Network modelling of gene regulation

Joshua W. K. Ho · Michael A. Charleston

Received: 1 September 2010 / Accepted: 4 November 2010 / Published online: 23 December 2010
© International Union for Pure and Applied Biophysics (IUPAB) and Springer 2010

Abstract Gene regulatory network (GRN) modelling has gained increasing attention in the past decade. Many computational modelling techniques have been proposed to facilitate the inference and analysis of GRN. However, there is often confusion about the aim of GRN modelling, and how a gene network model can be fully utilised as a tool for systems biology. The aim of the present article is to provide an overview of this rapidly expanding subject. In particular, we review some fundamental concepts of systems biology and discuss the role of network modelling in understanding complex biological systems. Several commonly used network modelling paradigms are surveyed with emphasis on their practical use in systems biology research.

Keywords Gene regulatory network · Systems biology · Bioinformatics

Introduction

Gene expression—the process by which information embedded in DNA is used to synthesise biologically functional molecules—is a complex and tightly regulated process, which lies at the heart of molecular biology. To illustrate the complexity of a gene regulatory system, let us consider a hypothetical system shown in Fig. 1. At the chromatin level, DNA accessibility is controlled by DNA modification (Suzuki and Bird 2008), histone modification (Strahl and Allis 2000), and nucleosome positioning (Jiang and Pugh 2009). At the transcriptional level, gene expression is regulated by a combination of transcription factors, cofactors, and insulators (Sutherland and Bickmore 2009). After transcription, the primary transcripts may be alternatively spliced, giving rise to multiple protein isoforms that may have distinct biological properties (Barash et al. 2010). The spliced mRNA transcripts are exported to the cytoplasm where some of them are translated into proteins whereas others are degraded (through microRNA-mediated silencing for example Bartel 2009). Furthermore, some proteins may be degraded by different mechanisms such as “nonsense-mediated decay” (Mendell et al. 2004). Eventually, some proteins are folded and post-translationally modified to become fully functional. Some of these functional protein molecules are involved in the signalling pathways that regulate the transcription of other genes and even their own genes. To further complicate matters, each cell constantly receives signals from the external environment through a range of signalling mechanisms (e.g. signals from the extracellular matrix or neighbouring cells). And this is only the story of

J. W. K. Ho (✉) · M. A. Charleston
School of Information Technologies,
The University of Sydney, Sydney, NSW 2006, Australia
e-mail: joshua@it.usyd.edu.au

M. A. Charleston
Centre for Mathematical Biology,
The University of Sydney, Sydney, NSW 2006, Australia

Present Address:

J. W. K. Ho
Department of Medicine, Brigham and Women's Hospital
and Harvard Medical School, Boston, MA 02115, USA
e-mail: jwho@partners.org

Fig. 1 This cartoon illustrates some of the key processes that regulate eukaryotic gene expression. The message is clear: gene expression is tightly regulated by a complex system of interconnected components inside and outside a cell

one of the tens of thousands of genes in a typical eukaryotic cell.

So how can one unravel and understand the complexity of this enormous regulatory system? To solve this problem, traditional molecular biology has adopted a bottom-up solution, which was neatly summarised by pioneer molecular biologist William Astbury: “[molecular biology] is concerned particularly with the forms of biological molecules and with the evolution, exploitation and ramification of those forms in the ascent to higher and higher levels of organisations.” (Astbury 1961). In other words, molecular biology is a divide-and-conquer approach to understand complex biological processes. This approach has been very successful in delineating many aspects of the cellular regulatory system for the past 50 years. Nonetheless, such an approach has limitations: it is inherently ineffective when interactions between molecules play a much more important role than the molecules themselves (Ahn et al. 2006). Such a limitation is clearly exemplified by various complex and chronic diseases, such as cancer, heart failure, diabetes mellitus, and asthma, whose pathogenesis is likely multifactorial and driven by complex interactions among many molecular constituents and environmental factors (Ahn et al. 2006). Besides this practical limitation, the reductionist approach also suffers from a deeper philosophical problem by implying a one-way causation from gene to phenotype, which implicitly down-plays the important role of the multi-scale interactions of molecular, cellular, physiological, and environmental components as part of the multifactorial causal agent (Noble 2008). Systems biology attempts to overcome these limitations by taking a top-

down approach to understanding the molecular system as a whole. Instead of viewing the complex network from the lens of individual molecules, systems biologists seek to study collective properties of molecules, such as system stability, robustness, and evolvability (Huang 2004; Kitano 2007b). In our view, the reductionist and the systems approach to understand complex biomolecular systems are largely complementary. Molecular biology is more powerful at revealing details of local molecular interactions whereas systems biology is better at understanding emergent properties that are not apparent when analyzing only a small number of molecules. Therefore, they should be viewed as two complementary approaches to deal with biological complexity: in a sense, molecular biology dissects complexity whereas systems biology embraces complexity.

While molecular modelling is central to molecular biology, network modelling is central to systems biology (Lander 2010). Network models are indeed very powerful for analysing complex biological systems, especially the gene expression system. A network model of gene regulation is commonly referred to as a gene regulatory network (GRN). It is important to note that a GRN is an abstract construct rather than a physical entity. Different authors have used the term “gene regulatory network” to refer to different aspects of gene regulation, such as transcription factor binding, cellular signalling, genetic epistatic interaction, and gene co-expression. All these network models are essentially only a small projection of the true biological regulatory system (Brazhnik et al. 2002), and each of them reveals different properties of the underlying system. A typical GRN models the states (activity or other

properties) of one of more kinds of these biomolecules (e.g. genes, transcription factors, and signalling molecules) as nodes. These nodes may be connected by direct and/or indirect edges. Edges may represent physical interactions, causal relations, or functional associations. For example, one class of GRN is called a transcriptional regulatory network (TRN) because it mainly models transcriptional factor-mediated regulations. In a typical TRN, two types of nodes are present: genes and transcription factors. Each edge in a TRN represents binding of transcription factor to the promoter of a target gene. In general, the choice of the type of nodes or edges typically depends on the availability of data and the application of the model.

