11-1-2007

Comparison of Probabilistic Boolean Network and Dynamic Bayesian Network Approaches for Inferring Gene Regulatory Networks

Peng Li
University of Southern Mississippi

Chaoyang Zhang
University of Southern Mississippi, Chaoyang.Zhang@usm.edu

Edward J. Perkins
U.S. Army Engineer Research and Development Center, edward.j.perskins@erdc.usace.army.mil

Ping Gong
SpecPro, Inc., Ping.Gong@usace.army.mil

Youping Deng
University of Southern Mississippi, youping.deng@usm.edu

BMC Bioinformatics

BioMed Central

Proceedings

Open Access

Comparison of probabilistic Boolean network and dynamic Bayesian network approaches for inferring gene regulatory networks

Peng Li[1], Chaoyang Zhang*[1], Edward J Perkins[2], Ping Gong[3] and Youping Deng*[4]

Address: [1]School of Computing, University of Southern Mississippi, Hattiesburg, MS 39406, USA, [2]Environmental Laboratory, U.S. Army Engineer Research and Development Center, 3909 Halls Ferry Rd. Vicksburg, MS, 39180, USA, [3]SpecPro Inc., 3909 Halls Ferry Rd, Vicksburg, MS, 39180, USA and [4]Department of Biological Sciences, University of Southern Mississippi, Hattiesburg, MS 39406, USA

Email: Peng Li - peng.li@usm.edu; Chaoyang Zhang* - chaoyang.zhang@usm.edu; Edward J Perkins - Edward.J.Perkins@erdc.usace.army.mil; Ping Gong - Ping.Gong@erdc.usace.army.mil; Youping Deng* - youping.deng@usm.edu

* Corresponding authors

from Fourth Annual MCBIOS Conference. Computational Frontiers in Biomedicine
New Orleans, LA, USA. 1–3 February 2007

Published: 1 November 2007

BMC Bioinformatics 2007, **8**(Suppl 7):S13 doi:10.1186/1471-2105-8-S7-S13

This article is available from: http://www.biomedcentral.com/1471-2105/8/S7/S13

Abstract

Background: The regulation of gene expression is achieved through gene regulatory networks (GRNs) in which collections of genes interact with one another and other substances in a cell. In order to understand the underlying function of organisms, it is necessary to study the behavior of genes in a gene regulatory network context. Several computational approaches are available for modeling gene regulatory networks with different datasets. In order to optimize modeling of GRN, these approaches must be compared and evaluated in terms of accuracy and efficiency.

Results: In this paper, two important computational approaches for modeling gene regulatory networks, probabilistic Boolean network methods and dynamic Bayesian network methods, are compared using a biological time-series dataset from the Drosophila Interaction Database to construct a Drosophila gene network. A subset of time points and gene samples from the whole dataset is used to evaluate the performance of these two approaches.

Conclusion: The comparison indicates that both approaches had good performance in modeling the gene regulatory networks. The accuracy in terms of recall and precision can be improved if a smaller subset of genes is selected for inferring GRNs. The accuracy of both approaches is dependent upon the number of selected genes and time points of gene samples. In all tested cases, DBN identified more gene interactions and gave better recall than PBN.

Background

The development of high-throughput genomic technologies (i.e., DNA microarrays), makes it possible to study dependencies and regulation among genes on a genome-

wide scale. In last decade, the amount of gene expression data has increased rapidly necessitating development of computational methods and mathematical techniques to analyze the resulting massive data sets. In order to understand the functioning of cellular organisms, why complicated response patterns to stressors are observed, and provide a hypothesis for experimental verification, it is necessary to model gene regulatory networks (GRNs). Currently, clustering, classification and visualization methods are used for reconstruction or inference of gene regulatory networks from gene expression data sets. These methods generally group genes based on the similarity of expression patterns. Based on large-scale microarray data retrieved from biological experiments, many computational approaches have been proposed to reconstruct genetic regulatory networks, such as Boolean networks [1,2], differential equations [1,3], Bayesian networks [4-6] and neural networks [7]. Among these approaches, Boolean network methods and Bayesian network methods have drawn the most interest in the field of systems biology.

Much recent work has been done to reconstruct gene regulatory networks from expression data using Bayesian networks and dynamic Bayesian network (DBN). Bayesian network approaches have been used in modeling genetic regulatory networks because of its probabilistic nature. However, drawbacks of Bayesian network approaches include failure to capture temporal information and modeling of cyclic networks. DBN is better suited for characterizing time series gene expression data than the static version. Perrin et al. [8] used a stochastic machine learning method to model gene interactions and it was capable of handling missing variables. Zou et al. [9] presented a DBN-based approach, in which the number of potential regulators is limited to reduce search space. Yu et al. [10] developed a simulation approach to improve DBN inference algorithms, especially in the context of limited quantities of biological data. In [11], Xing and Wu proposed a higher order Markov DBN to model multiple time units in a delayed gene regulatory network. Recently, likelihood maximization algorithms such as the Expectation-Maximization (EM) algorithm have been used to infer hidden parameters and deal with missing data [12].

