV – yeast microarray global viewer [26]	http://transcriptome.ens.fr/ymgv/index.php	1,544		2 yeasts	
Expression Connection	http://db.yeastgenome.org/cgi-bin/expression/expressionConnection.pl			yeast	
Webminer	http://genome-www.stanford.edu/cgi-bin/webminer/mkjavascript			yeast	
NASCArrays	http://affymetrix.arabidopsis.info/narrays/experimentbrowse.pl	234		*Arabidopsis* and other species	
Genevestigator Database [27]		183	2,317	Arabidopsis	
AtGenExpress	http://web.uni-frankfurt.de/fb15/botanik/mcb/AFGN/atgenex.htm	47	1,387	Arabidopsis	
MAEDA/RARGE at RIKEN [28]	http://rarge.gsc.riken.go.jp/microarray/microarray.pl			Arabidopsis	
SGED – Solanaceae Gene Expression Database	http://www.tigr.org/tdb/potato/SGED_index2.shtml	33	1,072	Solanaceae	

[a] as of January 2006

and metabolic profiles can facilitate the identification of candidate genes for biotechnology [43, 44]. In subsequent studies, combined metabolomics and transcriptomics data were mined and clusters of co-regulated genes and metabolites were determined that displayed coordinated behavior under given experimental conditions [45, 46]. Finally, the entire network of gene-metabolite correlations has been reconstructed from combined sets of transcript and metabolic profiles [47]. From such reconstructions, a global network of information exchange in a living organism is revealed allowing prediction of master controllers of homeostasis. Weckwerth and Morgenthal [48] recently summarized what biologists can gain when analyzing metabolite correlation networks. From studies on network topology putative regulators of underlying processes can be identified as highly connected nodes, or hubs. Metabolic correlation networks can be further superimposed on biochemical reaction networks; through this analysis unexpected pleiotropic changes in genetically modified plants can be identified and assigned to those parts of metabolism which are influenced by genetic manipulation [49]. Knowledge gained from the analysis of gene expression correlation networks is based on the underlying assumption that identified clusters of co-expressed genes are co-regulated. Gene expression at the level of transcription is regulated by transcription factors which bind to specific regulatory sequences in the promoter regions of regulated genes. That many genes are co-regulated suggests the presence of common regulatory sequences in the promoters of clustered genes and makes their analysis a priority in network studies. The validity of such promoter analysis was realized in early studies on correlations of patterns of gene expression [50]. To understand combinatorial control of gene expression, hierarchical and modular organization of regulatory DNA sequence elements in the promoters of co-expressed genes has been examined [34]. For such studies global gene expression correlation networks can be of extreme use, as they intrinsically contain and process the information encoded by transcription networks. Modern research on transcriptomics coupled to promoter analysis has allowed the identification of novel transcription factor target genes [51] and putative regulatory motifs [52], elucidation and prediction of complex regulatory events [53].

Signaling networks are often distinguished as another type of molecular network [54, 55]. These networks represent signal transduction pathways, where nodes are proteins or small molecules, and directed links are signal transduction events. The basic knowledge for reconstruction of such networks comes from low-throughput experiments on individual molecules. Resulting signaling networks are usually assembled around a single signaling cascade, as, for example, the signaling network of bacterial chemotaxis [56] or multiple studies on cancer signaling (reviewed by [57, 58]). In this sense such signaling pathways may be regarded as subnetworks, or network fragments of a global signaling network. Nevertheless their complexity is high due to a big number of the involved elements, branching, feedforward and feedback regulations and cross-talk with other signaling cascades [59, 60].

In plant biology several signaling networks have also been resolved at the molecular level, for example the signaling network of the plant immune system [61] or hydrogen peroxide signaling network that mediates plant programmed cell death [62]. Such studies can be concentrated also on signaling molecules, which may be

common for several signaling pathways. For example, nitric oxide and hydrogen peroxide are key signaling molecules produced in response to various stimuli and involved in a diverse range of plant signal transduction processes. One such process is stomatal closure controlled by guard cell signal transduction. By the combined efforts of several laboratories the whole signaling network which controls stomatal closure is being assembled molecule by molecule. Through the analysis of this network in its spatial and temporal resolution a close interrelationship between the involved molecules have been identified [63–66].

In spite of the fact that common signaling molecules have been identified, the present state of knowledge cannot say how molecular information is processed through a network of interlacing signal transduction pathways. Reconstruction of a whole network of interlacing signaling cascades remains a challenging task. In this direction, there are attempts to assemble the whole signaling network, although still limited to single processes. For example, Janes and co-workers [67] constructed a systems model of 7,980 intracellular signaling events that links response outputs associated with apoptosis. Due to globality of the model, it was possible to predict multiple responses induced by a combination of factors.

In what are often called gene networks (or genetic regulatory networks), nodes are genes that are connected with arrowed links directed from gene A to gene B, if for example a mutation (perturbed expression level) in gene A leads to changed expression of gene B. Thus, gene networks show the phenomenological interactions between gene activities. Although in this approach only the transcriptome is considered, gene relationships are basically mediated by proteins and metabolites, and in this way all biochemistry underlying gene–gene interactions is implicitly present in gene networks. Besides network connectivity, regulatory strengths of gene–gene interactions can be quantified from experimental data and represented by, e.g., a thickness of a connecting edge (for example, by an approach suggested by [68]), introducing quantitative aspects to gene networks. Gene networks can be reconstructed from single gene perturbations, as was done, for example, by modulating activin in mice [69], human fibroblast response [70], or by perturbing the action of a key regulator of floral asymmetry in *Arabidopsis* [71]. If perturbations were applied to all genes in a genome, the global gene network of an organism would be uncovered. On the way to such globalization, the repositories of compiled information of single-gene mutations of ‘post-genome’ organisms and resulting databases of essential genes, like DEG [72] could be used.

As summarized by Chan with colleagues [70], reconstruction of gene networks from gene expression data is useful for:

1. identifying important genes in relation to a disease or a biological function
2. gaining an understanding on the dynamic interaction between genes
3. predicting gene expression values at future time points
4. predicting drug effects over time.

Currently less utilized are protein sequence similarity-based networks. In patterns of protein domains the latter are connected if appearing in genome sequences in combinations [73]. Protein domain universe graphs (PDUG) are constructed by

representing the nonredundant set of protein structural domains as nodes and using the structural similarity between those domains to define the edges on the graph [74].

Other types of biological networks usually represent an integration of the above-described network types in different combinations based on multiple datasets, representing any relationship between a set of genes, mRNAs, metabolites or proteins. New types of network can be generated by an enrichment of any of these networks with data from diverse genetic sources. For example, Garten and colleagues [75] superimposed transcription network and gene expression correlation network of yeast to filter out false positive associations from so-called location data on transcription factor proteins with their spectrum of promoter-binding sites determined *in vivo*. In yeast cellular network modelled by Yu and Li [76], data on transcription factor, gene relationships, microarray data and prior biological knowledge are integrated. As distinguishing features resulting from this integration, the combinatorial nature of transcription regulation, an estimate of transcription factor activity and condition specificity of the relationships are considered. Lu and co-workers [77] integrated initial yeast protein interaction network with diverse sources of genomic evidence, ranging from coexpression relationships to similar phylogenetic profiles. As a result, they observed measurable improvement in prediction performance of protein networks. In another approach undertaken by Patil and Nielsen [78] integration of genome-scale metabolic network and gene expression data enabled systematic identification of so-called reporter metabolites, important in metabolic regulation. It was possible to identify also the significantly correlated metabolic subnetworks after direct or indirect perturbations of the metabolism. de Lichtenberg and colleagues [79] used gene expression data from different stages of the yeast cell cycle, integrated it with a protein network and discovered that most of the protein complexes are comprized of both periodically and constitutively expressed proteins, which suggests that the former control complex activity by a mechanism of just-in-time assembly. Ihmels and co-workers [80] integrated large-scale expression data with the structural description of yeast metabolic network and found that only distinct branches at metabolic branchpoints are coexpressed and that individual isozymes were often separately co-regulated with distinct processes. Ideker and co-workers [81] inferred models of transcriptional regulation through integrating the data on protein–protein and protein–DNA interactions, the directionality of signal transduction in protein–protein interactions, as well as signs of the immediate effects of these interactions in what they call physical networks.

Obviously, the list of integrated networks has increased dramatically in the last two years alone and may be continued with almost any combination of data.