Once the scope of the regulatory model (e.g. gene coexpression network, transcription factor-binding network, signalling pathway network, etc.) is defined and relevant data are collected, it is possible to computationally construct and represent the GRN using a network modelling paradigm such as graph models, Boolean networks, Petri nets, differential equations, or probabilistic Bayesian networks. Each modelling paradigm specifies rules for construction, representation, and manipulation of a model. These modelling paradigms differ in computational complexity, expressiveness, and the type of downstream analysis they can perform. Some models are good for network structure analysis whereas some are better for network dynamic analysis. The construction of a network model is sometimes called *network inference* or *network reconstruction*. Network models can be inferred from experimental data or constructed from pre-collected data in various molecular interaction databases. Once a GRN is computationally encoded using one of these mod-

elling approaches, network properties can be easily simulated or analysed using tools from graph theory, control theory, probability theory, and complex network science. We use the term *network modelling* to refer to the collective activity of network model inference, simulation, and analysis.

Many of the early network reconstruction algorithms focused on building network models from microarray gene expression data (Liang et al. 1998; Chen et al. 1999; D'haeseleer et al. 1999; Friedman et al. 2000; de la Fuente et al. 2002; Gardner et al. 2003). As experimental technology has rapidly evolved in the last decade, we can now perform a larger range of high-throughput experiments, including high-throughput genotyping, quantitative protein expression profiling by mass spectrometry, protein–protein interaction analysis by high-throughput yeast-two-hybrid systems, and protein–DNA interaction mapping by microarray and sequencing technologies. Perhaps the most breathtaking development of the last few years is the advancement of massively parallel sequencing technology (the so-called next-generation sequencing or NGS) (Metzker 2010). NGS technologies provide a means to rapidly sequence massive amounts of short DNA fragments using a relatively small amount of starting material. In combination with various experimental procedures, NGS has enabled many novel types of genome-wide profiling approaches including RNA-seq for transcriptomic analysis (Wang et al. 2009), ChIP-seq for profiling of transcription factor binding and chromatin modification analysis (Park 2009), and DNA methylation analysis (Laird 2010). In addition, there is a surge of high-quality pathway databases and sequence motif databases, such as KEGG (Kanehisa

Fig. 2 Network modelling provides a natural means to combine diverse experimental data and prior biological knowledge for data analysis in a principled manner

and Goto 2000), PID (Schaefer et al. 2009), and JASPAR (Sandelin et al. 2004). To capitalise on this mountain of data, the current trend is toward large-scale integrative analysis on multiple types of data and prior biological knowledge (Hawkins et al. 2010). Network models provide a natural means for such integration since many cellular regulatory and interaction events can be easily represented by a network of interconnected nodes (Fig. 2).

Many general reviews on GRN modelling techniques are already available (de Jong 2002; Brazhnik et al. 2002; Szallasi et al. 2006; Schwartz 2008; Karlebach and Shamir 2008). However, it is often difficult to navigate through the sea of modelling techniques. Instead of being another technical exposition, this review aims to clarify the motivation and principles of network modelling, and discuss the practical application of some widely used modelling paradigms.

Structural modelling and analysis

The use of graph models to represent the structure of GRNs is particularly attractive as they are intuitive, easy to visualise, and have access to a range of analytical tools from graph theory and complex network analysis (Strogatz 2001; Newman et al. 2006; Barabási 2007). Gene expression systems can be modelled such that nodes represent genes and directed signed (+ or −) edges represent activation or repression

between two genes by some cellular mechanisms (see Fig. 3 for an example). A GRN can be constructed by integrating transcriptional factor-binding data from databases, chromatin immunoprecipitation (ChIP) experiments and microarray profile data. Prominent early examples include the construction and analysis of GRNs of *Escherichia coli* (Shen-Orr et al. 2002; Ma et al. 2004; Salgado et al. 2006), yeast (Guelzim et al. 2002; Lee et al. 2002; Luscombe et al. 2004; Tong et al. 2004), and humans (Rodriguez-Caso et al. 2005; Franke et al. 2006).

The structure of a network is often characterised by a set of network statistics which summarise the global network properties. These statistics are often referred to as *topological features*. Commonly used topological features include node degree distribution, longest path length and average clustering coefficient. These topological features are mathematically motivated, but have been widely adopted to the analysis of real-life networks such as the Internet, social networks and biological networks. Many of these topological features are also biologically relevant. For example, the out-degree of a node in a GRN measures the number of genes a regulator directly controls. Similarly, network path length and clustering coefficient measure the signalling distance and the local communication density respectively. These two measures have been used as a measure of the local communication efficiency in a transcriptional regulatory network (Luscombe et al. 2004; Balaji et al. 2006). The study of network structures

Fig. 3 A transcription factor-binding network of *Escherichia coli* constructed using the data from RegulonDB (Salgado et al. 2006) (a). This network consists of 1,306 genes and 2,981 interactions. Network structure can be quantitatively assessed by topological

features and network motifs. Some common network motifs are shown (b). In this example, green nodes are regulators and yellow nodes are non-regulators

can also reveal the underlying robustness of a system, which has significance in the study of genome evolution (Wagner 2000, 2003; Ho and Charleston 2007) and drug discovery (Kitano 2007a).

At the local level, biological networks can be studied in terms of *network motifs*. A network motif is defined as a subgraph that is over-represented in a network compared with a randomised network of the same degree distribution (Milo et al. 2002). Commonly occurring biological network motifs include the single-input module (SIM), feed-forward loop (FFL), bi-fan, and multiple-input module (Alon 2007) (see Fig. 3b). The prevalence of these network motifs suggests that nature might have selected for these building blocks during the evolution of complex biological networks due to the selective advantage of these regulatory patterns (Conant and Wagner 2003), but the actual evolutionary processes involved are still being actively investigated.

The structure of a GRN is highly dynamic, in the sense that edges may be present or absent in response to different environmental or cellular conditions. This “rewiring” of a network is captured by a set of *condition-specific networks*. For example, by comparing the structure of the transcriptional subnetwork of yeast under exogenous and endogenous conditions, Luscombe et al. found that subnetworks have evolved to produce rapid, large-scale responses in exogenous states, while having coordinated processes in endogenous conditions (Luscombe et al. 2004). Such changes in network topology enable the cell to mount effective responses to cope with diverse and sudden changes of the external environment. At the molecular level, this change corresponds to activation or inhibition of certain transcription factors by metabolites or proteins induced by different environments.

