Computational methods in cancer gene networking

Edwin Wang

1. Biotechnology Research Institute, National Research Council of Canada, 6100 Royalmount Avenue, Montreal, Quebec, Canada, and 2. Center for Bioinformatics, McGill University, Montreal, Quebec, Canada

*To whom correspondence should be addressed.
Email: edwin.wang@cnrc-nrc.gc.ca
Fax: 1-514-496-5143
Tel: 1-514-496-0914

Biology is a science about relationships

In the past few years, many high-throughput techniques have been developed and applied to biological studies. These techniques such as "next generation" genome sequencing, chip-on-chip, microarray and so on can be used to measure gene expression and gene regulatory elements in a genome-wide scale. Moreover, as these technologies become more affordable and accessible, they have become a driving force in modern biology. As a result, huge amount biological data have been produced, with the expectation of increasing number of such datasets to be generated in the future. High-throughput data are more comprehensive and unbiased, but 'real signals' or biological insights, molecular mechanisms and biological principles are buried in the flood of data. In current biological studies, the bottleneck is no longer a lack of data, but the lack of ingenuity and computational means to extract biological insights and principles by integrating knowledge and high-throughput data.

To develop effectively computational tools, we must first understand what biology is. Biology deals with many kinds of relationships among genes, proteins, RNAs, cells, tissues, organs and environmental factors. For example, biological relationships include those encompassing gene regulatory, protein interaction, activation, genetic interaction, inhibitory and so on. In one word, biology is a science about relationships. Traditionally, biologists describe these relationships between a limited number of genes or proteins using a descriptive language. With the huge amount of data produced by high-throughput techniques, biologists have to deal thousands of biological relations in a single experiment. In this situation, the traditionally descriptive ways for biological relations are not sufficient to deal with the huge number of relations under study. The only way to deal with a large amount of relations is through mathematical representations and computations. Taking into account that the readers of this chapter are biologists, I would like to first introduce basic computational concepts and then illustrate the procedures and computational techniques for high-throughput data analysis, using examples from cancer research.

Network biology: concepts and biological meanings

Graph theory, a branch of mathematics, is designed to describe all kinds of relationships and complexity in mathematical language. Networks represent the principles of graph theory. Thus, it is reasonable to use networks to represent biological relationships. By doing so, we are able to transform the biological/descriptive language into a mathematical language, which is computable and capable of handling a large number of relationships. In summary, network biology involves the use of networks to represent complexity, computes biological relationships and seeks to uncover biological principles and insights. We can ask inspiring and fundamental questions and develop elegant computational methods using the principles of networks and statistics, which lead to new biological insights from high-throughput data. In turn, insightful results from these analyses can be used to ask new questions and design wet lab experiments.

Types and categories of cellular networks

Several types of cellular networks have been found in cells: protein interaction networks, metabolic networks, gene regulatory networks and signaling networks, genetic interaction network and gene co-expression network (Wang et al. 2007), which can be further classified into two categories based on whether the networks capture biological relations in genome-wide or a specific cellular condition. Protein interaction networks encode the information of proteins and their physical interactions. Protein interaction information in the network ranges from basic cellular machinery such as protein complexes for DNA synthesis, metabolic enzyme complexes, transcription factor complexes, to protein complexes involved in cellular signaling. A genome-wide protein interaction network encodes all the protein interaction information across all biological processes in a cell. A gene regulatory network describes regulatory relationships between transcription factors and genes. Similar to protein interaction networks, a gene regulatory network encodes the gene regulatory information for all biological processes and activities in a cell. Genetic interaction networks describe genetic interactions between genes that lead to certain phenotypes. A genetic interaction network encodes the genetic interaction information for all kinds of cellular phenotypes. Therefore, I have classified protein interaction networks, genetic interaction networks and gene regulatory networks