The Boolean Network model, originally introduced by Kauffman [1,13,14] is also very useful to infer gene regulatory networks because it can monitor the dynamic behaviour in complicated systems based on large amounts of gene expression data [15-17]. One of the main objectives of Boolean network models is to study the logical interactions of genes without knowing specific details [17,18]. In a Boolean network (BN), the target gene is predicted by other genes through a Boolean function. A probabilistic Boolean network (PBN), first introduced by Shmulevich et al. in [16,19] is the stochastic extension of Boolean network. It consists of a family of Boolean networks, each of which corresponds to a contextual condition determined by variables outside the model. As models of genetic regulatory networks, the PBN method has been further developed by several authors. In [20], a model for random gene perturbations was developed to derive an explicit formula for the transition probabilities in the new PBN model. In [21], intervention is treated via external control variables in a context-sensitive PBN by extending the results for instantaneously random PBN in several directions. Some learning approaches for PBN have also been explored [22-24]. Considering the same joint probability distribution over common variables, several fundamental relationships of two model classes (PBN and DBN) have been discussed in [25].

In this paper, two important computational approaches for modeling gene regulatory networks, PBN and DBN, are compared using a biological time-series dataset from the Drosophila Interaction Database [26] to construct a Drosophila gene network. We present the PBN and DBN approaches and GRN construction methods used and discuss the performance of the two approaches in constructing GRNs.

Results

A real biological time series data set (Drosophila genes network from Drosophila Interaction Database) was used to compare PBN and DBN approaches for modeling gene regulatory networks [27,28]. The raw data was preprocessed in the same way as given in [29]. There were 4028 gene samples with 74 time points available in *Drosophila melanogaster* genes network through the four stages of the life cycle: embryonic, larval, pupal and adulthood [27]. An example network of drosophila muscle development is given in [29], in which muscle-specific protein 300 (*Msp*-300) is treated as hub gene in their inferred network. We used a different subset of the genes which participate in the development of muscle. Particularly, *Mlp*84B and other genes which contribute to larval somatic muscle development were used to infer gene regulatory networks.

The *D. melanogaster* gene Muscle LIM protein at 84B (abbreviated as *Mlp*84B) has also been known in FlyBase as Lim3. It encodes a product with putative protein binding involved in myogenesis which is a component of the cytoplasm. It is expressed in the embryo (larval somatic muscle, larval visceral muscle, muscle attachment site, pharyngeal muscle and two other listed tissues). Table 1 shows the scores of *Mlp*84B interacting with other related genes [26].

Here, we first selected 12 genes to infer GRNs using PBN and DBN. The constructed GRNs are shown in Figure 1.

Table 1: The interactions and scores of Mlp84B with other genes

High Confidence	Scores	Other interactions	Scores
CG10722	0.5642	*Cdk7*	0.3569
CG13501	0.9005	*Impe1*	0.1108
CG17440	0.5811	*Pfk*	0.3155
CG7046	0.6626	*TfIIB*	0.2436
CG7447	0.5411	*Stck*	0.2523
CG11115 (*Ssl1*)	0.7917	*tup*	0.1094

There exists 18 interactions totally within this small larval somatic muscle network [26]. 10 and 12 interactions in the network have been successful identified. Most interactions between *Mlp*84B and genes with high confidence have been inferred.

More comparison results of PBN(n, e) and DBN(n, e) are given in Table 2, where n is the number of nodes (genes) in network and e the number of edges (interactions) among the nodes. PBN(30, 60) means that there are 30 nodes and 60 edges in that PBN simulation. To analyze the effect of network size on the inference accuracy, four (n, e) combinations, (12, 18), (20, 35), (30, 60) and (40, 80), were considered for inferring gene network. For each combination, we randomly selected five subsets of genes of the same numbers of genes and edges from the Drosophila gene network. For each subset genes, we inferred a gene network and retrieved the number of correct edges *Ce*, miss errors *Me*, and false alarm errors *Fe*. For each combination (n, e), the average and range of *Ce*, *Me* and *Fe* were calculated, as given in Table 2. A correct edge is the one that exists in a real network (i.e. the Drosophila gene network) and is successfully identified by the inference methods. Miss error is defined as the edge between two genes that exists in a real network, but the inference algorithms miss or make wrong orientations. False alarm error is the edge that the inference algorithms create but does not exist in the real network.