Direct network comparison is another active research endeavour, whose goal is to identify biologically interesting patterns by comparing the structure of two or more biological networks (Sharan and Ideker 2006). A new class of problem called network alignment seeks a node-to-node mapping that best matches two graphs based on certain optimization criteria. A score is then assigned to each alignment to assess the goodness of the match. Therefore, network alignment is a natural approach to measure distance (i.e. difference) between two networks. Phylogenetic trees can even be generated by aligning metabolic pathways (Ogata et al. 2000; Heymans and Singh 2003) and gene regulatory networks (Trusina et al. 2005). The conservation and divergence of pathways have also been discovered through protein interaction network alignment (Kelley et al. 2003; Liang et al. 2006). Research in this area is still in its infancy, but its development is anticipated to be as

important as that of the sequence alignment methods (Sharan and Ideker 2006).

Using the aforementioned techniques, simple structural analysis of a GRN can often yield surprising and valuable insights about its underlying gene regulatory system. Figure 3a shows a GRN of *E. coli* that was reconstructed by mining transcription factor-binding data from RegulonDB (Salgado et al. 2006). This network consists of 1,306 genes and 2,981 interactions. It has a relatively short longest path length (6; a gene can at most affect six levels of downstream genes), and a high proportion of FFLs (on average 1.05 FFLs per gene). This shows that local gene-to-gene communication is probably very efficient. There are 18 disconnected network components, in which the largest contains 92% of the genes. Upon visual inspection, a smaller number of interconnected hub genes are found in the core of each network, while a large number of SIM are found at the periphery of the network. This network structure is significantly different from randomly rewired networks that have the same degree distribution, indicating that these network properties may have some functional significance in genome evolution (Ho and Charleston 2007). This example illustrates the power of simple network structure analysis.

Besides the empirically derived GRNs described above, another similar but fundamentally different type of graph model has been used in the literature. It is called the *gene coexpression network* (Carter et al. 2004; Choi et al. 2005; Zhang and Horvath 2005). In a coexpression network, an undirected edge is placed between two genes if they have highly similar expression patterns across multiple measurements. A coexpression network is usually inferred by pairwise comparison of gene expression profile in a microarray dataset. The general idea is that genes with similar expression patterns are likely to be coregulated, based on the guilt-by-association principle (Quackenbush 2003; Wolfe et al. 2005). Although similar kinds of analytical tools can be used to study both GRNs and coexpression networks, their results should be interpreted differently since they generally have quite different network topologies (Balaji et al. 2006) and the meaning of the edges is very different. Each edge in a coexpression network implies that the two connected nodes have highly similar expression pattern, but do not directly convey information flow or causality. Figure 4 illustrates how condition-specific coexpression networks can be constructed from a microarray dataset of eight genes. In this toy example, the pattern of expression changes is not apparent when we visually inspect the heat map alone. We can construct a coexpression network

Fig. 4 Construction of two condition-specific gene coexpression networks from an eight-gene microarray dataset. At first, no obvious pattern emerges from visual inspection of the heat map of this simulated dataset. Nonetheless, comparison of the coexpress-

sion networks constructed from the two conditions (conditions 1 and 2) reveals a large shift in coexpression pattern among these eight genes. This simple example illustrates the power of gene coexpression network analysis

for each biological condition by placing an edge between every pair of nodes (genes) that have a high Pearson's correlation coefficient between their expression profiles in that condition. By comparing the structure of the inferred coexpression networks, alteration in global coexpression pattern may become apparent (see Fig. 4). Differential coexpression analysis is useful for complementing other microarray analysis tasks such as differential expression analysis and differential variability analysis (Ho et al. 2008b), and it is particularly useful for identifying large-scale gene dysregulation in human diseases (de la Fuente 2010). Besides the Pearson's correlation coefficient, other similarity measures have been proposed in the literature, including partial correlation (de la Fuente et al. 2004) and mutual information (Basso et al. 2005).

Simulation modelling and analysis

Simulation is the most straightforward way to understand the dynamics of a model. GRN simulation plays an important role in addressing various types of biological problem, including (1) prediction of the future states of a set of target genes given the initial condition of the system, (2) evaluation of system stability under various perturbation conditions, and (3) generation of synthetic gene expression data given the network structure of a GRN. Intuitively, simulation can be carried out in any GRN model that contains a precise description of how

the expression of one gene (or a set of genes) is quantitatively affected by its regulator(s). There are many different ways (i.e. models) in which gene expression dynamics can be simulated, such as rule-based models, regression-based models, logical models, and stochastic models.

Coupled ordinary differential equations (ODEs) have been widely used to model real-life systems in science and engineering. The application of ODEs to model GRNs is also widespread because ODEs can be easily simulated via a wide range of numerical techniques such as Euler and Runge–Kutta methods.

The central idea is to model the *rate of change* of each reaction as a set of differential equations. These differential equations are *coupled*, which means the rate of change of the expression of a gene is affected by the current expression level of other genes. When formulating an ODE model, it is important to choose the right balance between model complexity, identifiability, and interpretability. Generally, the parameters in the ODE models can either be determined experimentally or from prior knowledge. A complex model may require specification of many parameters, which may not be reliably estimated if only a limited amount of experimental data is available. Therefore it is usually impractical to model all the physical biochemical processes in a GRN (such as transcription, translation, translocation, and so on). Many have proposed to model functional relationships among genes instead. One simple approach is to model the influence

of the regulators of a target gene's expression as a linear function (e.g. Chen et al. 1999). If it is more desirable to model non-linear regulatory relationships, one can choose to use other functional formulations, such as the *Hill function* (Hill 1910; Hofmeyr and Cornish-Bowden 1997; Mendes et al. 2003) and the piecewise linear model (Glass and Kauffman 1973).

Boolean networks and Petri nets are two other interesting simulation modelling approaches. They have simple rules for simulation and in many cases are easy to construct. Moreover, the simple structure of these models enables us to analyse network dynamics without explicitly carrying out a simulation.

Boolean networks were introduced for gene network analysis by Stuart Kauffman over 40 years ago (Kauffman 1969). This model considers a set of genes as binary switches, and the state of each gene is governed by a set of well formed *transition functions* (Fig. 5). The *state* of a Boolean network is the ordered set of Boolean variables in the system. For example, a system consisting of two variables can have four different states $\{(00), (01), (10), (11)\}$. Since there is a finite number of states (2^n states for a network of n genes), and transitions among states are completely determined by the transition functions, it is possible to construct the *transition state space* of a Boolean network. Since a Boolean network is a deterministic model, each state can transition to at most one other state. By analysing the structure of the state space, it is possible to easily identify states of sets of states that does not transition to other states, and therefore form an *attractor* state of the system. These stable attractor states correspond to the steady states of a biological system. Changes of

steady states are typical of many biological processes, including stem cell differentiation and cancer metastasis (Huang 2010; Huang et al. 2009). Boolean networks therefore provide a simple tool to analyse steady state dynamics.