Types of representations of biological networks

With the use of high-throughput methods of modern biology the information on molecular interactions or co-behavior, cell regulation and signal transduction is rapidly accumulating. Although very complex by its nature, this data can be assembled in a simpler form of network graphs of interconnected elements. The informa-

tion contained in such graphs can be of varying precision, depending on the availability of underlying knowledge. For example, in the networks describing interactome edges are usually unambiguous: connection between two proteins represents the possibility of direct binding which has been experimentally proven. However, the symbols used in other network types may lack strict definitions (often reflecting a lack of exact knowledge). To illustrate this, Kitano and colleagues [82] give an example of a typical signal transduction diagram, in which an arrow symbol could be interpreted four different ways: activation, translocation, dissociation of protein complex and residue modification.

To be able to share and to exchange knowledge gained from network analysis, systems biologists need to 'speak the same language', i.e., apply similar sets of formalization rules in the process of building such networks. While, to date, no consensus has yet been reached several approaches such as that of Pirson and colleagues [83], who elaborated a simple symbolic representation set of 18 controls for signal transduction networks, have been attempted. This set of formalization rules was further extended by KW Kohn [84] to additionally cover protein interaction and transcription networks. The elaborated graphical method could deal with both 'heuristic' and 'explicit' diagrams. Heuristic diagrams are important to build networks, when detailed knowledge of all possible reaction paths is not available, while 'explicit' means that the diagrams are totally unambiguous and suitable for computer simulation. This work was a step forward in information standardization from human- to machine-readable form of representing and communicating biological networks. The innovation in this direction was the development of the Systems Biology Markup Language (SBML), an open XML-based format for representing biochemical reaction networks. With the help of SBML models common to research in many areas of computational biology, including cell signaling pathways, metabolic pathways, gene regulation networks and others can be described [85].

Network topologies

After it has become possible to assemble information around a biological system in the form of a network of molecular interactions, it's time now to get the knowledge on how the functioning of the entire biological system is accomplished by means of the analysis of assembled network. To make it clear why biologists need to study an assembly to understand a biosystem, an analogy with a comprehensive technical system consisting of a lot of pieces is often exploited. Indeed, to understand functioning of the entire biosystem from a sum of studies on functionality of individual molecules is similar to studying the ship components to obtain knowledge on how a ship retains buoyancy and moves in a desired direction. For conceiving the entire functioning of both systems, knowledge on functionality of separate components, although being absolutely necessary is not sufficient, it is rather a matter of assembling and interaction of the component parts. For biosystems these properties are indicated by the structure, or topology, of an assembled network.

Early topological studies of cellular networks revealed several common characteristic features. Assemblies of molecular interactions usually represent complex heterogeneous networks, with nests of more dense connections. These nests are recognized as network modules, allowing network fragmentation into functional subnetworks. Network structure often involves a hierarchy of levels.

Aspects of structure can be deduced from statistical analysis of several parameters of network topology, in particular a number of connections (connectivity) for network elements and connectivity degree distribution, the degree of node centralities, clustering coefficient, network diameter and average path length.

Connectivities

In a biological network representation two nodes are connected to each other by edges, if an information exchange between these nodes occurs. Each node may be connected to distinct numbers of other nodes. From multiple analyses of biological network topologies, it is well established that connectivities are distributed among nodes with high inhomogeneity: the majority of nodes have a small number of connections, while a minority have a big number of connections. In large networks, the probability function $P(k)$ for the connectivity degree k may follow a behavior, described by the formula $P(k) = Ak^{-\gamma}$, called a power law. In a logarithmic scale this function takes a shape of a line, with the slope reflected by γ. Such distribution of a connectivity degree means that none of the nodes can be chosen as a scale representative from connectivity degree of which the judgement on connectivities of the other nodes may be drawn. That is why the networks with such connectivity degree distribution are often referred to as scale-free networks. Scale-free property of large networks was first distinguished by Barabasi and Albert [86]. After that, numerous large networks were described as being scale-free. Among biological networks, approximate scale-freeness was detected for many systems including, among others, metabolic networks of 43 different organisms [87], a pattern of protein domain combinations occurring in 40 genomes [73] further expanded to a protein domain universe graph [74] and gene-metabolite correlation network of *Arabidopsis* [47].

Scale-free networks possess a set of universal properties. First, paths by which information from any node can reach any other node, are relatively short. This feature was called a ‘small-word’ property [88]. The consequence of this feature for topology of scale-free networks is their high density and relatively small diameter. This in turn, taken together with a vast number of weakly connected nodes, brings us to the next consequence that is high redundancy of network paths. This property is very important for network stability. Indeed, if information from one node can reach another node by many redundant paths, then the probability to break information exchange by disturbance of any casual node from these paths is low. This means that scale-free networks are very robust against casual disturbances [89]. High stress tolerance of biological systems can be deduced also from robustness of a scale-free network of stress information processing. However, this property has an evident underside. The network integrity can be easily disrupted by the disturbance

of highly connected nodes, called hubs. This determines the potential importance of elements with high numbers of connections in maintaining homeostasis of a biosystem. For biotechnology and biomedicine, such hubs represent target elements to influence system functioning. However, it has to be mentioned here that the latest well-defined studies on topologies of technological and biological networks clarify the relationship between scale-freeness and power law distribution and suggest that the connectivity degree distribution of many biological networks is often better described by distributions other than the popular power law. Affirmative conclusions, which are often deduced from scale-freeness of biological networks, have to be assessed critically for the quantitative understanding of complex biological processes [90, 91].

Centralities

The ranking of system elements (nodes) using centralities is another tool for estimating the importance, or influence strength, of a node. Such tools are mainly used in the analysis of social networks, where centrality measures are commonly described as indices of prestige, prominence, importance, and power – the four Ps [92]. Centrality is considered to weight indispensability of a node for information processing between distant nodes. A classical illustration implies a network of two clusters connected to each other with one node. This node is considered to be centrally positioned, or central. Although in a minimal case it may bear only two connections, one to each of the clusters (and thus is of low connectivity), it is nevertheless crucially important for keeping the integrity of the whole network. In terms of informational processing, information (a parcel) cannot be delivered from any node of one cluster to any other node of another cluster, bypassing the node which connects two clusters. Being central for information processing through the network, this node therefore is able to influence a lot of other nodes and consequently is of high importance for system functionality.

In network topology analysis, several centrality measures are utilized [93]. The degree centrality [94, 95] is interpreted as a measure of immediate influence. As opposed to connectivity, the degree centrality of a node considers not only a number of direct connections of this node, but also connectivities of its direct neighbors. Indeed, if a node has just a few connections, but through these connections is bound to a highly connected hub, then the probability of the information to be processed through this node is still high. The eigenvector centrality [96] can be considered as an extended degree centrality which is proportional to the sum of the centralities of the node's neighbors [93]. Another centrality measure, betweenness centrality [97], gives an estimation of how often a node appears on the way of an informational parcel between any two other nodes, and by this defines the control influence strength of the node whose centrality is being measured. Congenerous to this measure is the closeness centrality [94, 95], which in social networks is most frequently used to measure relative access to network resources and information, and can also be interpreted as measuring the degree of independence from others in the network

[98]. The subgraph centrality [93] characterizes the participation of each node in all subgraphs in a network, with smaller subgraphs having higher importance. To describe the centers of biological networks, further methods for geometric centrality measures were considered, namely excentricity, status, and centroid value that were originally used in the context of resource placement problems [99].

In biological networks the most important nodes are traditionally searched among those highest connected (hubs). However, this approach is not always successful, for example in the analysis of yeast protein interaction network the essentiality of a gene was poorly related to the number of interactors of the corresponding protein [100]. Centrality measures as an alternative to connectivity are increasingly attempted for this means. For example in the yeast protein interaction network, centrality of the genes was associated with the essential functions of the genes [101], and when compared with node connectivities, the ranking introduced by the subgraph centrality was more highly correlated with the lethality of individual yeast proteins [93]. Ma and Zeng [102] have identified the most central metabolites in a metabolic network by measuring the closeness centrality of the nodes, which correlated with the average path length. By the analysis of the betweenness centrality of protein domains in the graph of protein domain structures a gatekeeper protein domain, removal of which partitions the largest cluster into two large sub-clusters, was found. As was suggested, the loss of such gatekeeper protein domains in the course of evolution may be responsible for the creation of new fold families [103]. The centrality measure was recently also applied in biomedicine, where it helped to estimate, e.g., the importance of differentially expressed genes in lung cancer tissues [104], or the relevance of different mediators in the human immune cell network [105]. As was shown by a comparative study of protein interaction networks of three evolutionary distant eukaryotes: yeast, worm, and fly, the centrality of proteins had similar distributions; proteins that had a more central position in all three networks, regardless of the number of direct interactors, evolve more slowly and are more likely to be essential for survival [106].