We used the benchmark measures recall *R* and precision *P* to evaluate performances of inference algorithms for PBN and DBN. While different definitions for recall and precision exist [30], in this paper, *R* is defined as $Ce/(Ce + Me)$ and *P* is represented as $Ce/(Ce + Fe)$. The selection of subset genes in network was based on the current existing gene interactions and network diagram in the Drosophila genes network [26].

The results in Table 2 show that for the same (n, e) case, DBN reduces miss errors but increases false alarms errors slightly. For all cases, DBN can identify more corrected edges than PBN and hence improve recall. The precision

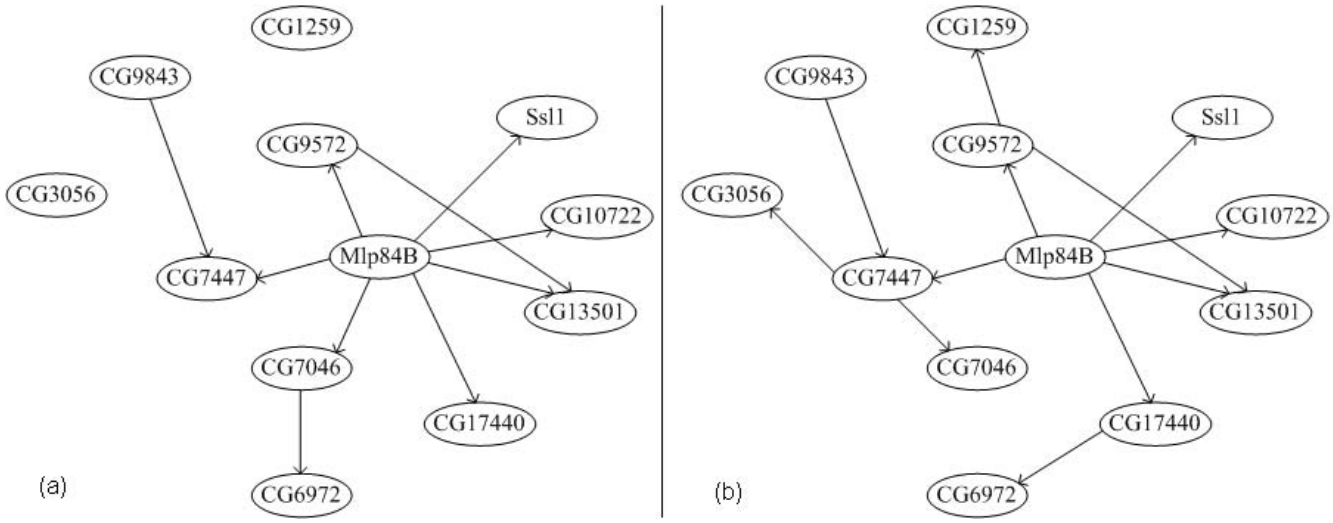

Figure 1
Drosophila larval somatic muscle development network. The genetic network inferred by PBN. (b)The genetic network inferred by DBN

Table 2: Comparison of PBN and DBN methods using different sample networks

	Miss errors *Me*			False alarm errors *Fe*			Correct edges *Ce*			Accuracy (%) (*R*, *P*)		Time(s) *T*
	min	max	avg	min	max	avg	min	max	avg	recall	precision	avg
PBN(12,18)	2	9	6.4	0	4	2.4	6	9	7.8	54.9	76.5	13.2
PBN(20,35)	12	22	16.8	3	6	4.8	11	15	13.6	44.7	73.9	19.7
PBN(30,60)	33	41	36.0	7	10	8.0	17	20	18.4	33.8	69.6	27.9
PBN(40,80)	48	63	55.4	4	6	5.6	18	22	19.6	26.1	77.8	39.2
DBN(12,18)	3	8	5.8	1	3	2.2	9	11	10.4	64.2	82.5	20.1
DBN(20,35)	13	17	15.2	4	7	5.4	14	18	16.8	52.5	75.7	36.0
DBN(30,60)	30	39	33.6	11	15	12.6	24	30	20.2	37.5	61.6	50.6
DBN(40,80)	46	57	51.2	5	9	7.4	28	34	22.8	30.8	75.5	87.6

of DBN is better in three cases but worse in one case than PBN. For both PBN and DBN methods, recall and precision decrease if the number of genes increases. One can see that if more genes are selected for inferring GRNs, the network contains more edges and it is more difficult to successfully identify the interactions among genes. While the DBN method can give better recall of identifying genetic network interactions, it is more time-consuming than PBN.