into the first category: general cellular network. The second category of networks identified is: cellular specific network, which encompasses metabolic networks, gene co-expression networks and signaling networks, describing relations in specific cellular conditions or specific cellular activities such as signaling and metabolism. A metabolic network collects all the metabolic reactions and metabolic flows, while a signaling network encodes signal information flows and biochemical reactions for signal transductions. Traditionally, both types of information are presented using pathways, e.g., metabolic pathways and signaling pathways. In a metabolic network, metabolic pathways are intertwined so that metabolic flows are transferable across different pathways. Certain metabolites can be shared and used by many different pathways, while certain end-product metabolites are able to be produced via bypassing one or several pathways. Signaling networks illustrate inter- and intracellular communications and information processing between signaling proteins. In fact, pathway concept gets fuzzy and many pathways lose their identities in networks (Patil et al. 2005;Spirin et al. 2006). Gene co-expression networks capture the co-expressed genes during certain cellular conditions. Co-expressed genes often represent a collection of genes that are involved in similar biological functions and activities. In addition, other types of networks can be constructed based on different biological contexts, i.e., disease-gene networks, drug-target networks, amino acid networks, gene-environment-factor network and so on. Nevertheless, these networks are represented at either a genome scale or under certain cellular conditions or specific cellular activities such as signaling.

Biological functions and mechanisms are encoded in network properties

A particular phenotype is the result of collaborations of a group of genes. This notion provides a structured network knowledge-based approach to analyze genome-wide data in the context of known functional interrelationships among genes, proteins and phenotypes. The biological relations and complexity are encoded in cellular networks (Wang et al. 2007). Therefore, a network view, or a systems-level view of cellular events emerges as an important concept. Cellular networks can be presented as either directed or undirected graphs. Usually in these networks, nodes represent proteins or genes and the links represent the physical interactions between proteins, gene regulatory relationships,

genetic interactions, gene co-expressions, or activation/inactivation signaling reaction relationships. Notably, signaling networks contain the most complicated relationships between proteins, e.g., nodes might represent different functional proteins such as kinases, growth factors, ligands, receptors, adaptors, scaffolds, transcription factors and so on, which all have different biochemical functions and are involved in many different types of biochemical reactions that characterize a specific signal transduction machinery (Wang et al. 2007).

The topology or architecture of cellular networks is 'scale-free', meaning that the network contains a small fraction of nodes acts as highly connected hubs, whereas most nodes have only a few links. As an example, an air transportation map is an analogy for the scale-free network. In the United States air transportation map, only a few big airports (hubs) in big cities such as Boston, New York, Chicago and Los Angles have many air routes (links) to other airports, while many small airports just have a few air routes to the nearby big airports. This common structural feature encodes a special property of these networks: they are robust but also very vulnerable to failure and attack (Barabasi et al. 1999;Barabasi et al. 2004). Randomly removal of a substantial fraction of the low-linked nodes will result in little damage to the network's connectivity and function, however, targeted removal of the hub nodes will easily disconnect and destroy the network completely. As illustrated by the air transportation map, disabling big airports (hubs) will wreak havoc in many ways, while damaging a few small airports will have little or no effect on overall air transportation.

Hub genes in gene regulatory networks are generally global transcription factors which govern a large number of genes in response to internal and external signals. Furthermore, mRNAs of the hub transcription factors have significantly faster decay rates than non-hubs in *Escherichia coli* and yeast gene regulatory networks (Batada et al. 2006;Wang et al. 2005), suggesting that hub transcription factors facilitate a rapid response of the network to external stimuli (Wang et al. 2005). In protein interaction networks, hub proteins take part in many biological processes and play a more important role in an organism's survival. Removal of hub proteins from an organism would have a

much broader effect on the organism than non-hubs (Babu et al. 2004;Jeong et al. 2001;Wuchty et al. 2003;Wuchty et al. 2006). Hub proteins are more evolutionarily conserved than non-hubs (Saeed et al. 2006). One explanation of these phenomena is that hub proteins are subject to selection pressure and constraints, due to their involvements in many biological processes and their multiple interacting protein partners. In signaling networks, hub proteins are the proteins most commonly used by multiple signaling pathways. They become information exchanging and processing centers of the network (Wang et al. 2007). Hubs are one of the global properties of networks. Other global network properties include: network diameter, shortest path, network density, average links, clustering coefficient, network centrality (degree centrality, closeness centrality, radiality, betweenness and pageRank), minimum spanning trees, network flows, and network bottle necks. These measures have been described in details in recent reports (Mason et al. 2007;Yu et al. 2007).