By analogy with the connectivity degree distribution, which follows a power law in most large biological networks, Goh and co-workers [107] found that the betweenness centrality in biological scale-free networks also displays a power law distribution, and an exponent of this distribution can be used as a discriminating factor to classify the scale-free networks. Power law distribution was demonstrated also for the betweenness centrality values of protein domains in the graph of protein domain structures [103].

Clustering coefficient

The clustering coefficient is another statistical measure to characterize large networks. It quantifies the cohesiveness of the neighborhood of a node, in other words, how well connected the neighbors of a vertex in a graph are. In real networks it decreases with the vertex degree connectivity [108]. The clustering coefficient of a node is defined as the ratio between the number of edges linking nodes adjacent to

this node and the total possible number of edges among them [88]. In other words, the clustering coefficient quantifies how close the local neighborhood of a node is to being part of a clique, a region of the graph (a subgraph) where every node is connected to every other node [109].

Real networks are generally characterized by a high clustering coefficient [88, 110]. For biological networks, a high average clustering coefficient was found, for example, in protein interaction and metabolic networks [111, 112], indicating a high level of redundancy and cohesiveness [109]. In gene expression networks generated from large model-organism expression datasets the average clustering coefficient was also several orders of magnitude higher than would be expected for similarly sized scale-free networks [113].

The diversity of cohesiveness of local neighborhoods is characterized by averaging the clustering coefficients of nodes that have the same connectivity degree. The function resulting from this procedure was decreasing in metabolic networks [114] and protein interaction networks [112]. This suggests that low-degree nodes tend to belong to highly cohesive neighborhoods whereas higher-degree nodes tend to have neighbors that are less connected to each other [109].

As an example application, in the recent study by Wei and colleagues [115] clustering coefficient was used to find out the superior one of the two possible mechanisms of the tRNA sequences evolution, namely point mutation and complementary duplication. From comparison of clustering coefficients in two alternative networks, which were constructed, based on these two possible mechanisms it was concluded that modern tRNA sequences evolved primarily by the mechanism of complementary method, and point mutation is an important and indispensable auxiliary mechanism during the evolutionary event.

Network diameter

In a graph theory, a network diameter is a global metric of its structure. It is defined as the average path length among all nodes. Together with average path lengths, the network diameter is considered as a measure of systems functionality, like, for example, in a study of robustness and vulnerability of the p53 protein interaction network [116]. In another example, using the path of shortest length, Said and coworkers [117] identified that the toxicity-modulating proteins in yeast have more interactions with other proteins, leading to a greater degree of metabolic adaptation upon modulating the functioning of these proteins.

Considering dynamics in biological networks

As a biological system is alive and ever-changing, it functions in time, or dynamically. Dynamical behavior is its intrinsic property and implies dynamical behavior of its constituting elements. Networks, now widely applied for systems biology, may be analyzed statically, or may consider this dynamical behavior, depending on the network type and on the nature of the datasets underlying network reconstruc-

tion. For metabolic, transcription and protein interaction networks, usual representation as graphs reflects the static properties of a system. The standard approach to model network dynamics is through sets of coupled differential equations, describing how the concentrations of the various products evolve over time [118]. However, such a model requires knowledge of the various reaction rates and rate-order kinetics. To overcome this drawback temporal data can be integrated into these networks. For example de Lichtenberg and colleagues analyzed the dynamics of protein complexes during the yeast cell cycle by means of integration of temporal data on protein interactions and gene expression [79], revealing previously unknown components and modules. In modeling the dynamics of another type of initially static network, a metabolic network, large-scale biochemical systems approaches, such as the network thermodynamics theory, biochemical systems theory, metabolic control analysis, and flux balance analysis are used. P Ao [119] modeled dynamics of a metabolic network by adding four dynamical structure elements: potential function, translocation matrix, degradation matrix, and stochastic force. Network dynamics was determined by these four elements being in balance, which gave rise to a special stochastic differential equation. This allowed experimental data being displayed stochasticity which carried important biological information.

As opposed to the above-mentioned networks, which are static by the nature of underlying data utilized, correlation networks are built from temporal (or sometimes concentrational) series of transcript or/and metabolite profiles. This defines the dynamical property of a resulting correlation network, which can be analyzed by cluster analysis and the systematic search for characteristic patterns of gene expression associated with a state of interest [120–123]. The dynamical property can also be implemented into the analysis of static networks by integrating with dynamical network types, as was demonstrated, for example, by Guthke and co-workers [124] in studies of the kinetics of the immune response to bacterial infection. In another study on yeast transcriptional regulatory network, molecular interactions in the cellular transcription, translation, and degradation machineries were incorporated into dynamic mathematical models of the biochemical system by finding the most changed parameters from yeast oligonucleotide microarray expression patterns in cases where a phenotype difference existed between two samples [125]. On a genomic scale, the dynamics of a biological network was analyzed for multiple conditions in yeast by integrating transcriptional regulatory information and gene expression data [126]. In another approach, which we would call vertical integration, dynamics is implemented into a biological network by combining different levels of system description. Applicability and limitations of modeling the dynamics of cellular networks with this approach were demonstrated by Vilar and colleagues [127] on the lac operon of *Escherichia coli* as a prototype system. Here, three levels (molecular, cellular, and that of cell population) were integrated into a single model, and by this dynamical aspects of the system were captured.

Several mathematical and computational approaches have been suggested for determining the dynamics of regulatory networks: including linear [128] and nonlinear [129] models, time-series analysis [130, 131] and Bayesian networks of dependencies [132, 133]. The dynamics of a biological system can be investigated by

computing kinetic curves for molecular components (RNA, proteins) using the method of generalized threshold models [134]. A dynamic network model can also be deduced from a simple discrete model by postulating logical rules that formally summarize legacy data, as was demonstrated by plant biologists for interaction of the so-called ABC homeotic floral genes in *Arabidopsis* floral organ determination [135].

Generally, the highly nonlinear dynamics exhibited by genetic regulatory systems can be predicted by either of two important theoretical approaches: the continuous approach, based on reaction-kinetics differential equations, and the Boolean approach, based on difference equations and discrete logical rules [136, 137]. With these approaches biological systems can be characterized into an ordered regime where the system is robust against perturbations, and a chaotic regime where the system is extremely sensitive to perturbations. In a case study of HeLa cells its underlying genetic network appeared to operate either in the ordered regime or at the border between order and chaos but did not appear to be chaotic [138].

Causal directionality in biological networks

Causality is another characteristic feature of a dynamically functioning system. Depending on the nature of underlying type of an informational exchange biological networks can be either directed or undirected. Causal directionality in the biological networks is subject for reconstruction, when cause-and-effect relationship of the interactions between two components is well defined, e.g., the direction of metabolic flow from substrates to products in metabolic networks, the information flow from transcription factors to the genes that they regulate in transcription networks, propagation of signal transduction events in signaling networks, or influence on gene expression in gene networks. Such networks are causally directed. In undirected networks, such as protein interaction networks or protein sequence similarity-based networks, the relationships are mutually equidirectional. Some biological networks, although possessing intrinsic causal directionality, stay as undirected graphs, because edge directions are difficult or even not possible to identify. This applies to a great extent to networks reconstructed from high-throughput metabolic, proteomic or genomic analysis. As can be illustrated by gene coexpression networks, although genes with similar expression profiles are likely to regulate each other or be regulated by another common gene, from co-response analysis it is impossible to infer any notion of causality – which gene is regulated and which gene is regulating. However, if such networks are built from dynamic measurements of responses, which yield hierarchical information about causal relations in the underlying system, then causal relationships in these networks can be inferred. This approach was probed, for example, on hormone and insulin signaling using tyrosine residues phosphorylation data [139]. Similarly, response dynamics elucidates causality, when the information is used regarding the time lag between species at which the highest correlation was found [122]. In the new multiscale fuzzy clustering method fuzzy cluster centers can be used to discover causal relationships between

groups of co-regulated genes. With this method applied to gene expression data, a new regulatory relationship concerning trehalose regulation of carbohydrate metabolism in *Arabidopsis* was found [140]. In another example, causal directionality was implemented to gene-metabolite correlation network with the use of *a priori* knowledge on the molecule, which excites the systems response and can thus be considered as a 'cause'. In such network propagation of the information flow from the exciter to physiological endpoints can be followed [47]. To derive causal influences in cellular signaling networks, machine learning was applied to the simultaneous measurement of multiple phosphorylated protein and phospholipid components in thousands of individual primary human immune system cells. Perturbing these cells with molecular interventions drove the ordering of connections between pathway components [141].