Discussion

It is challenging to infer GRNs from time series gene expression data. Among thousands of genes, each gene interacts with one or more other genes directly or indirectly through complex dynamic and nonlinear relationships, time series data used to infer genetic networks have low-sample size compared to the number of genes, and gene expression data may contain a substantial amount of noise. Different approaches may have different performances for different datasets. Moreover, inference accuracy depends not only upon models but also on inference schemes. In this paper, we only select two representative inference algorithms for PBN and DBN to model the GRNs, respectively. It is desirable to perform a more comprehensive evaluation of the two approaches with different inference methods and to develop the more robust algorithm and techniques to improve the accuracy of inferring GRNs.

Conclusion

PBN-based and DBN-based methods were used for inferring GRNs from Drosophila time series dataset with 74 time points obtained from the Drosophila Interaction Database. The results showed that accuracy in terms of recall and precision can be improved if a smaller subset of genes is selected for inferring GRNs. Both PBN and DBN approaches had good performance in modeling the gene regulatory networks. In all tested cases, DBN identified more gene interactions and gave better recall than PBN. The accuracy of inferring GRNs was not only dependent upon the model selection but also relied on the particular inference algorithms that were selected for implementation. Different inference schemes may be applied to improve accuracy and performance.

Methods

Boolean network and probabilistic Boolean network

In a BN, the expression level of a target gene is functionally related to the expression states of other genes using logical rules, and the target gene is updated by other genes through a Boolean function. There are only two gene expression levels (states) in a Boolean network (BN): on and off, which are represented as "activated" and "inhibited". A probabilistic Boolean network (PBN) consists of a family of Boolean networks and incorporates rule-based dependencies between variables. In a PBN model, BNs are allowed to switch from one to another with certain probabilities during state transitions. Since PBN is more suitable for GRN reconstruction from time series data and a Boolean network is just a special case of PBN and we only consider PBN for comparison.

Boolean network

We use the same definition as in [2,18] for a Boolean network. A Boolean network $G(V, F)$ is defined by a set of nodes (variables) representing genes $V = \{x_1, x_2, ..., x_n\}$ (where $x_i \in \{0, 1\}$ is a binary variable) and a set of Boolean functions $F = \{f_1, f_2, ..., f_n\}$, which represents the transitional relationships between different time points. A Boolean function $f(x_{j_1(i)}, x_{j_2(i)}, ..., x_{j_{k(i)}(i)})$ with $k(i)$ specified input nodes is assigned to node x_i. The gene status (state) at time point $t + 1$ is determined by the values of some other genes at previous time point t using one Boolean function f_i taken from a set of Boolean functions F. So we can define the transitions as

$$x_i(t+1) = f(x_{j_1(i)}(t), x_{j_2(i)}(t), ..., x_{j_{k(i)}(i)}(t))$$

where each x_i represents the expression value of gene i, if $x_i = 0$, gene i is inhibited; if $x_i = 1$, it is activated. The variable $j_{k(i)}$ represents the mapping between gene networks at different time points. Boolean function F represents the rules of regulatory interactions between genes.

Probabilistic Boolean network

Probabilistic Boolean network inference is the extension of Boolean network methods to combine more than one possible transition Boolean functions, so that each one can be randomly selected to update the target gene based on the selection probability, which is proportional to the coefficient of determination (COD) of each Boolean function. Here we briefly give the same notation of PBN as in [19]. The same set of nodes $V = \{x_1, x_2,..., x_n\}$ as in a Boolean network is used in a PBN $G(V, F)$, but the list of function sets $F = \{f_1, f_2,..., f_n\}$ is replaced by $F = \{F_1, F_2,... F_n\}$, where each function set $F_i = \{f_j^{(i)}\}_{j=1,2,...l(i)}$ composed of $l(i)$ possible Boolean functions corresponds to each node x_i. A realization of the PBN at a given time point is determined by a vector of Boolean functions. Each realization of the PBN maps one of the vector functions $f_k = (f_{k(1)}^{(1)}, f_{k(2)}^{(2)}, ... f_{k(n)}^{(n)})$, $1 \le k \le N$, $1 \le k(i) \le l(i)$, where $f_{k(i)}^{(i)} \in F_i$ and N is the number of possible realizations. Given the values of all genes in network at time point t and a realization f_k, the state of the genes after one updating step is expressed as

$$(x_1(t+1), x_2(t+1),... x_n(t+1)) = f_k(x_1(t), x_2(t),... x_n(t))$$

Let $f = (f^{(1)}, f^{(2)},... f^{(n)})$ denote a random vector taking values in $F_1 \times F_2 \cup \times F_n$. The probability that a specific transition function $f_j^{(i)}$, $(1 \le j \le l(i))$ is used to update gene i is equal to

$$c_j^{(i)} = \Pr\{f^{(i)} = f_j^{(i)}\} = \sum_{k: f_{k(i)}^{(i)} = f_j^{(i)}} \Pr\{f = f_k\}$$

Given genes $V = \{x_1, x_2,..., x_n\}$, each x_i is assigned to a set of Boolean functions $F_i = \{f_j^{(i)}\}_{j=1,2,...l(i)}$ to update target gene. The PBN will reduce to a standard Boolean network if $l(i) = 1$ for all genes. A basic building block of a PBN describing the updating mechanism is shown in Figure 2.