A complex network can be broken down into distinct regulatory patterns or basic units. Thus, local network properties can be defined. Network motif is such an example. Network motifs are the smallest functional modules in networks. These motifs are the statistically significant recurring structural patterns or small subgraphs or sub-networks that are found more often in a real network than would be expected by chance (Shen-Orr et al. 2002). These types of motifs have also been known as gene regulatory loops in biology. In gene regulatory networks, three major motifs are found in gene regulatory networks (Figure 1): Single Input Module (SIM), bi-fan and Feedforward Loop (FFL).

Positive feedback loops lean to emergent network properties such as ultrasensitivity, bistability and switch-like behavior, while negative feedback loops perform adaptation, desensitization, and preservation of homeostasis (Ferrell 2002). These motifs are enriched with the transcription factors whose mRNAs have fast decay rates, suggesting that motif structures encode a regulatory behavior: they are able to rapidly respond to internal and external stimuli and decrease cell internal noise (Wang et al. 2005). Network motifs also bear distinct regulatory functions and particular kinetic properties that determine the temporal program of gene expression (Mangan et al.

2003b). Thus, the frequencies and types of network motifs with which cells use indicate the regulatory strategies that are selected under different cellular conditions (Balazsi et al. 2005;Wang et al. 2005). For example, FFLs are buffers that respond only to persistent input signals (Mangan et al. 2003a), which makes them well-suited for responding to endogenous conditions, while the motifs whose key regulator's transcripts have a fast mRNA decay rate are preferentially used for responding to exogenous or environmental conditions (Wang et al. 2005). In signaling networks, network motifs such as switches (Bhalla et al. 2002), gates (Blitzer et al. 1998), and positive or negative feedback loops provide specific regulatory capacities in decoding signal strength, processing information and controlling noise (Dublanche et al. 2006). Network motifs often form large aggregated structures, called network themes that perform specific functions by forming collaborations between a large number of motifs (Zhang et al. 2005). A higher level of aggregation of network themes can be regarded as network modules.

A practical guide for network analysis in cancer biology

Cancer is an extremely complex, heterogeneous disease and represents a typical example of biology gone awry. More and more efforts in "omics" are being made in the cancer research community. As a result, tremendous amount of money has been poured into cancer research in the past few years. Relatively speaking, more high-throughput data have been generated in cancer biology than in any other field of biology. However, the complexity of cancer forms a major obstacle for a comprehensive understanding of underlying molecular mechanisms of tumorigenesis. To crack the cancer code, computational methods have been developed and applied to cellular networks of cancer. Using a few examples of cancer based studies, I will demonstrate the procedures and computational analysis involved in cancer cell networks.

Data sources and quality for cancer cell networks

Public databases collect and assemble literature-mined datasets describing human protein interactions, metabolic and signaling pathways, cancer driving mutated genes such as COSMIC (http://www.sanger.ac.uk/genetics/CGP/cosmic/), and tumor gene expression profiles such as Oncomine (http://www.oncomine.org/), human protein