The problem of causality in biological networks can be accessed also by means of integrating with directed networks. In a causal inference approach transcriptional regulatory networks of yeast were constructed using gene expression data, promoter sequences and information on transcription factor binding sites [142]. In this method identified active transcription factors provide the causal effect as 'treatments' measured quantitatively, and gene expression levels are viewed as 'responses'. In a study of the pheromone response in yeast, causal relationships were implemented into the non-directed network of protein–protein interactions by integrating with the directed networks of protein–DNA interactions and signal transduction [81].

Comparative network analysis

Now, as enormous amounts of data are available on molecular interaction networks, the next cognition step for system biologists implies new questions about network dynamics and evolution. These questions can be approached by means of network comparison. In such analysis communication networks for steady state and perturbation, or for organisms of different evolutionary distance in normal growth and in response to the same perturbing agent, can be compared. By comparing topologies of the resulting alternative communication networks constitutive and exciter-specific communication paths can be revealed, as well as hubs as specific controllers of the response development. Moreover, network comparisons can be used systematically to catalog conserved network regions, each representing a functionally homologous mechanism or pathway [143]. This approach also helps to resolve some technical aspects of network analysis. One of the major such problems is generally the high noise component in biological networks. This problem can be approached, for example, by comparing a network reconstructed from real data with a network built from the same dataset, subjected to shuffling procedure and thus assumed to be information-free. As a result of such comparison, noise component can be subtracted from the real data-based network. Comparative analysis of real networks also helps to address the problem of noise. Thus, by comparing networks drawn from different species or conditions [144–146], it was possible to reinforce the common signal present in both networks while reducing the noise component. Network

comparison was helpful also in separating true protein–protein and protein–DNA interactions from false positives [147], annotating interactions with functional roles and, ultimately, organizing large-scale interaction data into models of cellular signaling and regulatory machinery [148].

In biological applications, network comparison is becoming increasingly fruitful. We shall illustrate this with several examples. As was shown by the analysis of metabolic networks, comparison of network topologies for 43 organisms revealed hierarchical modularity in the network organization [114]. Pairwise comparison of protein interaction networks of bacteria and yeast allowed detection of evolutionarily conserved pathways [149] and significantly conserved protein complexes [150]. Further cross-species study of protein–protein interaction networks, now of worm, fly and yeast, revealed remarkable similarities in network structures [106], and identified previously not described protein functions and interactions [151]. Network comparison was applied also to gene coexpression networks. In cancer research, studies on two distinct coexpression networks: a tumor network and normal network showed that cancer affected many coexpression relationships accompanied with functional changes [40]. These case studies demonstrate that network comparisons provide essential biological information beyond what is gained from the analysis of separate networks.

A growing demand for statistical techniques and tools applicable for network comparison meets with a growing response by bioinformaticians. In this vein a technique for finding branching structure shared by a set of phylogenetic networks was recently introduced [152]. Kelley and co-workers [149] implemented a strategy for aligning two protein–protein interaction networks that combines interaction topology and protein sequence similarity, which was further developed into a PathBLAST tool for alignment of protein interaction network**s** [143]. Another tool called Cfinder allows finding overlapping dense groups of nodes in networks [153]. The reader will find the collection of corresponding links in Table 3 below.

Testing biological networks

Analysis of biological networks gives life to new global hypotheses on systems functionality and reductionist findings of novel molecular interactions. The reliability of these hypotheses will be based on the general reliability of the network reconstruction procedure. If among numerous findings revealed through network analysis a significant number matches with prior experimental knowledge, this can generally serve as a validation of the network analysis methodology employed. However this approach evidently cannot validate each individual finding and as such cannot substitute for wet-laboratory experimentation.

The use of *a priori* knowledge is best illustrated by the studies on the yeast integrated regulatory network. Its reliability was tested on datasets related to the pheromone response pathway, and the resulting model showed consistence with previous studies on the pathway [81]. Similarly, in the network model of bacteria and yeast protein complexes several of these complexes matched well with prior ex-

Table 3. Networking tools

Tool name [Reference #]	Designation	Web link
Pajek [157]	analysis of large networks	http://vlado.fmf.uni-lj.si/pub/networks/pajek/
Cytoscape [158]	visualizing molecular interaction networks and integrating these interactions with gene expression profiles and other state data	http://www.cytoscape.org/
VANTED [159]	Visualization and Analysis of Networks containing Experimental Data	http://vanted.ipk-gatersleben.de/
VisANT [160]	visualizing and analyzing many types of biological networks	http://visant.bu.edu/
BiNGO [161]	assessing overrepresentation of gene ontology categories in biological networks	http://www.psb.ugent.be/cbd/papers/BiNGO/
Centibin [162]	calculation and visualization of centralities for biological networks	http://centibin.ipk-gatersleben.de/
Cfinder [153]	finding overlapping dense groups of nodes in networks	http://angel.elte.hu/%7Evicsek/
PathBLAST [143]	alignment of protein interaction networks	http://www.pathblast.org/
TopNet [163]	comparing network topologies	http://networks.gersteinlab.org/genome/interactions/networks/core.html
CellDesigner [82]	diagrammatic network editing software	http://www.celldesigner.org/

perimental knowledge on complexes in yeast only and thus served for validation of the methodology [150]. In biomedical studies, the importance of identified hubs for network function was supported by the severe phenotypes exhibited by human patients and animal models when these genes were mutated [154].

Similarly the use of direct experimentation for validation of biological networks has also been applied to the yeast integrated regulatory network: whereby the knockout of genes and subsequent phenotyping confirmed the effects which were predicted by the network model [81]. Mutation has also been used strategically in cancer research in order to test the significance of the results drawn from the network analysis [116]. In the same study another method of experimental testing was tried, namely the effects of tumor inducing viruses were compared with those derived from network analysis. Protein interaction networks were tested by two hybrid experiments in which approximately half of 60 inferred interaction predictions were confirmed [151]. However, in spite of the general acceptance of the reductionist

methods of experimental confirmation in biology, the problem of testing the reliability of a reconstructed biological network cannot be fully approached by such methods for all network types. Where it is possible, network construction as the method for analysis of the entire system's functionality by means of assembling coherence between the elements in complex systems can be reliably tested by the assembly of an alternative network. The expected experiments on this may imply, e.g., analysis of information conductivity in a network reconstructed from the similar data source, but obtained on a system with a hub gene/protein knocked out, and therefore will lay in a field of network comparison. Here, matching of the predicted information conductivity to that one in an alternative network will work for confirmation of the reliability of the reconstructed network.

Intrinsic properties of biological networks – are there any?

Recent advances in networking studies allow a comparative analysis of many large networks of biological, social and technological nature (e.g., [153, 155, 156]). In these studies a question is asked on the existence of common properties for these large networks and systems they describe. It was found, that, while on the one hand, complex systems, indeed, share several common properties, on the other hand, each system is characterized by unique parameters. Identification of regularities being specific for biosystems may lead to better understanding of the uniqueness of life phenomenon and may imply also a practical interest in developing the new information technologies of complex systems management.

Software solutions for network visualization and analysis with useful links

Modern software networking tools can handle multiple data types from distinct technologies. Some of these tools are multifunctional developments for general networking studies like Pajek, Cytoscape, and VANTED. The others represent more specialized tools created for the analysis of separate network properties, like network centralities (Centibin), or overrepresented gene ontologies (BiNGO). Network comparison studies can be approached with PathBLAST and TopNet. The reader will find short descriptions of functionality and applicability for the major networking tools with the corresponding links and references in Table 3.

Furthermore, the set of software tools helpful in networking studies, which has been developed for pathway analysis, is given in Table 4. Among these tools, AraCyc, a collection of biochemical pathways described in *Arabidopsis*, is designated for the networking of plant biosystems.

Among the other useful software developments, in Table 5 we provide the list of those, which are the most commonly used as data sources for network reconstructions. The last two links are devoted to the universal networking language SBML and the data integration tool Pointillist.