An interesting application of a Boolean network is the study of feedback cycles in GRNs (Kwon and Cho 2008; Kim et al. 2008). By studying the structure of the state space and particular trajectories, the dynamic behaviour of the system can be analysed without any actual simulation of the system. For example, Aldana et al. studied the Boolean transition state space of the dynamics in a scale-free network, and the effect of gene duplication on the evolvability and robustness of a gene network (Aldana 2003; Aldana et al. 2007).

Another advantage of Boolean network models is that they are often very easy to simulate, which facilitates efficient and large-scale network simulation analysis (see Helikar et al. 2008 for an example of signal transduction network simulation using a Boolean network). Inference of Boolean network transition functions from time-series microarray data is made possible by a number of reverse engineering algorithms (Liang et al. 1998; Akutsu et al. 2000; Martin et al. 2007). The general principle is to extract the rules of state transitions from the expression profiles of multiple time points or gene perturbation experiments. These inference algorithms were shown to reliably produce the gene expression dynamics observed in the expression profiles. The ability to build a model that captures the dynamic behaviour of a gene expression system has made Boolean networks a very attractive tool for microarray analysis.

Fig. 5 A four-gene Boolean network is shown in (a) and its corresponding state transition function is shown in (b). By emulating all position state transitions trajectories, we can construct the entire state-space of a given Boolean network model. The state space of our example network is shown in (c). In this state-space diagram, attractor states are represented by green nodes while transient states are represented by blue nodes. Analysis of the

state space network structure can reveal interesting dynamical properties of the Boolean network. (d) An example Petri net. The orange circles represent places, and the blue rectangles represent transitions. The black circles inside a place are the tokens, and the distribution of tokens among all the places represents a marking of a Petri net

A Boolean network assumes that state transition is a deterministic process, but sometimes it is more suitable to model state transition as a non-deterministic process. In this case, a Petri net is more suitable (Fig. 5d). A Petri net (Petri 1962) is a well-established model for describing a system of concurrent processes. The basic concept of Petri net modelling is reviewed by Reisig et al. (Reisig 1985; Reisig and Rozenberg 1998), and has recently been applied to model cellular regulatory networks (Goss and Peccoud 1998; Pinney et al. 2003; Steggles et al. 2007).

In modelling gene expression systems, a gene is represented by a place and its gene expression level is represented by the number of tokens it holds. The state of a Petri net is called a *marking*, and is defined as the distribution of tokens among the places. Tokens are moved by *transitions*. Whenever a transition fires, non-deterministically, tokens are consumed from one place and put in others. Similar to the Boolean network model, one of the most important applications of Petri nets is to analyse the properties of its state space (in this case, the marking space). Since this is a non-deterministic model, a marking can move to a number of possible markings.

Beyond the basic definition of Petri nets, the model can be extended to cope with some more sophisticated cases. In particular, places can be *weighted* such that transitions are enabled only when the number of tokens in the input places exceeds its weight. Arcs can be weighted to specify the number of tokens to be consumed/transferred in one firing event. Another extension is called a hybrid Petri net, which allows the number of tokens in a place to be either an integer number or a real number. A hybrid Petri net is useful in modelling and simulating gene regulatory systems (Matsuno et al. 2000; Nagasaki et al. 2006). While the hybrid Petri net representation is very similar to the

standard representation of biochemical reaction pathways, it can also be used for quantitative simulation of system dynamics. Another interesting recent example of Petri net in GRN modelling is provided by Küffner et al. (2010) where they combined Petri network with Fuzzy logic to infer GRNs.

Probabilistic modelling and analysis

The Bayesian network is an entirely different modelling scheme. In a typical Bayesian network model, the expression of a gene can be modelled as a discrete or a continuous random variable, X_i . The distribution of expression patterns of all the genes in a system can be captured by their joint probability density (JPD), $P(X_1, X_2, \dots, X_n)$, where n is the total number of genes in the system. If each variable has two possible states (i.e. it is a binary random variable), a complete JPD of n variables needs to specify the probability of 2^n different combination of states, which is prohibitively large for large n . One key idea in Bayesian network modelling is to decompose the JPD into a product of many lower dimensional conditional probability densities (CPDs) by exploiting conditional independence among variables.

The conditional dependency structure of the random variables can be summarised by a directed acyclic graph (for example, Fig. 6a). It is important to note that the directed edges in a Bayesian network merely convey information about conditional dependency among variables, which has a very different meaning from an empirically determined GRN or a gene coexpression network. Therefore direct comparison of a Bayesian network structure to those that do not have a probabilistic meaning can be misleading. Given a Bayesian network, one is able to perform various *probabilistic queries* about the marginal distribution (hence the

Fig. 6 An example Bayesian network with three nodes. The directed acyclic graph shown in (a) depicts the conditional dependency structure among the three variables. Dynamic information can be encoded in a dynamic Bayesian network (b), where the dependency structure between successive time points (t and $t+1$) is explicitly modelled. An intervention to a variable in a Bayesian network, say node B in (c), has the effect of removing all its incoming edges in the network (say $A \rightarrow B$). A Bayesian network

which is consistent with all interventional events is called a causal Bayesian network. A Bayesian network is only one member of a larger family of probabilistic graphical models. Other popular probabilistic graphical models include Markov random fields (d) and factor graphs (e). All these probabilistic graphical models are well suited for integrative data analysis of multiple data types since they are flexible and are inherently capable of dealing with noise in the experimental data

expected value) of any set of variables. The use of the Bayesian network formalism was first proposed by Pearl (1988), and applied to analyse gene expression data by Friedman et al. in 2000 (Friedman et al. 2000). Since then, it has attracted a lot of attention mainly due to its natural ability to deal with experimental noise in a principled manner (Friedman 2004).