interaction database, HPRD (http://www.hprd.org/), IntAct (http://www.ebi.ac.uk/intact/site/index.jsf), MINT (http://mint.bio.uniroma2.it/mint/Welcome.do) and DIP (http://dip.doe-mbi.ucla.edu/), signaling pathway database, BioCarta (http://www.biocarta.com/) and Reactome (http://reactome.org/, and so on. Datasets can be found in research articles that conduct high-throughput studies such as microarray profiling of tumors, genome-wide sequencing of tumor samples. Useful datasets are also found in research articles that manually curate data from literature for constructing cellular networks (Oda et al. 2005;Oda et al. 2006), i.e., we have manually curated a large human signaling network containing more than 1,600 proteins and 5,000 signaling relations (Cui et al. 2007). When using these datasets, the quality of data should be carefully examined. For example, it is well-known that false positives are in the protein interaction data derived from high-throughput studies. Relevant computational methods have been developed to clean these false positives as much as possible (Mahdavi et al. 2007). However, it is still a challenge task. On the other hand, data are often incomplete. To overcome these problems, sensitivity analysis can be applied: we can mimic false positives and false negatives by randomly adding or removing extra 10 and 20% of network nodes into the networks, and then re-examine the analysis (Cui et al. 2006).

Network construction and visualization

Networks can be constructed using the information contained in relationships and other information. Genes and proteins are often used as nodes, relations between them are edges (undirected links) or arcs (directed links). In gene regulatory network, links are directed and represent regulatory relationships. In protein interaction networks, links are undirected links and represent physical interactions between proteins. In signaling networks, there are undirected and directed links, furthermore, signs (i.e., activating and inhibiting) can be assigned to the directed links (Wang et al. 2007). In certain situation, links can be assigned numbers such as protein interacting affinity. The resulting network is called weighted network. Free software programs such as Pajek (http://vlado.fmf.uni-lj.si/pub/networks/pajek/) and Cytoscape (http://www.cytoscape.org/) are available to visualize networks.

Data integration onto networks

The integration of experimental techniques with the information technology provides a powerful approach to address and dissect the complexity of cancer and other biological problems at various levels in a systems-based manner.

It is often necessary to integrate different sources of data onto the network to perform analysis. The selection of data sources are based on the questions asked. For example, we integrated cancer driver mutated genes determined by large scale genome sequencing of tumors onto a human signaling network to understand cancer signaling mechanisms (Cui et al. 2007). Chuang et al. integrated tumor gene microarray expression profiles onto a human protein interaction network to find network clusters as breast cancer biomarkers (Chuang et al. 2007).

Network property analysis

Global and local network properties can be analyzed. Almost all of the network properties are analyzed using statistical methods. Global network properties can be investigated and drawn biological insights. For example, by integrating cancer driver mutated genes and cancer associated methylated genes onto a human signaling network, we uncovered that cancer causal mutations most likely occur in signaling proteins that are acting as signaling hubs (i.e. actively sending or receiving signals) rather than in nodes that are simply involved in passive physical interactions with other proteins (Cui et al. 2007). These results can be interpreted in light of the fact that since signaling hubs are focal nodes that are shared by, and/or are central in, many signaling pathways, alterations of these nodes, or signaling hubs, are predicted to affect more signaling events, resulting in cancer or other diseases. We further showed that activating and inhibitory network flows enhance and alleviate oncogenic signaling flows, respectively. Network node connectivity is correlated with cancer biomarkers, i.e., a protein which has more links with cancer mutated genes, has a higher chance to be cancer-associated genes or biomarker (Cui et al. 2007). We also found that the downstream cancer mutating genes of the network, especially the genes of the output layer of the network, have a higher

mutation frequency, which complements with another finding that cancer-associated genes are enriched in the nuclear proteins (Awan et al. 2007). In contrast, the distributions of the cancer associate methylated genes have no such preference, suggesting that genes subject to genomic silencing do not tend to directly affect the output layer of the network. These results strongly suggest that the genes in the output layer of the network, which play direct and important roles in determining phenotypic outputs, are frequent targets for activating mutations. The importance of this output layer is reinforced by another observation that the expression of the output layer genes of the signaling network is heavily regulated by microRNAs (Cui et al. 2006).