Construction of GRNs from PBN

The Coefficient of Determination (COD) is used to select a list of predictors for a given gene [19,23]. So far, most learning methods for reconstructing gene regulatory network use COD to select predictors for each target gene at any time point t. COD has also been used previously for the steady state data sets. Here we use upper case letters to represent random variables: Let X_i be the target gene, $X_1^{(i)}, X_2^{(i)}, \ldots, X_{l(i)}^{(i)}$ be sets of genes and $f_1^{(i)}, f_2^{(i)}, \ldots, f_{l(i)}^{(i)}$ be available Boolean functions. Thus, the optimal predictors of X_i can be defined by $f_1^{(i)}(X_1^{(i)}), f_2^{(i)}(X_2^{(i)}), \ldots, f_{l(i)}^{(i)}(X_{l(i)}^{(i)})$ and the probabilistic error measure can be represented as $\varepsilon(X_i, f_k^{(i)}(X_k^{(i)}))$. For each k, the COD for X_i relative to the conditioning set $X_k^{(i)}$ is defined by

$$\omega_k^i = \frac{\varepsilon_i - \varepsilon(X_i, f_k^{(i)}(X_k^{(i)}))}{\varepsilon_i}$$

where ε_i is the error of the best estimate of X_i [23].

Now, if a class of gene sets $X_1^{(i)}, X_2^{(i)}, \ldots, X_{l(i)}^{(i)}$ which have high CODs has been selected, we can use the optimal Boolean functions $f_1^{(i)}, f_2^{(i)}, \ldots, f_{l(i)}^{(i)}$ as the rule set for gene X_i, with the probability of $f_j^{(i)}$ being chosen (see(3)). Then the approximations are given by

$$c_k^{(i)} = \frac{\omega_k^i}{\sum_{j=1}^{l(i)} \omega_j^i}$$

According to the above expressions [19,23], those Boolean functions corresponding to the highest CODs will be selected in the probabilistic network. The selected Boolean functions are used to predict the gene expression status at the next time point, and they also will be used to reconstruct gene regulatory networks.

Bayesian networks and dynamic Bayesian networks

Among the many computational approaches that infer gene regulatory networks from time series data, Bayesian network analysis draws significant attention because of its probabilistic nature. DBN is the temporal extension of Bayesian network analysis. It is a general model class that is capable of representing complex temporal stochastic processes. It captures several other often used modeling frameworks as its special cases, such as hidden Markov models (and its variants) and Kalman filter models.

