BIOINFORMATICS ORIGINAL PAPER
Vol. 27 no. 19 2011, pages 2686–2691
doi:10.1093/bioinformatics/btr454

Systems biology

Advance Access publication August 4, 2011

Large-scale dynamic gene regulatory network inference combining differential equation models with local dynamic Bayesian network analysis

Zheng Li[1,*], Ping Li[1,†], Arun Krishnan[2] and Jingdong Liu[1,*]

[1]Monsanto Company, Mail zone CC1A, Chesterfield, MO 63017, USA and [2]Monsanto Research Centre, #44/2A, Vasant Business Park, Hebbal, Bangalore 560092, India

Associate Editor: John Quackenbush

ABSTRACT

Motivation: Reverse engineering gene regulatory networks, especially large size networks from time series gene expression data, remain a challenge to the systems biology community. In this article, a new hybrid algorithm integrating ordinary differential equation models with dynamic Bayesian network analysis, called Differential Equation-based Local Dynamic Bayesian Network (DELDBN), was proposed and implemented for gene regulatory network inference.

Results: The performance of DELDBN was benchmarked with an *in vivo* dataset from yeast. DELDBN significantly improved the accuracy and sensitivity of network inference compared with other approaches. The local causal discovery algorithm implemented in DELDBN also reduced the complexity of the network inference algorithm and improved its scalability to infer larger networks. We have demonstrated the applicability of the approach to a network containing thousands of genes with a dataset from human HeLa cell time series experiments. The local network around BRCA1 was particularly investigated and validated with independent published studies. BRAC1 network was significantly enriched with the known BRCA1-relevant interactions, indicating that DELDBN can effectively infer large size gene regulatory network from time series data.

Availability: The R scripts are provided in File 3 in Supplementary Material.

Contact: zheng.li@monsanto.com; jingdong.liu@monsanto.com

Supplementary information: Supplementary data are available at *Bioinformatics* online.

Received on April 19, 2011; revised on July 18, 2011; accepted on July 31, 2011

1 INTRODUCTION

The increasing availability of high-throughput measurements of transcripts has presented a golden opportunity for elucidating gene regulatory pathways and networks for important traits such as human diseases. Time series datasets such as the temporal measurements in yeast (Spellman *et al.*, 1998) and human cell cycles (Whitfield *et al.*, 2002) are of particular interests for two reasons. First, the temporal measurements capture the expression profile of the genes in a dynamic process, which allows for efficient deconvolution of the regulatory interactions from the data. Second, the time information of a temporal measurement provides an important extra dimension for causality studies based on the intuition that past events has causal effects on current and future events and not vice versa.

It is the motivation of this article to develop a systems biology approach to reverse engineer gene regulatory network from time series data with higher accuracy and better scalability than has been achieved using currently existing methodologies. Various approaches have been developed and applied for this purpose, which generally fall into one of the two categories, namely the model-based approaches and the machine learning-based approaches. Ordinary differential equation (ODE) models were used in the model-based approaches to explicitly represent the dependence of the concentration of one gene's transcripts on that of other genes. The representative example algorithms in this category include network identification by regression (NIR) (Gardner *et al.*, 2003), singular value decomposition and regression analysis (Yeung *et al.*, 2002), mode-of-action by network identification (MNI) (di Bernardo *et al.*, 2005), time series network identification (TSNI) (Cantone *et al.*, 2009) and Inferator (Bonneau *et al.*, 2006). Examples in the second category include the use of various machine learning approaches for network learning, such as partial correlation (Schafer and Stimmer, 2005), Graphic Gaussian Models (GGM) (Schafer and Stimmer, 2005), Dynamic Bayesian network analysis (DBN) (Murpy and Mian, 1999; Yu *et al.*, 2004; Zou and Conen, 2005), state space model (Li *et al.*, 2006a) and Granger causality (Mukhopadhyay and Chatterjee, 2007; Nagarajan and Upreti, 2010). However, the success of these approaches have been limited due to technical challenges such as the difficulty in estimating the parameters in the ODE models and the computational complexity for machine learning approaches like DBN for large size networks and lack of gold standard datasets for the benchmarking of methodologies.