Table 4. Pathways: databases and analysis tools

Tool/Database name [Reference #]	Designation	Link
KEGG PATHWAY [5]	collection of manually drawn pathway maps for the molecular interaction and reaction networks	http://www.genome.ad.jp/kegg/pathway.html
BioCyc [2]	collection of pathway/genome databases plus the BioCyc open chemical database	http://www.biocyc.org/
AraCyc [3]	biochemical pathway database for *Arabidopsis*	http://www.arabidopsis.org/tools/aracyc/
MetaCyc [4]	database of nonredundant, experimentally elucidated metabolic pathways	http://metacyc.org/
PaVESy [164]	Pathway Visualization Editing System	http://pavesy.mpimp-golm.mpg.de/PaVESy.htm
KnowledgeEditor [165]	interactive modeling and analyzing biological pathways based on microarray data	

Table 5. Databases of molecular interactions and other

Name [Reference #]	Designation	Web link
RegulonDB [7]	database on mechanisms of transcription regulation and operon organization in *Escherichia coli*	http://regulondb.ccg.unam.mx
GRID [166]	database of genetic and physical interactions in yeast, fly and worm	http://biodata.mshri.on.ca/grid
Ospray [167]	visualization of complex interaction networks	http://biodata.mshri.on.ca/osprey
BIND [8]	Biomolecular Interaction Network Database	http://www.bind.ca/Action
DIP [22]	Database of Interacting Proteins	http://dip.doe-mbi.ucla.edu/
PPI [19]	Human protein–protein interaction network database	http://141.80.164.19/neuroprot/ppi_search.php
KEGG [168]	Kyoto Encyclopedia of Genes and Genomes	http://www.genome.ad.jp/kegg/
DEG [72]	Database of Essential Genes	http://tubic.tju.edu.cn/deg/
SBML [85]	Systems Biology Markup Language	http://www.sbml.org/
Pointillist [169]	inferring the set of elements affected by a perturbation of a biological system	http://magnet.systemsbiology.net/software/Pointillist/

References

1. Nikiforova VJ, Kopka J, Tolstikov V, Fiehn O, Hopkins L, Hawkesford MJ, Hesse H, Hoefgen R (2005) Systems re-balancing of metabolism in response to sulfur deprivation, as revealed by metabolome analysis of *Arabidopsis* plants. *Plant Physiol* 138: 304–318
2. Karp PD, Ouzounis CA, Moore-Kochlacs C, Goldovsky L, Kaipa P, Ahren D, Tsoka S, Darzentas N, Kunin V, Lopez-Bigas N (2005) Expansion of the BioCyc collection of pathway/genome databases to 160 genomes. *Nucleic Acids Res* 33: 6083–6089
3. Zhang PF, Foerster H, Tissier CP, Mueller L, Paley S, Karp PD, Rhee SY (2005) MetaCyc and AraCyc. Metabolic pathway databases for plant research. *Plant Physiol* 138: 27–37
4. Krieger CJ, Zhang PF, Mueller LA, Wang A, Paley S, Arnaud M, Pick J, Rhee SY, Karp PD (2004) MetaCyc: a multiorganism database of metabolic pathways and enzymes. *Nucleic Acids Res* 32: D438–D442
5. Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M (1999) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res* 27: 29–34
6. Sweetlove L, Fernie AR (2005) Tansley Review: Regulation of metabolic networks. Understanding metabolic complexity in the systems biology era. *New Phytol* 168: 9–23
7. Salgado H, Gama-Castro S, Martinez-Antonio A, Diaz-Peredo E, Sanchez-Solano F, Peralta-Gil M, Garcia-Alonso D, Jimenez-Jacinto V, Santos-Zavaleta A, Bonavides-Martinez C et al. (2004) RegulonDB (version 4.0): transcriptional regulation, operon organization and growth conditions in *Escherichia coli* K-12. *Nucleic Acids Res* 32: D303–D306
8. Alfarano C, Andrade CE, Anthony K, Bahroos N, Bajec M, Bantoft K, Betel D, Bobechko B, Boutilier K, Burgess E et al. (2005) The Biomolecular Interaction Network Database and related tools 2005 update. *Nucleic Acids Res* 33: D418–D424
9. Shen-Orr SS, Milo RM, Alon U (2002) Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nature Genet* 31: 64–68
10. Lee TI, Rinaldi NJ, Robert F, Odom DT, Bar-Joseph Z, Gerber GK, Hannett NM, Harbison CT, Thompson CM, Simon I et al. (2002) Transcriptional regulatory networks in *Saccharomyces cerevisiae. Science* 298: 799–804
11. Zhong JH, Zhang HM, Stanyon CA, Tromp G, Finley RL (2003) A strategy for constructing large protein interaction maps using the yeast two-hybrid system: Regulated arrays and two-phase mating. *Genome Res* 13: 2691–2699
12. Cusick ME, Klitgord N, Vidal M, Hill DE (2005) Interactome: gateway into systems biology. *Hum Mol Gen* 14: R171–R181
13. Skipper M (2005) A protein network of one's own proteins. *Nature Rev Mol Cell Biol* 6: 824–825
14. Uetz P, Giot L, Cagney G, Mansfield TA, Judson RS, Knight JR, Lockshon D, Narayan V, Srinivasan M, Pochart P et al. (2000) A comprehensive analysis of protein–protein interactions in *Saccharomyces cerevisiae*. *Nature* 403: 623–627
15. Giot L, Bader JS, Brouwer C, Chaudhuri A, Kuang B, Li Y, Hao YL, Ooi CE, Godwin B, Vitols E et al. (2003) A protein interaction map of *Drosophila melanogaster*. *Science* 302: 1727–1736
16. Hoebeke M, Chiapello H, Noirot P, Bessieres P (2001) SPiD: a subtilis protein interaction database. *Bioinformatics 17: 1209–1212*
17. Li S, Armstrong CM, Bertin N, Ge H, Milstein S, Boxem M, Vidalain PO, Han JD, Chesneau A, Hao T et al. (2004) A map of the interactome network of the metazoan *C. elegans*. *Science* 303: 540–543

18. LaCount DJ, Vignali M, Chettier R, Phansalkar A, Bell R, Hesselberth JR, Schoenfeld LW, Ota I, Sahasrabudhe S, Kurschner C et al. (2005) A protein interaction network of the malaria parasite *Plasmodium falciparum*. *Nature* 438: 103–107
19. Stelzl U, Worm U, Lalowski M, Haenig C, Brembeck FH, Goehler H, Stroedicke M, Zenkner M, Schoenherr A, Koeppen S et al. (2005) A human protein–protein interaction network: a resource for annotating the proteome. *Cell* 122: 957–968
20. Rual J-F, Venkatesan K, Hao T, Hirozane-Kishikawa T, Dricot A, Li N, Berriz GF, Gibbons FD, Dreze M, Ayivi-Guedehoussou N et al. (2005) Towards a proteome-scale map of the human protein–protein interaction network. *Nature* 437: 1173–1178
21. de Folter S, Immink RGH, Kieffer M, Parenicova L, Henz SR, Weigel D, Busscher M, Kooiker M, Colombo L, Kater MM et al. (2005) Comprehensive interaction map of the Arabidopsis MADS box transcription factors. *Plant Cell 17: 1424–1433*
22. Salwinski L, Miller CS, Smith AJ, Pettit FK, Bowie JU, Eisenberg D (2004) The database of interacting proteins: 2004 update. *Nucleic Acids Res* 32: D449–451
23. Barrett T, Suzek TO, Troup DB, Wilhite SE, Ngau WC, Ledoux P, Rudnev D, Lash AE, Fujibuchi W, Edgar R (2005) NCBI GEO: mining millions of expression profiles – database and tools. *Nucleic Acids Res* 33: D562–566
24. Parkinson H, Sarkans U, Shojatalab M, Abeygunawardena N, Contrino S, Coulson R, Farne A, Garcia Lara G, Holloway E, Kapushesky M et al. (2005) ArrayExpress – a public repository for microarray gene expression data at the EBI. *Nucleic Acids Res* 33: D553–D555
25. Fellenberg K, Hauser NC, Brors B, Hoheisel JD, Vingron M. (2002) Microarray data warehouse allowing for inclusion of experiment annotations in statistical analysis. *Bioinformatics* 18: 423–433
26. Le Crom S, Devaux F, Jacq C, Marc P (2002) yMGV: helping biologists for yeast microarray data mining. *Nucleic Acid Res* 30: 76–79
27. Zimmermann F, Hirsch-Hoffmann M, Hennig L, Gruissem W (2004) GENEVESTIGATOR. *Arabidopsis* microarray database and analysis toolbox. *Plant Physiol* 136: 2621–2632
28. Seki M, Narusaka M, Ishida J, Nanjo T, Fujita M, Oono Y, Kamiya A, Nakajima M, Enju A, Sakurai T et al. (2002) Monitoring the expression profiles of 7000 *Arabidopsis* genes under drought, cold, and high-salinity stresses using a full-length cDNA microarray. *Plant J* 31: 279–292
29. Oliver S (1996) A network approach to the systematic analysis of yeast gene function. *Trends in Genetics* 12: 241–242
30. Hodgman TC (2000) A historical perspective on gene/protein functional assignment. *Bioinformatics* 16: 10–15
31. Blochzupan A, Decimo D, Loriot M, Mark MP, Ruch JV (1994) Expression of nuclear retinoic acid receptors during mouse odontogenesis. *Differentiation* 57: 195–203
32. Yamazaki M, Majeska RJ, Yoshioka H, Moriya H, Einhorn TA (1997) Spatial and temporal expression of fibril-forming minor collagen genes (types V and XI) during fracture healing. *J Orthopaedic Res* 15: 757–764
33. Cho RJ, Campbell MJ, Winzeler EA, Steinmetz L, Conway A, Wodicka L, Wolfsberg TG, Gabrielian AE, Landsman, Lockhart DJ et al. (1998) A genome-wide transcriptional analysis of the mitotic cell cycle. *Mol Cell* 2: 65–73
34. Zhang MQ (1999) Promoter analysis of co-regulated genes in the yeast genome. *Comput Chem* 23: 233–250
35. Eisen MB, Spellman PT, Brown PO, Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* 95: 14863