The standard Bayesian network models the steady-state probability distribution of genes expression levels. To incorporate dynamical information from time-series data and gene perturbation data, a *dynamic Bayesian network* model can be used (Smith et al. 2003; Husmeier 2003; Zou and Conzen 2005; Dojer et al. 2006). In a dynamic Bayesian network, each random variable denotes the gene expression levels at successive time points (time t and $t + 1$, see Fig. 6b). Learning the structure of a large dynamic Bayesian network is usually computationally expensive and often requires a large time-series dataset for accurate estimation.

It must be emphasised that the directionality of an edge in a Bayesian network does not necessarily imply causality. The edge, $B \rightarrow C$, implies that B and C are not conditionally independent, and their dependency is encoded in a CPD, $P(C|B)$, which does not have the same meaning as “ C is caused by B ”. To resolve this problem, Pearl (2000) introduced a crucial concept: the probability of a causal event must be described by interventional events, not by conditional events alone. He introduced a new notation, $do(X = x)$ to denote an intervention event that set the value of X to x . In general, $P(C|B = b) \neq P(C|do(B = b))$ since $do(B = b)$ implies that $B = b$ with probability 1, whereas the probability of the event $B = b$ is dependent on its parents through the CPD $P(B|parents(B))$ in a Bayesian network. In other words, an intervention, $do(B = b)$, has the effect of removing all incoming edges of node B in the corresponding Bayesian network graph (Fig. 6c), whereas, an observation, $B = b$, does not alter the graph structure at all. A Bayesian network is said to be a *causal Bayesian network* if the network structure (and all the subgraphs induced by all possible intervention events) is consistent with the observed interventions. Therefore, it is necessary to have interventional data (in the case of gene expression system, data from gene knock-out or knock-down experiments) are required to learn a causal Bayesian network model. An elegant example of causal Bayesian network learning can be found in Sachs et al. (2005), Pe’er (2005).

Bayesian networks actually belong to a large family of modelling techniques called probabilistic graphical modelling (Koller and Friedman 2009). The Bayesian network is a directed graphical model and the directionality of the edges are encoded by the set of CPDs. There

are other undirected graphical models, and two of the most common undirected graphical models are Markov random fields (Fig. 6d) and factor graphs (Fig. 6e). These models allow more complicated relationships between nodes (such as feedback loops) to be modelled since they allow the JPD to be decomposed into any arbitrary positive functions (called potential functions, or factors) instead of restricting those functions to be probability density functions. With well-designed potential functions, one can perform inference and learning of these undirected networks efficiently using a range of exact or approximate algorithms.

Probabilistic graphical models are gaining popularity in systems biology due to their flexibility and a large range of well-studied analysis tools (Friedman 2004). Probabilistic modelling is particularly attractive for integrating prior biological knowledge (such as signalling pathway structure) and diverse experimental data (gene expression profiles, protein–protein interactions) in a principled manner, and this can be exemplified by several recent applications in the area: (Segal et al. 2003; Ho et al. 2008a; Gat-Viks et al. 2006; Vaske et al. 2010).

Discussion

The success of modelling any biological system relies on having a well-posed biological question and an appropriate choice of modelling strategy (Endy and Brent 2001; Noble 2002). As genetic and biochemical knowledge about how individual genes and biomolecules interact becomes increasingly available, the challenge is to construct models at the right level of abstraction with the right modelling tools for a specific task (Bornholdt 2005). A GRN model can be used for several purposes: (1) hypothesis generation, (2) critical evaluation of multiple alternative biological hypotheses when direct experimentation is impossible or expensive, and (3) summarisation and communication of knowledge about a system. A model is a tool for information extraction, and it is only useful if it is carefully utilised throughout the planning and execution of a research study. A GRN model by itself is of little value if it does not affect the data interpretation or the design of downstream experiments.

Prior to modelling, one should examine the type of analysis that will be of the greatest value. Some modelling paradigms are better suited for model inference (like Bayesian networks), some are better for model analysis (like graphs or Boolean networks), and some are more suitable for simulation (like ODEs). It is important to keep in mind that a model is a representation

of the underlying system according to some characteristics: no model tells you everything about the system, simply because it is not the actual system, and a lot of rational decisions have to be made during the modelling process. This principle is perhaps best summarised by the eminent statistician George E. Box: “Essentially, all models are wrong, but some are useful.” (Box and Draper 1986).

Objective model evaluation is a major challenge in GRN modelling since it is often difficult to test whether a model is good enough. There is often no ground truth about the underlying regulatory network structure. This evaluation problem is particularly problematic when comparing a new network inference algorithm with other existing approaches. One evaluation criterion is based on network structure correctness, such as counting the number of true or false positive edges according to our existing biological knowledge, or checking whether the degree distribution is similar to other experimentally validated biological networks with similar characteristics (such as a power-law node degree distribution). It is also possible to assess a network inference algorithm using simulated expression data of some artificially created networks (Husmeier 2003; Mendes et al. 2003). Although this method is objective, it is hard to test the assumption that the artificial network and the simulated data are biologically realistic. Recent collaborative efforts in gene network inference evaluation, such as the Dialogue on Reverse Engineering Assessment and Methods challenge (Stolovitzky et al. 2007), are gaining momentum and are likely to catalyse the development of better gene network inference methodology.

Besides evaluation of models and network inference algorithms, a more subtle issue arises in verification and validation of the implementation of various network inference, simulation, and analysis algorithms. A theoretically sound algorithm may be implemented incorrectly, yet it is intriguing to find that many scientists tend not to see software as a separate entity from the underlying algorithms or models (Kelly and Sanders 2008). However, it is increasingly recognised that there are novel challenges in testing bioinformatics software due to the large size of the input data, stochastic elements, and the use of sophisticated and computationally intensive algorithms. These problems are particularly relevant in the area of biological network simulation since in general there is no simple means to check whether the output generated by a deterministic or stochastic simulator is consistent with the programme specification. There are some initial efforts to systematically test newly developed biomolecular network simulators (Bergmann and Sauro 2008;

Evans et al. 2008). The current approach involves executing each simulator with multiple existing simulators on some well-studied input models and evaluating the consistency of the simulation results (Bergmann and Sauro 2008). However, it is often difficult to judge what is the correct output when different simulators generate contradicting results, as has been noted that “As no authoritative result set exists, it is hard to devise a metric based on the simulation results, that would tell us whether a given simulation result is ‘correct’ or not” (Bergmann and Sauro 2008). Some recent works have proposed to test this type of network simulation software by verifying that the outputs of multiple test cases adhere to some known necessary properties of the underlying models or algorithms (Chen et al. 2009; Xie et al. 2009). Clearly more work in this area is needed to provide reliable software for network modelling.