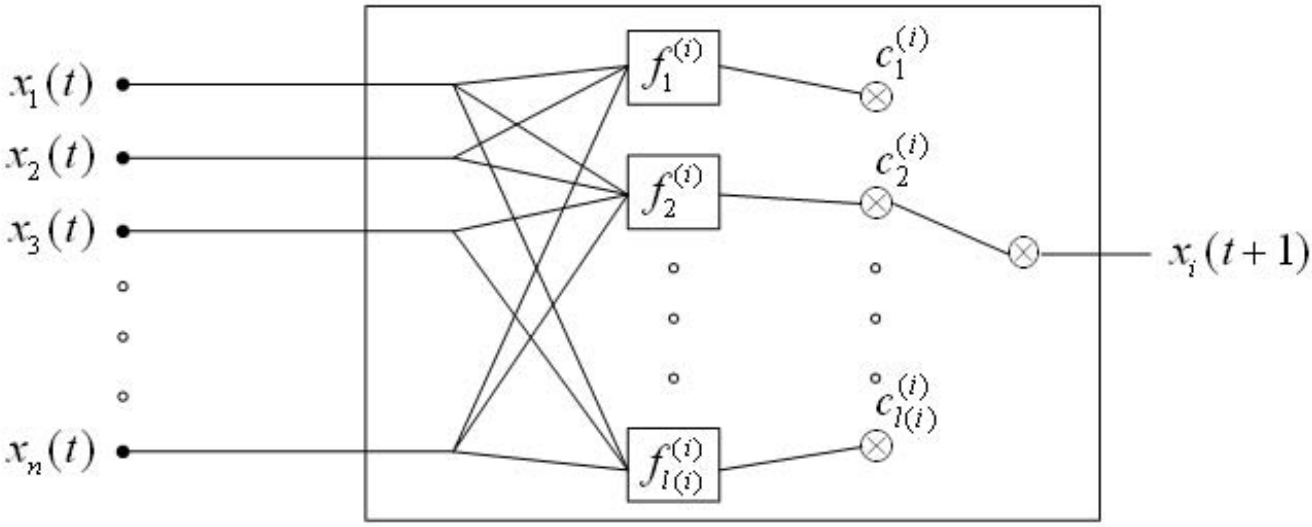

Figure 2
A basic building block of a PBN.

Bayesian network

Given a set of variables $U = \{x_1, x_2, \ldots x_n\}$ in gene network, a Bayesian network, for U is a pair $B = (G, \Theta)$ which encodes a joint probability distribution over all states of U. It is composed of a directed acyclic graph (DAG) G whose nodes correspond to the variables in U and Θ which defines a set of local conditional probability distributions (CPD) to qualify the network. Let $\mathrm{Pa}(x_i)$ denote the parents of the variables x_i in the acyclic graph G and $\mathrm{pa}(x_i)$ denote the values of the corresponding variables. Given G and Θ, a Bayesian network defines a unique joint probability distribution over U given by

$$\Pr\{x_1, x_2, \ldots x_n\} = \prod_{i=1}^{n} \Pr\{x_i \mid \mathrm{pa}(x_i)\}$$

For more detail on Bayesian networks, see [24].

Dynamic Bayesian network

A DBN is defined by a pair (B_0, B_1) represents the joint probability distribution over all possible time series of variables $\mathrm{X} = \{X_1, X_2, \ldots X_n\}$, where $X_i (1 \le i \le n)$ represents the binary-valued random variables in the network, besides, we use lower case $x_i (1 \le i \le n)$ denotes the values of variable X_i. It is composed of an initial state of Bayesian network $B_0 = (G_0, \Theta_0)$ and a transition Bayesian network $B_1 = (G_1, \Theta_1)$, where B_0specifies the joint distribution of the variables in $X(0)$ and B_1 represents the transition probabilities $\Pr\{\mathrm{X}(t+1) \mid \mathrm{X}(t)\}$ for all t. In slice 0, the parents of $X_i(0)$ are assumed to be those specified in the prior network B_0, which means $\mathrm{Pa}(X_i(0)) \subseteq \mathrm{X}(0)$ for all $1 \le i \le n$; in slice $t + 1$, the parents of $X_i(t + 1)$ are nodes in slices t, $\mathrm{Pa}(X_i(t+1)) \subseteq \mathrm{X}(t)$ for all $1 \le i \le n$ and $t \ge 0$, as stated in [25], the connections only exist between consecutive slices. The joint distribution over a finite list of random variables $\mathrm{X}(0) \cup \mathrm{X}(1) \cup \cup \cup \mathrm{X}(\mathrm{T})$ can be expressed as [24,25]

$$\begin{aligned}
&\Pr\{\mathrm{x}(0), \mathrm{x}(1), \ldots, \mathrm{x}(\mathrm{T})\} \\
&= \Pr\{\mathrm{x}(0)\} \prod_{\mathrm{t}=0}^{\mathrm{T}-1} \Pr\{\mathrm{x}(t+1) \mid \mathrm{x}(t)\} \\
&= \prod_{i=1}^{n} \Pr\{x_i(0) \mid \mathrm{pa}(X_i(0))\} \times \prod_{t=0}^{\mathrm{T}-1} \prod_{j=1}^{n} \Pr\{x_j(t+1) \mid \mathrm{pa}(X_j(t+1))\}
\end{aligned}$$

An example of a DBN is shown in Figure 3.

Construction of GRNs from DBN

Given a set of training gene data, how the network structure is found that best fits the data is called learning the structure of a dynamic Bayesian network. The goal of constructing a network is to find the model with maximum likelihood (i.e., REVEAL algorithm in [3] and its improvement in [9]). The network we want to learn is the transition network, i.e., the network defining dependencies between the adjacent time slices $X(t)$ and $X(t + 1)$. The training set of data is composed of all adjacent time-slices $X(t)$ and $X(t + 1)$.