A study of integration of literature-mined cancer genes onto a human protein interaction network showed that cancer proteins have more interaction partners than other proteins in the network, suggesting that cancer proteins may be involved in significantly more biological processes and play a central role in the protein network (Jonsson et al. 2006). As another example, Platzer et al. systematically investigated 22 individual network measures for the cancer gene networks constructed by combining a human protein interaction network with tumor-associated differentially expressed genes (Platzer et al. 2007). Network measures rang from closeness centrality, network diameter, index of aggregation, assortative mixing coefficient, connectivity and sum of the Wiener number to modified vertex distance number, representing several key network properties: size, distribution, relevance, density, modularity, and cycles. The analysis showed that genes showing significant differential expressions in cancer appear to be interlinked on the human protein interaction network. These cancer gene networks show a low density, indicating that they bear high robustness.

Biological insights are also encoded in local network properties. We have uncovered that cancer associated mutated and methylated genes are enriched in positive and negative regulatory loops, respectively. These results uncovered not only an overall picture of the network architecture where the oncogenic stimuli are embedded but also the signaling regulatory mechanisms involving cancer mutated and methylated genes (Cui et al. 2007). Furthermore, cancer genes are enriched in the convergent target nodes

of most network motifs, and form network motif clusters or cancer gene hotspots on the network (Awan et al. 2007). We have extracted a giant subgraph, which we named oncogenic map, containing connected cancer mutated and methylated genes (Cui et al. 2007). In the map, we further defined three network regions, which are enriched in cancer dependent signaling events, based on the overall frequency of cancer associated gene mutations. These results uncover an overall cancer signaling network architecture and highlight the most frequently used cancer signaling cascades. The map was further decomposed into 12 network community modules with different signaling functions: cell surface receptor-linked signaling, intracellular signaling cascades and apoptotic signaling. Similarly, cancer genes on protein networks also form network communities (Jonsson et al. 2006), representing distinct biological processes.

Most of the methods for network property analysis have been implemented: Pajek, Cytoscape and Network Workbench (http://nwb.slis.indiana.edu/) are used to explore both global and local properties. Mfinder (http://www.weizmann.ac.il/mcb/UriAlon/groupNetworkMotifSW.html), FANMOD (http://www.minet.uni-jena.de/~wernicke/motifs/index.html) and MAVisto (http://mavisto.ipk-gatersleben.de/) are designed for finding network motifs. CFinder (http://www.cfinder.org/) can be used to define network communities. More network measures and computational tools are expected to be developed to interpret biological meanings in the future.

Network dynamics and modeling

Dynamics of cancer gene networks can be constructed by mapping expressed cancer gene in different stages of tumor progression such as tumors in the early developmental stages and metastasis onto all kinds of cellular networks such as protein interaction networks. Chuang et al. mapped tumor expressed genes onto a human protein interaction network to identify subnetworks as cancer biomarkers. The resulting cancer subnetwork markers are more reproducible than individual marker genes selected without network information, and that they achieve higher accuracy in the classification of metastatic versus non-metastatic tumors (Chuang et al. 2007).

Although cancer is considered a very heterogeneous disease, querying mutating genes in tumor samples using the network communities defined in a human cancer signaling map reveals that one common network module occurs in most of tumor samples. Furthermore, the common module seems to collaborate with one or more tumor specific modules for tumorigenesis in different tumor types (Cui et al. 2007). In particular, breast and lung cancers have more complex oncogenic signaling block collaborative patterns than other cancer types examined, highlighting their heterogeneous nature.

Tumor gene co-expression networks can be constructed using different types or subtypes of tumors. Weighted gene coexpression networks have been constructed using gene microarray profiles of glioblastoma samples (Horvath et al. 2006). Analysis of such networks provides a blueprint for leveraging genomic data to identify key control networks and molecular targets for cancer.