With the rapid advancement in high-throughput profiling technologies, data generation is no longer a bottleneck in systems biology. On the other hand, data analysis becomes much more involved. Network modelling not only provides a principled means to combine heterogeneous data and prior biological knowledge, but more importantly, it forces us to better appreciate the true beauty of biological complexity. On this note, we hope the concepts and techniques reviewed in this article can further stimulate development of network modelling techniques and wider application of network modelling in systems biology research.

Acknowledgement We thank Professor Cristobal dos Remedios for critically revising an earlier draft of this manuscript and for his valuable comments on this work.

References

- Ahn AC, Tewari M, Poon CS, Phillips RS (2006) The clinical applications of a systems approach. *PLoS Med* 3:e209
- Akutsu T, Miyano S, Kuhara S (2000) Inferring qualitative relations in genetic networks and metabolic pathways. *Bioinformatics* 16:727–734
- Aldana M (2003) Boolean dynamics of networks with scale-free topology. *Phys Nonlinear Phenom* 185:45–66
- Aldana M, Balleza E, Kauffman S, Resendiz O (2007) Robustness and evolvability in genetic regulatory networks. *J Theor Biol* 245:433–448
- Alon U (2007) Network motifs: theory and experimental approaches. *Nat Rev Genet* 8:450–461
- Astbury WT (1961) Molecular biology or ultrastructural biology. *Nature* 190:1124
- Balaji S, Iyer LM, Aravind L, Babu MM (2006) Uncovering a hidden distributed architecture behind scale-free transcriptional regulatory networks. *J Mol Biol* 360:204–212
- Barabási AL (2007) Network medicine—from obesity to the “diseaseome”. *N Engl J Med* 357:404–407

- Barash Y, Calarco JA, Gao W, Pan Q, Wang X, Shai O, Blencowe BJ, Frey BJ (2010) Deciphering the splicing code. *Nature* 465:53–59
- Bartel DP (2009) MicroRNAs: target recognition and regulatory functions. *Cell* 136:215–233
- Basso K, Margolin AA, Stolovitzky G, Klein U, Dalla-Favera R, Califano A (2005) Reverse engineering of regulatory networks in human B cells. *Nat Genet* 37:382–390
- Bergmann FT, Sauro HM (2008) Comparing simulation results of SBML capable simulators. *Bioinformatics* 24:1963–1965
- Bornholdt S (2005) Less is more in modeling large genetic networks. *Science* 310:449–450
- Box GEP, Draper NR (1986) Empirical model-building and response surface. John Wiley and Sons, Inc
- Brazhnik P, de la Fuente A, Mendes P (2002) Gene networks: how to put the function in genomics. *Trends Biotechnol* 20:467–472
- 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
- Chen T, He HL, Church GM (1999) Modeling gene expression with differential equations. *Pac Symp Biocomput* 4:29–44
- Chen TY, Ho JWK, Liu H, Xie X (2009) An innovative approach for testing bioinformatics programs using metamorphic testing. *BMC Bioinformatics* 10:24
- Choi JK, Yu U, Yoo OJ, Kim S (2005) Differential coexpression analysis using microarray data and its application to human cancer. *Bioinformatics* 21:4348–4355
- Conant GC, Wagner A (2003) Convergent evolution of gene circuits. *Nat Genet* 34:264–266
- de Jong H (2002) Modeling and simulation of genetic regulatory systems: a literature review. *J Comput Biol* 9:67–103
- de la Fuente A (2010) From ‘differential expression’ to ‘differential networking’—identification of dysfunctional regulatory networks in diseases. *Trends Genet* 26:326–333
- de la Fuente A, Brazhnik P, Mendes P (2002) Linking the genes: Inferring quantitative gene networks from microarray data. *Trends Genet* 18:395–398
- de la Fuente A, Bing N, Hoeschele I, Mendes P (2004) Discovery of meaningful associations in genomic data using partial correlation coefficients. *Bioinformatics* 20:3565–3574
- D’haeseleer P, Wen X, Fuhrman S, Somogyi R (1999) Linear modeling of mRNA expression levels during CNS development and injury. *Pac Sym Biocomput* 4:41–52
- Dojer N, Gambin A, Mizera A, Wilczynski B, Tiuryn J (2006) Applying dynamic Bayesian networks to perturbed gene expression data. *BMC Bioinformatics* 7:249
- Endy D, Brent R (2001) Modelling cellular behaviour. *Nature* 409:391–395
- Evans TW, Gillespie CS, Wilkinson DJ (2008) The SBML discrete stochastic models test suite. *Bioinformatics* 24:285–286
- Franke L, van Bakel H, Fokkens L, de Jong ED, Egmont-Petersen M, Wijmenga C (2006) Reconstruction of a functional human gene network, with an application for prioritizing positional candidate genes. *Am J Hum Genet* 78:1011–1025
- Friedman N (2004) Inferring cellular networks using probabilistic graphical models. *Science* 303:799–805
- Friedman N, Linial M, Nachman I, Pe’er D (2000) Using bayesian networks to analyze expression data. *J Comput Biol* 7:601–620
- Gardner TS, di Bernardo D, Lorenz D, Collins JJ (2003) Inferring genetic networks and identifying compound mode of action via expression profiling. *Science* 301:102–105
- Gat-Viks I, Tanay A, Raijman D, Shamir R (2006) A probabilistic methodology for integrating knowledge and experiments on biological networks. *J Comput Biol* 13:165–181
- Glass L, Kauffman SA (1973) The logical analysis of continuous, non-linear biochemical control networks. *J Theor Biol* 39:103–129
- Goss PJ, Peccoud J (1998) Quantitative modeling of stochastic systems in molecular biology by using stochastic Petri nets. *Proc Natl Acad Sci USA* 95:6750–6755
- Guelzim N, Bottani S, Bourguin P, Képès F (2002) Topological and causal structure of the yeast transcriptional regulatory network. *Nat Genet* 31:60–63
- Hawkins RD, Hon GC, Ren B (2010) Next-generation genomics: an integrative approach. *Nat Rev Genet* 11:476–486
- Helikar T, Konvalina J, Heidel J, Rogers JA (2008) Emergent decision-making in biological signal transduction networks. *Proc Natl Acad Sci USA* 105:1913–1918
- Heymans M, Singh AK (2003) Deriving phylogenetic trees from the similarity analysis of metabolic pathways. *Bioinformatics* 19:i138–i146
- Hill AV (1910) The possible effects of the aggregation of the molecules of haemoglobin on its dissociation curves. *J Physiol* 40:iv–vii
- Ho JWK, Charleston MA (2007) Modeling the evolution of gene regulatory networks. In: *Proceedings of the 8th international conference on systems biology (ICSB’07)*, p 44
- Ho JWK, Koundinya R, Caetano T, dos Remedios CG, Charleston MA (2008a) Inferring differential leukocyte activity from antibody microarrays using a latent variable model. *Genome Inform* 21:126–137
- Ho JWK, Stefani M, dos Remedios CG, Charleston MA (2008b) Differential variability analysis of gene expression and its application to human diseases. *Bioinformatics* 24:i390–i398
- Hofmeyr JHS, Cornish-Bowden A (1997) The reversible Hill equation: how to incorporate cooperative enzymes into metabolic models. *Comput Appl Biosci* 13:377–385
- Huang S (2004) Back to the biology in systems biology: what can we learn from biomolecular networks? *Brief Funct Genomic Proteomic* 2:279–297
- Huang S (2010) Cell lineage determination in state space: a systems view brings flexibility to dogmatic canonical rules. *PLoS Biol* 8:e1000380
- Huang S, Ernberg I, Kauffman S (2009) Cancer attractors: a systems view of tumors from a gene network dynamics and developmental perspective. *Semin Cell Dev Biol* 20:869–876
- Husmeier D (2003) Sensitivity and specificity of inferring genetic regulatory interactions from microarray experiments with dynamic bayesian networks. *Bioinformatics* 19:2271–2282
- Jiang C, Pugh BF (2009) Nucleosome positioning and gene regulation: advances through genomics. *Nat Rev Genet* 10:161–172
- Kanehisa M, Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucl Acids Res* 28:27–30
- Karlebach G, Shamir R (2008) Modelling and analysis of gene regulatory networks. *Nat Rev Mol Cell Biol* 9:770–780
- Kauffman S (1969) Metabolic stability and epigenesis in randomly constructed gene nets. *J Theor Biol* 44:167–190
- Kelly D, Sanders R (2008) Assessing the quality of scientific software. In: *Proceedings of the 1st international workshop on software engineering for computational science and engineering*
- 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* 100:11,394–11,399