36. Chu S, DeRisi J, Eisen M, Mulholland J, Botstein D, Brown PO, Herskowitz I (1998) The transcriptional program of sporulation in budding yeast. *Science* 282: 699–705
37. Kim SK, Lund J, Kiraly M, Duke K, Jiang M, Stuart JM, Eizinger A, Wylie BN, Davidson GS (2001) A gene expression map for *Caenorhabditis elegans*. *Science* 293: 2087
38. Stuart JM, Segal E, Koller D, Kim SK (2003) A gene-coexpression network for global discovery of conserved genetic modules**.** *Science* 302: 249–255
39. Snel B, van Noort V, Huynen MA (2004) Gene co-regulation is highly conserved in the evolution of eukaryotes and prokaryotes. *Nucleic Acids Res* 32: 4725–4731
40. Choi JK, Yu US, Yoo OJ, Kim S (2005) Differential coexpression analysis using microarray data and its application to human cancer. *Bioinformatics* 21: 4348–4355
41. Kose F, Weckwerth W, Linke T, Fiehn O (2001) Visualizing plant metabolomic correlation networks using clique-metabolite matrices. *Bioinformatics* 17: 1198–1208
42. Steuer R, Kurths J, Fiehn O, Weckwerth W (2003) Observing and interpreting correlations in metabolomic networks. *Bioinformatics* 19: 1019–1026
43. Askenazi M, Driggers EM, Holtzman DA, Norman TC, Iverson S, Zimmer DP, Boers ME, Blomquist PR, Martinez EJ, Monreal AW et al. (2003) Integrating transcriptional and metabolite profiles to direct the engineering of lovastatin-producing fungal strains. *Nature Biotechnol* 21: 150–156
44. Urbanczyk-Wochniak E, Luedemann A, Kopka J, Selbig J, Roessner-Tunali U, Willmitzer L, Fernie AR (2003) Parallel analysis of transcript and metabolic profiles: a new approach in systems biology. *EMBO Rep* 4: 989–993
45. Hirai MY, Yano M, Goodenowe DB, Kanaya S, Kimura T, Awazuhara M, Arita M, Fujiwara T, Saito K (2004) Integration of transcriptomics and metabolomics for understanding of global responses to nutritional stresses in *Arabidopsis thaliana***.** *Proc Natl Acad Sci USA* 101: 10205–10210
46. Hirai MY, Klein M, Fujikawa Y, Yano M, Goodenowe DB, Yamazaki Y, Kanaya S, Nakamura Y, Kitayama M, Suzuki H et al. (2005) Elucidation of gene-to-gene and metabolite-to-gene networks in *Arabidopsis* by integration of metabolomics and transcriptomics. *J Biol Chem* 280: 25590–25595
47. Nikiforova VJ, Daub CO, Hesse H, Willmitzer L, Hoefgen R (2005) Integrative gene-metabolite network with implemented causality deciphers informational fluxes of sulphur stress response. *J Exp Bot* 56: 1887–1896
48. Weckwerth W, Morgenthal K (2005) Metabolomics: from pattern recognition to biological interpretation. *Drug Discov Today* 10: 1551–1558
49. Weckwerth W, Loureiro ME, Wenzel K, Fiehn O (2004) Differential metabolic networks unravel the effects of silent plant phenotypes. *Proc Natl Acad Sci USA* 101: 7809–7814
50. Spellman PT, Sherlock G, Zhang MQ, Iyer VR, Anders K, Eisen MB, Brown PO, Botstein D, Futcher B (1998) Comprehensive identification of cell cycle-regulated gene of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol Biol Cell* 9: 3273–3297
51. Vlieghe K, Vuylsteke M, Florquin K, Rombauts S, Maes S, Ormenese S, Van Hummelen P, Van de Peer Y, Inze D, De Veylder L (2003) Microarray analysis of E2Fa-DPa-overexpressing plants uncovers a cross-talking genetic network between DNA replication and nitrogen assimilation. *J Cell Sci* 116: 4249–4259
52. Liu FL, VanToai T, Moy LP, Bock G, Linford LD, Quackenbush J (2005) Global transcription profiling reveals comprehensive insights into hypoxic response in *Arabidopsis***.** *Plant Physiol* 137: 1115–1129
53. Venter M, Botha FC (2004) Promoter analysis and transcription profiling: Integration of genetic data enhances understanding of gene expression. *Physiol Plant* 120: 74–83

54. Xia Y, Yu HY, Jansen R, Seringhaus M, Baxter S, Greenbaum D, Zhao HY, Gerstein M (2004) Analyzing cellular biochemistry in terms of molecular networks. *Annu Rev Biochem* 73: 1051–1087
55. Schlitt T, Brazma A (2005) Modelling gene networks at different organisational levels. *FEBS Letters* 579: 1859–1866
56. Kollmann M, Lovdok L, Bartholome K, Timmer J, Sourjik V (2005) Design principles of a bacterial signalling network. *Nature* 438: 504–507
57. Bagnato A, Spinella F, Rosano L (2005) Emerging role of the endothelin axis in ovarian tumor progression. *Endocr Relat Cancer* 12: 761–772
58. Kundu JK, Surh YJ (2005) Breaking the relay in deregulated cellular signal transduction as a rationale for chemoprevention with anti-inflammatory phytochemicals. *Mutat Res – Fund Mol Mech Mut* 591: 123–146
59. Katagiri F (2004) A global view of defense gene expression regulation – a highly interconnected signaling network. *Curr Opin Plant Biol* 7: 506–511
60. Zhao J, Davis LC, Verpoorte R (2005) Elicitor signal transduction leading to production of plant secondary metabolites. *Biotech Adv* 23: 283–333
61. Feechan A, Kwon E, Yun BW, Wang YQ, Pallas JA, Loake GJ (2005) A central role for S-nitrosothiols in plant disease resistance. *Proc Natl Acad Sci USA* 102: 8054–8059
62. Gechev TS, Minkov IN, Hille J (2005) Hydrogen peroxide-induced cell death in *Arabidopsis*: Transcriptional and mutant analysis reveals a role of an oxoglutarate-dependent dioxygenase gene in the cell death process. *IUBMB Life* 57: 181–188
63. Murata Y, Pei ZM, Mori IC, Schroeder J (2001) Abscisic acid activation of plasma membrane Ca2+ channels in guard cells requires cytosolic NAD(P)H and is differentially disrupted upstream and downstream of reactive oxygen species production in abi1-1 and abi2-1 protein phosphatase 2C mutants. *Plant Cell* 13: 2513–2523
64. MacRobbie EAC (2002) Evidence for a role for protein tyrosine phosphatase in the control of ion release from the guard cell vacuole in stomatal closure. *Proc Natl Acad Sci USA* 99: 11963–11968
65. Mustilli AC, Merlot S, Vavasseur A, Fenzi F, Giraudat J (2002) *Arabidopsis* OST1 protein kinase mediates the regulation of stomatal aperture by abscisic acid and acts upstream of reactive oxygen species production. *Plant Cell* 14: 3089–3099
66. Bright J, Desikan R, Hancock JT, Weir IS, Neill SJ (2006) ABA-induced NO generation and stomatal closure in *Arabidopsis* are dependent on H2O2 synthesis. *Plant J* 45: 113–122
67. Janes KA, Albeck JG, Gaudet S, Sorger PK, Lauffenburger DA, Yaffe MB (2005) Systems model of signaling identifies a molecular basis set for cytokine-induced apoptosis. *Science* 310: 1646–1653
68. de la Fuente A, Brazhnik P, Mendes P (2002) Linking the genes: inferring quantitative gene networks from microarray data. *Trends Genetics* 18: 395–398
69. Mazhawidza W, Winters SJ, Kaiser UB, Kakar SS (2006) Identification of gene networks modulated by activin in L beta T2 cells using DNA microarray analysis. *Histol Histopathol* 21: 167–178
70. Chan ZSH, Kasabov N, Collins L (2006) A two-stage methodology for gene regulatory network extraction from time-course gene expression data. *Expert Systems with Applications* 30: 59–63
71. Costa MMR, Fox S, Hanna AI, Baxter C, Coen E (2005) Evolution of regulatory interactions controlling floral asymmetry. *Development* 132: 5093–5101
72. Zhang R, Ou HY, Zhang CT (2004) DEG, a Database of Essential Genes. *Nucleic Acids Res* 32: D271–D272