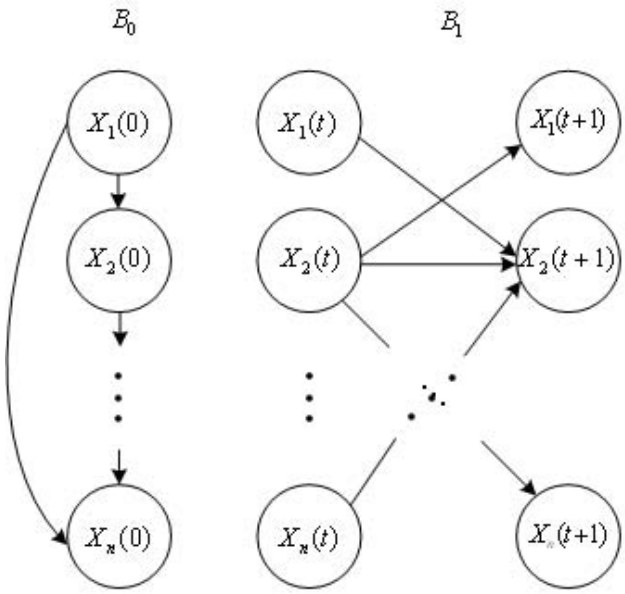

Figure 3
A basic building block of a DBN.

Algorithms for learning gene network structure have focused on networks with complete data. Structural Expectation Maximization (SEM) is developed to handle data with hidden variables and missing values. One of the algorithms to infer network structure from training data is based on the mutual information analysis of the data. For each node, this algorithm learns the optimal parent set independently by choosing the parent set that maximizes a scoring function. The scoring function is defined by

$$\mathrm{I}(X, \mathrm{Pa}(X))/\max\{H(X), H(\mathrm{Pa}(X))\}$$

where $I(X, Y)$ is the mutual information between X and Y, and $H(X)$ is the entropy of X. With parent set of genes in DBN, GRNs can be constructed [31].

For each inferred network, scoring metrics are used to evaluate the probabilistic scores which explain relationships in the given data sets. There are two popular Bayesian scoring metrics: the BDe (Bayesian Dirichlet equivalence) score [32] and the BIC (Bayesian information criterion) score [33]. Then, the network with highest score will be identified using search heuristics, which have three widely used methods: greedy search, simulated appealing and a genetic algorithm [10].

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
PL implemented the algorithms and inferred gene networks. PL and CZ performed the statistical analysis and drafted the manuscript. CZ and YD coordinated the study. EP, PG and YD gave suggestions to improve the methods and revised the manuscript. All authors read and approved the final manuscript.

Acknowledgements
This work was supported by the Army Environmental Quality Program of the US Army Corps of Engineers under contract #W912HZ-05-P-0145. Permission was granted by the Chief of Engineers to publish this information. The project was also supported by the Mississippi Functional Genomics Network (DHHS/NIH/NCRR Grant# 2P20RR016476-04).

This article has been published as part of *BMC Bioinformatics* Volume 8 Supplement 7, 2007: Proceedings of the Fourth Annual MCBIOS Conference. Computational Frontiers in Biomedicine. The full contents of the supplement are available online at http://www.biomedcentral.com/1471-2105/8?issue=S7.

References
1. Kauffman SA: **Metabolic stability and epigenesis in randomly constructed genetic nets.** *J Theor Biol* 1969, **22:**437-467.
2. Akutsu T, Miyano S, Kuhara S: **Identification of genetic networks from a small number of gene expression patterns under the Boolean network model.** *Pacific Symposium on Biocomputing* 1999, **4:**17-28.
3. Chen T, He HL, Church GM: **Modeling gene expression with differential equations.** *Pacific Symposium Biocomputing* 1999, **4:**29-40.
4. Liang S, Fuhrman S, Somogyi R: **REVEAL, A general reverse engineering algorithm for inference of genetic network architectures.** *Pacific Symposium on Biocomputing* 1998, **3:**18-29.
5. Friedman N, Goldszmidt M, Wyner A: **Data analysis with Bayesian networks: A bootstrap approach.** *Proc Fifteenth Conf on Uncertainty in Artificial Intelligence (UAI)* 1999.
6. Imoto S, Goto T, Miyano S: **Estimation of Genetic Networks and Functional Structures Between Genes by Using Bayesian Networks and Nonparametric Regression.** *Pacific Symposium on Biocomputing* 2002, **7:**175-186.
7. Weaver DC, Workman CT, Stormo GD: **Modeling regulatory networks with weight matrices.** *Pacific Symposium on Biocomputing* 1999, **4:**112-123.
8. Perrin BE, Ralaivola L, *et al.*: **Gene networks inference using dynamic Bayesian networks.** *Bioinformatics* 2003, **19(Suppl 2):**II138-II148.
9. Zou M, Conzen SD: **A new dynamic Bayesian network (DBN) approach for identifying gene regulatory networks from time course microarray data.** *Bioinformatics* 2005, **21(1):**71-79.
10. Yu J, *et al.*: **Advances to Bayesian network inference for generating causal networks from observational biological data.** *Bioinformatics* 2004, **20(18):**3594-3603.
11. Xing ZZ, Wu D: **Modeling Multiple Time Units Delayed Gene Regulatory Network Using Dynamic Bayesian Network.** *ICDM Workshops* 2006:190-195.
12. Zhang L, Samaras D, Alia-Klein N, Volkow N, Goldstein R: **Modeling neuronal interactivity using dynamic Bayesian networks.** In *Advances in Neural Information Processing Systems Volume 18*. Edited by: *Weiss Y, Scholkopf B, Platt J. Cambridge, MA: MIT Press*; 2006.
13. Glass K, Kauffman SA: **The logical analysis of continuous, nonlinear biochemical control networks.** *J Theoret Biol* 1973, **39:**103-129.
14. Kauffman SA: **The large scale structure and dynamics of genetic control circuits: an ensemble approach.** *J Theoret Biol* 1974, **44:**167-190.
15. Huang S: **Gene expression profiling, genetic networks and cellular states: An integrating concept for tumorigenesis and drug discovery.** *Journal of Molecular Medicine* 1999, **77:**469-480.
16. Shmulevich I, Gluhovsky I, Hashimoto RF, Dougherty ER, Zhang W: **Steady-state analysis of genetic regulatory networks mod-**