- Kim JR, Yoon Y, Cho KH (2008) Coupled feedback loops form dynamic motifs of cellular networks. *Biophys J* 94:359–365
- Kitano H (2007a) A robustness-based approach to systems-oriented drug design. *Nat Rev Drug Design* 6:202–210
- Kitano H (2007b) Towards a theory of biological robustness. *Mol Syst Biol* 3:137
- Koller D, Friedman N (2009) Probabilistic graphical models: principles and techniques. The MIT Press, Cambridge
- Küffner R, Petri T, Windhager L, Zimmer R (2010) Petri nets with fuzzy logic (pnfl): reverse engineering and parametrization. *PLoS One* 5:e12807
- Kwon YK, Cho KH (2008) Quantitative analysis of robustness and fragility in biological networks based on feedback dynamics. *Bioinformatics* 24:987–994
- Laird PW (2010) Principles and challenges of genome-wide DNA methylation analysis. *Nat Rev Genet* 11:191–203
- Lander A (2010) The edges of understanding. *BMC Biology* 8:40
- Lee TI, Rinaldi NJ, Robert F, Odom DT, Bar-Joseph Z, Gerber GK, Hannett NM, Harbison CT, Thompson CM, Simon I, Zeitlinger J, Jennings EG, Murray HL, Gordon B, Ren B, Wyrick JJ, Tagne JB, Volkert TL, Fraenkel E, Gifford DK, Young RA (2002) Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* 298:799–804
- Liang S, Fuhrmann S, Somogyi R (1998) REVEAL, a general reverse engineering algorithm for inference of genetic network architectures. *Pac Symp Biocomput* 3:18–29
- Liang Z, Xu M, Teng M, Niu L (2006) Comparison of protein interaction networks reveals species conservation and divergence. *BMC Bioinformatics* 7:457
- Luscombe NM, Babu MM, Yu H, Snyder M, Teichmann SA, Gerstein M (2004) Genomic analysis of regulatory network dynamics reveals large topological changes. *Nature* 431:308–312
- Ma HW, Kumar B, Ditges U, Gunzer F, Buer J, Zeng AP (2004) An extended transcriptional regulatory network of *Escherichia coli* and analysis of its hierarchical structure and network motifs. *Nucl Acids Res* 32:6643–6649
- Martin S, Zhang Z, Martino A, Faulon JL (2007) Boolean dynamics of genetic regulatory networks inferred from microarray time series data. *Bioinformatics* 23:866–874
- Matsuno H, Doi A, Nagasaki M, Miyano S (2000) Hybrid petri net representation of gene regulatory network. *Pac Symp Biocompt* 5:338–349
- Mendell JT, Sharifi NA, Meyers JL, Martinez-Murillo F, Dietz HC (2004) Nonsense surveillance regulates expression of diverse classes of mammalian transcripts and mutes genomic noise. *Nat Genet* 36:1073–1078
- Mendes P, Sha W, Ye K (2003) Artificial gene networks for objective comparison of analysis algorithms. *Bioinformatics* 19:ii122–ii129
- Metzker ML (2010) Sequencing technologies—the next generation. *Nat Rev, Genet* 11:31–46
- Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovshii D, Alon U (2002) Network motifs: simple building blocks of complex networks. *Science* 298:824–827
- Nagasaki M, Yamaguchi R, Yoshida R, Imoto S, Doi A, Tamada Y, Matsuno H, Miyano S, Higuchi T (2006) Genomic data assimilation for estimating hybrid functional Petri net from time-course gene expression data. *Genome Inform* 17:46–61
- Newman M, Barabási AL, Watts DJ (2006) The structure and dynamics of networks. Princeton University Press, Princeton, NJ
- Noble D (2002) The rise of computational biology. *Nat Rev Mol Cell Biol* 3:459–463
- Noble D (2008) Genes and causation. *Phil Trans R Soc A* 366:3001–3015
- Ogata H, Fujibuchi W, Goto S, Kanehisa M (2000) A heuristic graph comparison algorithm and its application to detect functionally related enzyme clusters. *Nucl Acids Res* 28:4021–4028
- Park PJ (2009) ChIP-seq: advantages and challenges of a maturing technology. *Nat Rev Genet* 10:669–680
- Pearl J (1988) Probabilistic reasoning in intelligent systems. Morgan Kaufmann, Massachusetts
- Pearl J (2000) Causality: models, reasoning, and inference. Cambridge University Press, Cambridge
- Pe'er D (2005) Bayesian network analysis of signaling networks: a primer. *Sci STKE* 2005:14
- Petri CA (1962) Kommunikation mit automaten. Ph.D. thesis, Institut für Instrumentelle Mathematik, Bonn
- Pinney J, Westhead D, McConkey G (2003) Petri net representations in systems biology. *Biochem Soc Trans* 31:1513–1515
- Quackenbush J (2003) Microarrays—guilt by association. *Science* 302:240–241
- Reisig W (1985) Petri nets: an introduction. Monographs on Theoretical Computer Science. Springer, Berlin
- Reisig W, Rozenberg G (eds) (1998) Lectures on Petri nets I: basic models. Lecture notes in computer science. Springer, Berlin
- Rodriguez-Caso C, Medina MA, Solé RV (2005) Topology, tinkering and evolution of the human transcription factor network. *FEBS J* 272:6423–6434
- 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
- Salgado H, Gamma-Castro S, Peralta-Gil M, Díaz-Peredo E, Sánchez-Solano F, Santo-Zavaleta A, Martínez-Flores I, Jiménez-Jacinto V, Bonavides-Martínez C, Segura-Salazar J, Martínez-Antonio A, Collado-Vides J (2006) RegulonDB (version 5.0): *Escherichia coli* k-12 transcriptional regulatory network, operon organization, and growth conditions. *Nucl Acids Res* 34:D394–D397
- Sandelin A, Alkema W, Engstrom P, Wasserman WW, Lenhard B (2004) JASPAR: an open-access database for eukaryotic transcription factor binding profiles. *Nucl Acids Res* 32:D91–D94
- Schaefer CF, Anthony K, Krupa S, Buchoff J, Day M, Hannay T, Buetow KH (2009) PID: the pathway interaction database. *Nucl Acids Res* 37:D674–D679
- Schwartz R (2008) Biological modeling and simulation. The MIT Press, Cambridge
- Segal E, Wang H, Koller D (2003) Discovering molecular pathways from protein interaction and gene expression data. *Bioinformatics* 19:i264–i272
- Sharan R, Ideker T (2006) Modeling cellular machinery through biological network comparison. *Nat Biotechnol* 24:427–433
- Shen-Orr SS, Milo R, Mangan S, Alon U (2002) Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nat Genet* 31:64–68
- Smith VA, Jarvis ED, Hartemink AJ (2003) Influence of network topology and data collection on network inference. *Pac Symp Biocomput* 8:164–175
- Steggles LJ, Banks R, Shaw O, Wipat A (2007) Qualitatively modelling and analysing genetic regulatory networks: a Petri net approach. *Bioinformatics* 23:336–343
- Stolovitzky G, Monroe D, Califano A (2007) Dialogue on reverse-engineering assessment and methods: the DREAM of high-throughput pathway inference. *Ann N Y Acad Sci* 1115:1–22