73. Apic G, Gough J, Teichmann SA (2001) Domain combinations in archaeal, eubacterial and eukaryotic proteomes. *J Mol Biol* 310: 311–325
74. Dokholyan NV, Shakhnovich B, Shakhnovich EI (2002) Expanding protein universe andits origin from the biological Big Bang. *Proc Natl Acad Sci USA* 99: 14132–14136
75. Garten Y, Kaplan S, Pilpel Y (2005) Extraction of transcription regulatory signals from genome-wide DNA – protein interaction data. *Nucleic Acids Res* 33: 605–615
76. Yu T, Li K-C (2005) Inference of transcriptional regulatory network by two-stage constrained space factor analysis. *Bioinformatics* 21: 4033–4038
77. Lu LJ, Xia Y, Paccanaro A, Yu H, Gerstein M (2005) Assessing the limits of genomic data integration for predicting protein networks. *Genome Res* 15: 945–953
78. Patil KR, Nielsen J (2005) Uncovering transcriptional regulation of metabolism by using metabolic network topology. *Proc Natl Acad Sci USA* 102: 2685–2689
79. de Lichtenberg U, Jensen LJ, Brunak S, Bork P (2005) Dynamic complex formation during the yeast cell cycle. *Science* 307: 724–727
80. Ihmels J, Levy R, Barkai N (2004) Principles of transcriptional control in the metabolic network of *Saccharomyces cerevisiae*. *Nat Biotechnol* 22: 86–92
81. Yeang CH, Ideker T, Jaakkola T (2004) Physical network models. *J Comput Biol* 11: 243–262
82. Kitano H, Funahashi A, Matsuoka Y, Oda K (2005) Using process diagrams for the graphical representation of biological networks. *Nat Biotechnol* 23: 961–966
83. Pirson I, Fortemaison N, Jacobs C, Dremier S, Dumont JE, Maenhaut C (2000) The visual display of regulatory information and networks. *Trends Cell Biol* 10: 404−408
84. Kohn KW (2001) Molecular interaction maps as information organizers and simulation guides. *Chaos* 11: 84−97
85. Hucka M, Finney A, Sauro HM, Bolouri H, Doyle JC, Kitano H, Arkin AP, Bornstein BJ, Bray D, Cornish-Bowden A et al. (2003) The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics* 19: 524–531
86. Barabasi AL, Albert R (1999) Emergence of scaling in random networks. *Science* 286: 509–512
87. Jeong H, Tombor B, Albert R, Oltvai ZN, Barabasi A-L (2000) The large-scale organization of metabolic networks. *Nature* 407: 651–654
88. Watts DJ, Strogatz SH (1998) Collective dynamics of 'small-world' networks. *Nature* 393: 440–442
89. Albert R, Jeong H, Barabási AL (2000) Error and attack tolerance of complex networks. *Nature* 406: 378–382
90. Stumpf MPH, Ingram PJ (2005) Probability models for degree distributions of protein interaction networks. *Europhysics Letters* 71: 152–158
91. Arita M (2005) Scale-freeness and biological networks. *J Biochem* 138: 1–4
92. Borgatti SP (1995) Centrality and AIDS. *Connections* 18: 112–115
93. Estrada E, Rodriguez-Velazquez JA (2005) Subgraph centrality in complex networks. *Physical Review E* 7105: 6103
94. Freeman LC (1979) Centrality in social networks conceptual clarification. *Soc Networks* 1: 215–239
95. Albert R, Jeong H, Barabasi AL (1999) Internet – Diameter of the World-Wide Web. *Nature* 401: 130–131
96. Bonacich P (1972) Factoring and weighting approaches to status scores and clique identification. *J Math Sociol* 2: 113–120

97. Freeman LC (1977) A set of measures of centrality based on betweenness. *Sociometry* 40: 35–41
98. Hagen G, Killinger DK, Streeter RB (1997) An analysis of communication networks among Tampa Bay economic development organizations. *Connections* 20: 13–22
99. Wuchty S, Stadler PF (2003) Centers of complex networks. *J Theor Biol* 223: 45–53
100. Coulomb S, Bauer M, Bernard D, Marsolier-Kergoat MC (2005) Gene essentiality and the topology of protein interaction networks. *Proc R Soc Lond [Biol]* 272: 1721–1725
101. Jeong H, Mason SP, Barabasi AL, Oltvai ZN (2001) Lethality and centrality in protein networks. *Nature* 411: 41–42
102. Ma HW, Zeng AP (2003) The connectivity structure, giant strong component and centrality of metabolic networks. *Bioinformatics* 19: 1423–1430
103. Dokholyan NV (2005) The architecture of the protein domain universe. *Gene* 347: 199–206
104. Wachi S, Yoneda K, Wu R (2005) Interactome-transcriptome analysis reveals the high centrality of genes differentially expressed in lung cancer tissues. *Bioinformatics* 21: 4205–4208
105. Tieri P, Valensin S, Latora V, Castellani GC, Marchiori M, Remondini D, Franceschi C (2005) Quantifying the relevance of different mediators in the human immune cell network. *Bioinformatics* 21: 1639–1643
106. Hahn MW, Kern AD (2005) Comparative genomics of centrality and essentiality in three eukaryotic protein-interaction networks. *Mol Biol Evol* 22: 803–806
107. Goh KI, Oh E, Jeong H, Kahng B, Kim D (2002) Classification of scale-free networks. *Proc Natl Acad Sci USA* 99: 12583–12588
108. Soffer SN, Vázquez A (2005) Network clustering coefficient without degree-correlation biases. *Physical Review E* 71: 057101
109. Albert R (2005) Scale-free networks in cell biology. *J Cell Sci* 118: 4947–4957
110. Albert R, Barabasi A-L (2002) Statistical mechanics of complex networks. *Rev Mod Phys* 74: 47–97
111. Wagner A, Fell DA (2001) The small world inside large metabolic networks. *Proc R Soc Lond [Biol]* 268: 1803–1810
112. Yook SH, Oltvai ZN, Barabási AL (2004) Functional and topological characterization of protein interaction networks. *Proteomics* 4: 928–942
113. Carter SL, Brechbuhler CM, Griffin M, Bond AT (2004) Gene co-expression network topology provides a framework for molecular characterization of cellular state. *Bioinformatics* 20: 2242–2250
114. Ravasz E, Somera AL, Mongru DA, Oltvai ZN, Barabasi AL (2002) Hierarchical organization of modularity in metabolic networks. *Science* 297: 1551–1555
115. Wei FP, Meng M, Li S, Ma HR (2006) Comparing two evolutionary mechanisms of modern tRNAs. *Mol Phylogenet Evol* 38: 1–11
116. Dartnell L, Simeonidis E, Hubank M, Tsoka S, Bogle IDL, Papageorgiou LG (2005) Robustness of the p53 network and biological hackers. *FEBS Letters* 579: 3037–3042
117. Said MR, Begley TJ, Oppenheim AV, Lauffenburger DA, Samson LD (2004) Global network analysis of phenotypic effects: protein networks and toxicity modulation in Saccharomyces cerevisiae. *Proc Natl Acad Sci USA* 101: 18006–18011
118. Voit E (2000) *Computational Analysis of Biochemical Systems*. Cambridge University Press, Cambridge
119. Ao P (2005) Metabolic network modelling: Including stochastic effects. *Computers & Chem Eng* 29: 2297–2303