eled by probabilistic Boolean networks. *Comparative and Functional Genomics* 2003, **4:**601-608.
17. Kim H, Lee JK, Park T: **Boolean networks using the chi-square test for inferring large-scale gene regulatory networks.** *BMC Bioinformatics* 2007, **8:**37.
18. Shmulevich I, Dougherty ER, Zhang W: **From Boolean to Probabilistic Boolean Networks as Models of Genetic Regulatory Networks.** *Proceeding of the IEEE* 2002, **90(11):**1778-1792.
19. Shmulevich I, Dougherty ER, Seungchan K, Zhang W: **Probabilistic Boolean networks: a rule-based uncertainty model for gene regulatory networks.** *Bioinformatics* 2002, **18(2):**261-274.
20. Shmulevich I, Dougherty ER, Zhang W: **Gene perturbation and intervention in probabilistic Boolean networks.** *Bioinformatics* 2002, **18(10):**1319-1331.
21. Lähdesmäki H, Shmulevich I, Yli-Harja O: **On Learning Gene Regulatory Networks Under the Boolean Network Model.** *Machine Learning* 2003, **52:**147-167.
22. Zhou X, Wang X, Dougherty ER: **Construction of genomic networks using mutual information clustering and reversible jump Markov-Chain-Monte-Carlo predictor design.** *Signal Processing* 2003, **83(4):**745-761.
23. Dougherty ER, Kim S, Chen Y: **Coefficient of determination in nonlinear signal processing.** *Signal Processing* 2000, **80(10):**2219-2235.
24. Friedman N, Murphy K, Russell S: **Learning the structure of dynamic probabilistic networks.** *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence (UAI)* 1998:139-147.
25. Lahdesmki H, Hautaniemi S, Shmulevich I, Yli-Hrja O: **Relationships between probabilistic Boolean networks and dynamic Bayesian networks as models of gene regulatory networks.** *Signal Processing* 2006, **86(4):**814-834.
26. **Drosophila Interaction Database** [http://portal.curagen.com/cgi-bin/interaction/flyHome.pl]
27. Arbeitman MN, *et al.*: **Gene expression during the life cycle of Drosophila melanogaster.** *Science* 2002, **297:**2270-2275.
28. Giot L, *et al.*: **A protein interaction map of Drosophila melanogaster.** *Science* 2003, **302:**1727-1736.
29. Zhao W, Serpedin E, Dougherty ER: **Inferring gene regulatory networks from time series data using the minimum description length.** *Bioinformatics* 2006, **22(17):**2129-2135.
30. Zhang X, Baral C, Kim S: **An Algorithm to Learn Causal Relations Between Genes from Steady State Data: Simulation and Its Application to Melanoma Dataset.** *Proceedings of 10th Conference on Artificial Intelligence in Medicine (AIME 05), Aberdeen, Scotland* 2005:524-534.
31. Murphy K, Mian S: **Modelling gene expression data using dynamic Bayesian networks.** In *Technical report, Computer Science Division University of California, Berkeley, CA*; 1999.
32. Heckerman D, Geiger D, Chickering DM: **Learning Bayesian networks: The combination of knowledge and statistical data.** *Mach Learning* 1995, **20:**197-243.
33. Schwarz G: **Estimating the dimension of a model.** *Ann Stat* 1978, **6:**461-464.