- Strahl BD, Allis CD (2000) The language of covalent histone modifications. *Nature* 403:41–45
- Strogatz SH (2001) Exploring complex networks. *Nature* 410:268–276
- Sutherland H, Bickmore WA (2009) Transcription factories: gene expression in unions? *Nat Rev Genet* 10:457–466
- Suzuki MM, Bird A (2008) DNA methylation landscapes: provocative insights from epigenomics. *Nat Rev Genet* 9:465–476
- Szallasi Z, Stelling J, Periwal V (eds) (2006) *System modeling in cell biology: from concept to nuts and bolts*. The MIT Press, Cambridge
- Tong AHY, Lesage G, Bader G, Ding H, Xu H, Xin X, Young J, Berriz GF, Brost RL, Chang M, Chen YQ, Cheng X, Chua G, Friesen H, Goldberg DS, Haynes J, Humphries C, He G, Hussein S, Ke L, Krogan N, Li Z, Levinson JN, Lu H, Menard P, Munyana C, Parsons A, Ryan O, Tonikian R, Roberts T, Sdicu AM, Shapiro J, Sheikh B, Suter B, Wong SL, Zhang LV, Zhu H, Burd CG, Munro S, Sander C, Rine J, Greenblatt J, Roth FP, Brown GW, Andrews B, Bussey H, Boone C (2004) Global mapping of the yeast genetic interaction network. *Science* 303:808–813
- Trusina A, Sneppen K, Dodd IB, Shearwin KE, Egan JB (2005) Functional alignment of regulatory networks: a study of temperate phages. *PLoS Comput Biol* 1:e74
- Vaske CJ, Benz SC, Sanborn JZ, Earl D, Szeto C, Zhu J, Haussler D, Stuart JM (2010) Inference of patient-specific pathway activities from multi-dimensional cancer genomics data using PARADIGM. *Bioinformatics* 26:i237–i245
- Wagner A (2000) Robustness against mutations in genetic networks of yeast. *Nat Genet* 24:355–361
- Wagner A (2003) How the global structure of protein interaction networks evolves. *Proc R Soc Lond B* 270:457–466
- Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* 10:57–63
- Wolfe CJ, Kohane IS, Butte AJ (2005) Systematic survey reveals general applicability of “guilt-by-association” within gene coexpression networks. *BMC Bioinformatics* 6:227
- Xie X, Ho JWK, Murphy C, Kaiser G, Xu B, Chen TY (2009) Application of metamorphic testing to supervised classifiers. In: *Proceedings of the 9th international conference on quality software (QSIC’09)*. pp. 135–144
- Zhang B, Horvath S (2005) A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol* 4:17
- Zou M, Conzen SD (2005) A new dynamic Bayesian network (DBN) approach for identifying gene regulatory networks from time course microarray data. *Bioinformatics* 21: 71–79