In this article, we propose a new approach, named DELDBN (Differential Equation based Local Dynamic Bayesian Network), to integrate ODE models with dynamic Bayesian analysis to infer gene regulatory networks. ODE models represent a parametric version of mechanistic gene regulations with the advantage of accurately capturing the underlying data generation process. To avoid the challenge of estimating parameters in solving ODE models, we applied a DBN analysis step to connect the left-hand side (transcription rate of a gene) to the right-hand side of an ODE (expression level of regulatory genes). To improve the speedup

*To whom correspondence should be addressed.
†Present address: School of Life Science, Nantong University, P.R.China.

and scalability of the DBN analysis, we implemented a new DBN learning algorithm using local causality discovery algorithm by identifying a Markov Blanket (Margaritis and Thrun, 1999) of the transcription rate. To benchmark the method, we first estimated the performance of the DELDBN algorithm using an *in vivo* benchmark dataset from yeast with a known synthetic gene regulatory network. Our algorithm demonstrated superior accuracy when compared with other known approaches that have previously been applied to the same dataset. Next, we applied the algorithm to the time series expression measurements for thousands of genes of human HeLa cells within a cell cycle. We examined the network around a well studied gene, BRCA1, and found that the BRAC1 network was significantly enriched with the known BRCA1-relevant interactions, indicating that DELDBN can effectively infer large-sized gene regulatory network from time series data.

2 METHODS

2.1 Modeling dynamic gene regulation with ordinary differential equation

Following previous work on the ODE modeling of gene regulation (Bonneau *et al.*, 2006), we model the dynamic gene expression as follows

$$\frac{dX_i(t)}{dt} = \sum \beta_{ij} X_j(t) \quad (1)$$

in which *Xi(t)*, *Xj(t)* represent the expression level of gene i and j at time t, respectively. The items on the right represent the combined effects of all other genes on gene i with β_{ij} representing the effect of gene j on gene i. We use

$$\frac{X_i(t+1) - Xi(t)}{\Delta t}$$

to approximate the transcription rate of gene i when expression data was sampled densely enough, e.g. every 10 min in yeast cell (Cantone *et al.*, 2009). Equation (1) can then be transformed into the following format

$$\frac{X_i(t+1) - X_i(t)}{\Delta t} = \sum \beta_{ij} X_j(t) \quad (2)$$

If time interval is long (e.g. in hours) and transcription rate dx/dt cannot be accurately estimated from data, we use an autoregression model to replace Equation (1)

$$X_i(t+1) = \sum \beta_{ij} X_j(t) \quad (3)$$

While in the steady state case, we learn the gene regulations by uncovering the genes at time t that regulate the expression of gene i at $t+1$, in the dynamic case, we learn the gene regulations by identifying the genes that regulate the transcription rate of gene i at time t. Hence, learning the gene regulatory network is thus reduced to solving the Equations (2) and (3) as described before. Unlike previous studies (Bonneau *et al.*, 2006), we use DBN analysis to learn the dependence of the left-hand side of the equations to the right-hand side.

2.2 Local causality-based

DBN **analysis**

DBN was first introduced and implemented by Murphy et al. to model the gene regulatory network (Murphy and Mian, 1999); however, the intensive computational cost limited the application of DBN to small size gene regulatory networks with 100 or less genes. The search and score-based DBN inference tool Banjo (Yu *et al.*, 2004) is the only DBN tool to our knowledge that can be applied to large-sized networks; however, multiple equivalent structures could have the same scores, thereby making it difficult to obtain a unique, globally optimized solution.

In this article, we propose a new algorithm based on local causal discovery algorithms which have been developed in recent years to address the problem of scalability of Bayesian network learning approaches (Aliferis *et al.*, 2010). Aliferis et al. presented a theoretical analysis of the robustness of local causal discovery algorithms which we briefly summarize here. Local causal discovery algorithm learns the local neighborhood of a specific target variable instead of learning the full causal network. In this study, we use a Markov blanket to identify the local neighborhood. The Markov blanket of a target variable T was defined as a minimal set of variables, by conditioning on which all other variables are independent of T. Sound algorithms have been developed for Markov blanket identification, an example of which is the growth-shrink algorithm (Margaritis and Thrun, 1999). In the growth phase, all variables that are dependent on T are added to MB(T), while in the shrink phase all variables were tested for independence to T within MB(T) to remove the independent variables. However, the GS algorithm uses weak heuristic and requires the sample size to grow exponentially to the size of the Markov blanket (Aliferis *et al.*, 2010). Improved algorithms such as incremental association Markov blanket (IAMB) and its variants from Aliferis's group are needed for large-scale dataset analysis. Therefore, the GS algorithm was used in our first application of yeast synthetic network data and the IAMB algorithm was used in the second application to human HeLa cell cycle data analysis.

Our DBN learning is based on Markov blanket identification. The pseudo algorithm is described as follows.