Tumor gene microarary profiles can also be used to reversely construct gene networks. We constructed a gene regulatory network using the time course microarray profiles from a mouse epithelial breast cell line (BRI-JM01) (Wang et al. 2007), which undergo an epithelial to mesenchymal transition (EMT) when they are treated with TGF-β (Lenferink et al. 2004). Notably, clusterin, one of the genes that are up-regulated in the middle and late time-points shows many regulatory links to other genes in the network. During the EMT process, clusterin is secreted by the BRI-JM01 cells. Interestingly, by applying anti-clusterin antibodies to the TGF-β treated BRI-JM01 cells, we were able to block the TGF-β induced EMT (Wang et al. 2007). Reverse-engineered gene network approach has been applied to the expression profiles of prostate cancer to identify genetic mediators and mediating pathways (Ergun et al. 2007).

Oncogene predictions can be made by constructing networks and modeling of multiple sources of data. Oncogenes of the malignant B-cell phenotypes have been predicted by generating a network from multiple sources: a genome-wide compendium of

human B-cell molecular interactions, in combination with a large set of microarray expression profiles (Mani et al. 2008). The resulting network contains protein–protein, protein–DNA and post-translational interactions. This approach allows capturing several different mechanisms of action associated with oncogenic lesions, specific tumor phenotypes and biochemical perturbations.

In summary, the analysis of the cancer phenomenon using a network approach is still in its infancy. However, network approaches provide new ways to explore the complexity of biological systems and lead to discover new insights into high-throughput data.

Acknowlegements

I would like to thank Dr. Maria Luz Jaramillo for crital reading the manuscript.

References

Awan, A. Bari, H. Yan, F. Moksong, S. Yang, S. Chowdhury, S. Cui, Q. Yu, Z. Purisima, E.O. Wang, E. 2007. Regulatory network motifs and hotspots of cancer genes in a mammalian cellular signaling network. *IET Systems Biology* 1(5):292-297.

Babu, Madan M. Luscombe, Nicholas M. Aravind, L. Gerstein, Mark Teichmann, Sarah A. 2004. Structure and evolution of transcriptional regulatory networks. *Current Opinion in Structural Biology* 14(3):283-291.

Balazsi, G. Barabasi, A. L. Oltvai, N. Z. 2005. Topological units of environmental signal processing in the transcriptional regulatory network of *Escherichia coli*. *Proceedings of the National Academy of Science USA* 102(22):7841-7846.

Barabasi, Albert-László. Albert, Réka**.** 1999. Emergence of scaling in random networks. *Science* 286:509-512.

Barabasi, Albert-László. Oltvai, Zoltán N. 2004. Network biology: understanding the cell's functional organization. *Nature Review Genetics* 5(2):101-113.

Batada, Nizar N. Hurst, Laurence D. Tyers, Mike. 2006. Evolutionary and physiological importance of hub proteins. *PLoS Computational Biology* 2(7):e88.

Bhalla, Upinder S. Ram, Prahlad T. Iyengar, Ravi**.** 2002. MAP kinase phosphatase as a locus of flexibility in a mitogen-activated protein kinase signaling network. *Science* 297(5583):1018-1023.

Blitzer, Robert D. Connor, John H. Brown, George P. Wong, Tony. Shenolikar, Shirish**.** Iyengar, Ravi. Landau, Emmanuel M**.** 1998. Gating of CaMKII by cAMP-regulated protein phosphatase activity during LTP. *Science* 280(5371):1940-1942.

Chuang, Han Yu. Lee, Eunjung. Liu, Yu Tsueng. Lee, Doheon. Ideker, Trey. 2007. Network-based classification of breast cancer metastasis. *Molecular Systems Biology* 3:140.

Cui, Qinghua. Ma, Yun. Jaramillo, Maria. Bari, Hamza. Awan, Arif. Yang, Song. Zhang, Simo. Liu, Lixue. Lu, Meng. O'Connor-McCourt, Maureen. Purisima, Enrico O. Wang, Edwin. 2007. A map of human cancer signaling. *Molecular Systems Biology* 3:152.

Cui, Qinghua. Yu, Zhengbao. Purisima, Enrico O. Wang, Edwin. 2006. Principles of microRNA regulation of a human cellular signaling network. *Molecular Systems Biology* 2:46.