120. Holstege FC, Jennings EG, Wyrick JJ, Lee TI, Hengartner CJ, Green MR, Golub TR, Lander ES, Young R (1998) Dissecting the regulatory circuitry of a eukaryotic genome. *Cell* 95: 717–728
121. Tavazoie S, Hughes JD, Campbell MJ, Cho RJ, Church GM (1999) Systematic determination of genetic network architecture. *Nat Genet* 22: 281–285
122. D'haeseleer P, Liang S, Somogyi R (2000) Genetic network inference: from co-expression clustering to reverse engineering. *Bioinformatics* 16: 707–726
123. Wagner A (2001) How to reconstruct a large genetic network from *n* gene perturbations in fewer than n^2 easy steps. *Bioinformatics* 17: 1183–1197
124. Guthke R, Moller U, Hoffmann M, Thies F, Topfer S (2005) Dynamic network reconstruction from gene expression data applied to immune response during bacterial infection. *Bioinformatics* 21: 1626–1634
125. Cavelier G, Anastassiou D (2005) Phenotype analysis using network motifs derived from changes in regulatory network dynamics. *Proteins* 60: 525–546
126. Luscombe NM, Babu MM, Yu HY, Snyder M, Teichmann SA, Gerstein M (2004) Genomic analysis of regulatory network dynamics reveals large topological changes. *Nature* 431: 308–312
127. Vilar JMG, Guet CC, Leibler S (2003) Modeling network dynamics: the lac operon, a case study. *J Cell Biol* 161: 471–476
128. Tegner J, Yeung MKS, Hasty J, Collins JJ (2003) Reverse engineering gene networks: Integrating genetic perturbations with dynamical modeling. *Proc Natl Acad Sci USA* 100: 5944–5949
129. Wahde M, Hertz J (2000) Coarse-grained reverse engineering of genetic regulatory networks. *Biosystems* 55: 129–136
130. Arkin A, Shen P, Ross J (1997) A test case of correlation metric construction of a reaction pathway from measurements. *Science* 277: 1275–1279
131. Remondini D, O'Connell B, Intrator N, Sedivy JM, Neretti N, Castellani GC, Cooper LN (2005) Targeting c-Myc-activated genes with a correlation method: Detection of global changes in large gene expression network dynamics. *Proc Natl Acad Sci USA* 102: 6902–6906
132. Friedman N, Linial M, Nachman I, Pe'er D (2000) Using Bayesian networks to analyze expression data. *J Comput Biol* 7: 601–620
133. Tamada Y, Kim S, Bannai H, Imoto S, Tashiro K, Kuhara S, Miyano S (2003) Estimating gene networks from gene expression data by combining Bayesian network model with promoter element detection. *Bioinformatics* 19: II227–II236
134. Tchuraev RN, Galimzyanov AV (2001) Modeling of actual eukaryotic control gene subnetworks based on the method of generalized threshold models. *Mol Biol* 35: 933–939
135. Espinosa-soto C, Padilla-Longoria P, Alvarez-Buylla ER (2004) A gene regulatory network model for cell-fate determination during *Arabidopsis* thalianal flower development that is robust and recovers experimental gene expression profiles. *Plant Cell* 16: 2923–2939
136. Shmulevich I, Dougherty ER, Kim S, Zhang W (2002) Probabilistic Boolean networks: a rule-based uncertainty model for gene regulatory networks. *Bioinformatics* 18: 261–274
137. Ramo P, Kesseli J, Yli-Harja O (2005) Stability of functions in Boolean models of gene regulatory networks. *Chaos* 15: 34101
138. Shmulevich I, Kauffman SA, Aldana M (2005) Eukaryotic cells are dynamically ordered or critical but not chaotic. *Proc Natl Acad Sci USA* 102: 13439–13444
139. Kam Z (2002) Generalized analysis of experimental data for interrelated biological measurements. *Bull Math Biol* 64: 133–145

140. Du P, Gong H, Wurtele ES, Dickerson JA (2005) Modeling gene expression networks using fuzzy logic. *IEEE T Syst Man Cy B* 35: 1351–1359
141. Sachs K, Perez O, Pe'er D, Lauffenburger DA, Nolan GP (2005) Causal protein-signaling networks derived from multiparameter single-cell data. *Science* 308: 523–529
142. Xing B, van der Laan MJ (2005) A causal inference approach for constructing transcriptional regulatory networks. *Bioinformatics* 21: 4007–4013
143. Kelley BP, Yuan BB, Lewitter F, Sharan R, Stockwell BR, Ideker T (2004) PathBLAST: a tool for alignment of protein interaction networks. *Nucleic Acids Res* 32: W83–W88
144. Forst CV, Schulten K (1999) Evolution of metabolisms: a new method for the comparison of metabolic pathways using genomics information. *J Comput Biol* 6: 343–360
145. Ogata H, Fujibuchi W, Goto S, Kanehisa M (2000) A heuristic graph comparison algorithm and its application to detect functionally related enzyme clusters. *Nucleic Acids Res* 28: 4021–4028
146. Dandekar T, Schuster S, Snel B, Huynen M, Bork P (1999) Pathway alignment: application to the comparative analysis of glycolytic enzymes. *Biochem J* 343: 115–124
147. von Mering C, Krause R, Snel B, Cornell M, Oliver SG, Fields S, Bork P (2002) Comparative assessment of large-scale data sets of protein–protein interactions. *Nature* 417: 399–403
148. Ideker T, Lauffenburger DA (2003) Building with a scaffold: emerging strategies for high- to low-level cellular modeling. *Trends Biotechnol* 21: 255–262
149. Kelley BP, Sharan R, Karp RM, Sittler T, Root DE, Stockwell BR, Ideker T (2003) Conserved pathways within bacteria and yeast as revealed by global protein network alignment. *Proc Natl Acad Sci USA* 100: 11394–11399
150. Sharan R, Ideker T, Kelley B, Shamir R, Karp RM (2005) Identification of protein complexes by comparative analysis of yeast and bacterial protein interaction data. *J Comput Biol* 12: 835–846
151. Sharan R, Suthram S, Kelley RM, Kuhn T, McCuine S, Uetz P, Sittler T, Karp RM, Ideker T (2005) Conserved patterns of protein interaction in multiple species. *Proc Natl Acad Sci USA* 102: 1974–1979
152. Choy C, Jansson J, Sadakane K, Sung WK (2005) Computing the maximum agreement of phylogenetic networks. *Theor Comput Sci* 335: 93–107
153. Palla G, Derenyi I, Farkas I, Vicsek T (2005) Uncovering the overlapping community structure of complex networks in nature and society. *Nature* 435: 814–818
154. Clipsham R, Zhang YH, Huang BL, McCabe ERB (2002) Genetic network identification by high density, multiplexed reversed transcriptional (HD-MRT) analysis in steroidogenic axis model cell lines. *Mol Genet Metab* 77: 159–178
155. Girvan M, Newman MEJ (2002) Community structure in social and biological networks. *Proc Natl Acad Sci USA* 99: 7821—7826
156. Barabasi AL, de Menezes MA, Balensiefer S, Brockman J (2004) Hot spots and universality in network dynamics. *Eur Physical J B* 38: 169–175
157. Batagelj V, Mrvar A (2003) Pajek – Analysis and visualization of large networks. In: M Jünger, P Mutzel (eds): *Graph Drawing Software.* Springer, Berlin, 77–103
158. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T (2003) Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res* 13: 2498–2504
159. Junker BH, Klukas C, Schreiber F (2006) VANTED: A system for advanced data analysis and visualization in the context of biological networks. *BMC Bioinformatics* 7: 109
160. Hu Z, Mellor J, Wu J, Yamada T, Holloway D, DeLisi C (2005) VisANT: data-integrating visual framework for biological networks and modules. *Nucleic Acids Res* 33: W352–W357

161. Maere S, Heymans K, Kuiper M (2005) BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* 21: 3448–3449
162. Koschützki D, Lehmann KA, Peeters L, Richter S, Tenfelde-Podehl D, Zlotowski O (2005) Centrality Indices. In: U Brandes, T Erlebach (eds): *Network Analysis. LNCS Tutorial 3418.* Springer, 16–61
163. Yu HY, Zhu XW, Greenbaum D, Karro J, Gerstein M (2004) TopNet: a tool for comparing biological sub-networks, correlating protein properties with topological statistics. *Nucleic Acids Res* 32: 328–337
164. Ludemann A, Weicht D, Selbig J, Kopka J (2004) PaVESy: Pathway visualization and editing system. *Bioinformatics* 20: 2841–2844
165. Toyoda T, Konagaya A (2003) KnowledgeEditor: a new tool for interactive modeling and analyzing biological pathways based on microarray data. *Bioinformatics 19: 433–434*
166. Breitkreutz BJ, Stark C, Tyers M (2003) The GRID: the General Repository for Interaction Datasets. *Genome Biol* 4: R23
167. Breitkreutz BJ, Stark C, Tyers M (2003) Osprey: A network visualization system. *Genome Biol* 4: R22
168. Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M (2004) The KEGG resources for deciphering the genome. *Nucleic Acids Res* 32: D277–D280
169. Hwang D, Rust AG, Ramsey S, Smith JJ, Leslie DM, Weston AD, Atauri PD, Aitchison JD, Hood L, Siegel AF et al. (2005) A data integration methodology for systems biology. *Proc Natl Acad Sci USA* 102: 17296–17301