Dublanche, Yann. Michalodimitrakis, Konstantinos. Kummerer, Nico. Foglierini, Mathilde. Serrano, Luis. 2006. Noise in transcription negative feedback loops: simulation and experimental analysis. *Molecular Systems Biology* 2:41.

Ergun, Ayla. Lawrence, Carolyn A. Kohanski, Michael A. Brennan, Timothy A. Collins, James J. 2007. A network biology approach to prostate cancer. *Molecular Systems Biology* 3:82.

Ferrell, James E**.** 2002. Self-perpetuating states in signal transduction: positive feedback, double-negative feedback and bistability. *Current Opinion in Cell Biology* 14:140-148.

Horvath, S. Zhang, B. Carlson, M. Lu, K. V. Zhu, M. Felciano, R. M. Laurance, F. 2006. Analysis of oncogenic signaling networks in glioblastoma identifies ASPM as a molecular target. *Proceedings of the National Academy of Science USA* 103(46):17402-17407.

Jeong, H. Mason, S. P. Barabasi, A. L. Oltvai, Z. N. 2001. Lethality and centrality in protein networks. *Nature* 411(6833):41-42.

Jonsson, Pall F**.** Bates, Paul A**.** 2006. Global topological features of cancer proteins in the human interactome. *Bioinformatics* 22(18):2291-2297.

Lenferink, Anne E. Magoon, Joanne. Cantin, Christiane. O'Connor-McCourt, Maureen. 2004. Investigation of three new mouse mammary tumor cell lines as models for transforming growth factor (TGF)-beta and Neu pathway signaling studies: identification of a novel model for TGF-beta-induced epithelial-to-mesenchymal transition. *Breast Cancer Research* 6(5):R514-R530.

Mahdavi, Mahmoud. Lin, Yen Han. 2007. False positive reduction in protein-protein interaction predictions using gene ontology annotations. *BMC Bioinformatics* 8(1):262.

Mangan, S. Alon, U. 2003a. Structure and function of the feed-forward loop network motif. *Proceedings of the National Academy of Science USA* 100(21):11980-11985.

Mangan, S. Zaslaver, A. Alon, U. 2003b. The Coherent Feedforward Loop Serves as a Sign-sensitive Delay Element in Transcription Networks. *Journal of Molecular Biology* 334(2):197-204.

Mani, Kartik M. Lefebvre, Celine. Wang, Kai. Lim, Wei Keat. Basso, Katia. la-Favera, Riccardo. Califano, Andrea. 2008. A systems biology approach to prediction of oncogenes and molecular perturbation targets in B-cell lymphomas. *Molecular Systems Biology* 4:169.

Mason, O. Verwoerd, M. 2007. Graph theory and networks in Biology. *IET Systems Biology* 1(2):89-119.

Oda, Kanae. Kitano, Hiroaki. 2006. A comprehensive map of the toll-like receptor signaling network. *Molecular Systems Biology* 2:2006.0015.

Oda, Kanae. Matsuoka, Yukiko. Funahashi, Akira. Kitano, Hiroaki. 2005. A comprehensive pathway map of epidermal growth factor receptor signaling. *Molecular Systems Biology* 1:2005.0010.

Patil, K. R. Nielsen, J. 2005. Uncovering transcriptional regulation of metabolism by using metabolic network topology. *Proceedings of the National Academy of Science USA* 102(8):2685-2689.

Platzer, Alexander. Perco, Paul. Lukas, Arno. Mayer, Bernd. 2007. Characterization of protein-interaction networks in tumors. *BMC Bioinformatics* 8:224.

Saeed, Ramazan. Deane, Charlotte M. 2006. Protein protein interactions, evolutionary rate, abundance and age. *BMC Bioinformatics* 7:128.

Shen-Orr, Shai S. Milo, Ron. Mangan, Shmoolik. Alon, Uri. 2002. Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nature Genetics* 31(1):64-68.

Spirin, Victor. Gelfand, Mikhail S. Mironov, Andrey A. Mirny, Leonid A. 2006. A metabolic network in the evolutionary context: multiscale structure and modularity. *Proceedings of the National Academy of Science USA* 103(23):8774-8779.

Wang, E. Lenferink, A. O'Connor-McCourt, M. 2007. Cancer systems biology: exploring cancer-associated genes on cellular networks. *Cellular and Molecular Life Science* 64(14):1752-1762.

Wang, Edwin. Purisima, Enrico. 2005. Network motifs are enriched with transcription factors whose transcripts have short half-lives. *Trends in Genetics* 21(9):492-495.

Wuchty, Stefan. Barabasi, Albert-Laszlo. Ferdig, Michael T. 2006. Stable evolutionary signal in a yeast protein interaction network. *BMC Evolutionary Biology* 6:8.

Wuchty, Stefan. Oltvai, Zoltán N. Barabasi, Albert-Laszlo. 2003. Evolutionary conservation of motif constituents in the yeast protein interaction network. *Nature Genetics* 35(2):176-179.

Yu, Haiyuan. Kim, Philip M. Sprecher, Emmett. Trifonov, Valery. Gerstein, Mark. 2007. The importance of bottlenecks in protein networks: correlation with gene essentiality and expression dynamics. *PLoS Computational Biology* 3(4):e59.

Zhang, Lan V. King, Oliver D. Wong, Sharyl L. Goldberg, Debra S. Tong, Amy H. Y. Lesage, Guillaume. Andrews, Brenda. Bussey, Howard. Boone, Charles. Roth, Frederick P. 2005. Motifs, themes and thematic maps of an integrated *Saccharomyces cerevisia*e interaction network. *Journal of Biology* 4(2):6.

Figure legends:

Figure 1. Network motifs in gene regulatory networks.
A, Single Input Module (SIM): a transcription factor (TF) regulates a group of genes (G1, G2, G3 and G4). B, Feedforward Loop (FFL): a transcription factor (TF1) regulates the second transcription factor (TF2), both TF1 and TF2 regulate a target gene (G1). C, Bi-fan: both transcription factors TF1 and TF2 regulate both target genes (G1 and G2).

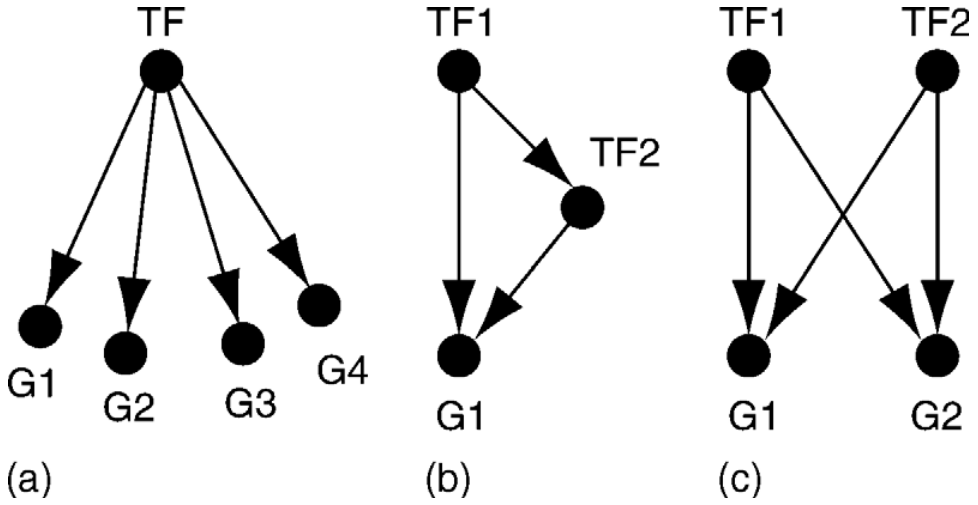

Figure 1