Algorithm DELDBN
Input: A time series data $X(t=1:N)$ with P genes, N time points
Output: Net_DELDBN with dimension *pxp*

For $i=1:P$//loop through every gene

1: Target $= X(i, 2:N)$ expression of gene i from $t=2$ to N, for steady state measurements with long time intervals, or Target $= X(i, 2:N) - T(i, 1:N-1)$, transcription rate from $t=2$ to N, for dynamic profiles with small time intervals;

2: Select expression data for potential regulators $R = X(:, 1:N-1)$, expression of all genes from $t=1$ to $N-1$.

3: Solve Equation (1) using local causal discovery algorithm through identifying the Markov blanket of target variable T, MB(Target) from its potential regulators.

4: Parent (i) = MB(Target)//Direction assignment. Directions were assigned based upon the intuition that the regulatory effects always go from previous events to the present or future events. Thus, the direction were naturally assigned as MB(T) to T.

5: Net_DELDBN(i, parent(i)) = 1;

End //for loop

The algorithm is illustrated in Figure 1. The BNLearn R package (Scutari, 2010) was used for Markov Blanket identification. The Hygecdf function in Matlab was used for hypergeometric test in evaluating the enrichment score of a subset of genes in a local network.

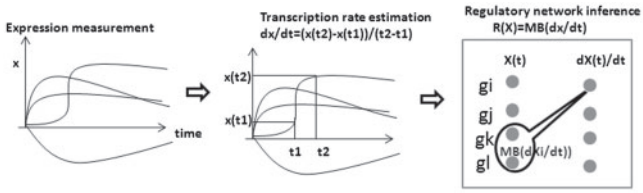

Fig. 1. Schematic illustration of DELDBN. DELDBN infers the causal factors regulating gene x from its potential regulators genes by learning the Markov Blanket of transcription rate gene x at time (T) from expression of its potential regulators at time $(T-1)$.

3 RESULTS

3.1 Improved accuracy of a reverse engineered yeast IRMA network from time series data

In order to evaluate the performance of DELDBN, we first applied the algorithm to a newly published *in vivo* benchmark dataset from yeast (Cantone *et al.*, 2009). A synthetic gene regulatory network was built into the *Saccharomyces cerevisiae*, thus we know *a priori* the underlying network structure. Dynamic profiles of the mRNA expression were measured every 10 min after a perturbation of culture medium caused by a switch between galactose and glucose. The switch-off time series data were used for reconstructing the network structure.

Considering the short time intervals in the data, we used the transcription rate of a target gene for inferring the connections to the target gene. The inference results were shown in Figure 2. Overall our algorithm was able to uncover 10 connections, of which 7 are part of the original network. Following the metrics used by the authors (Cantone *et al.*, 2009), we calculated positive predictive value, PPV = 0.7, and sensitivity, SE = 0.875. We compared the performance to the other algorithms, including TSNI, and DBN in Banjo, as described by Cantone *et al.* (2009) and Gaussian process as described in Aijo and Lahdesmaki (2010) in Table 1. From the comparison, we can see that our algorithm achieved the highest sensitivity with accuracy comparable with other top performing algorithms. We also noticed that one of the three false connections was between connected pair of genes, SWI5-CBF1, but with a reversed direction. It indicated that it was difficult to tell direction between SWI5-CBF1 based on the data, while strong connections were suggested by the data.

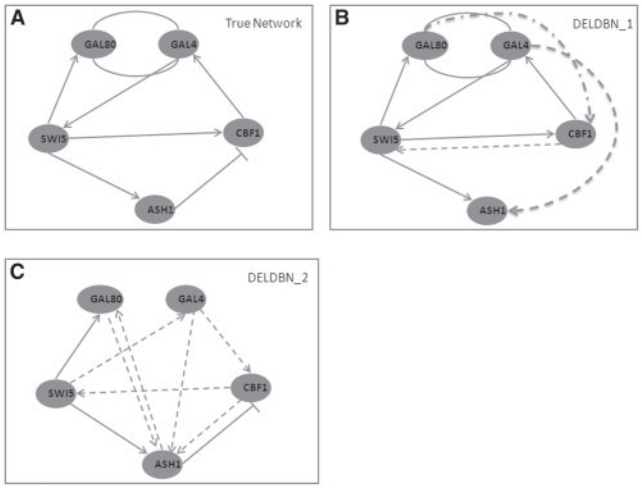

Fig. 2. DELDBN inference of IRMA network. (**A**) True network structure of IRMA network as described in Cantone *et al.* (2009). (**B**) Reverse engineered IRMA network with DELDBN using transcription rate of genes. (**C**) Reverse engineered IRMA network with DELDBN using expression of genes. Comparison indicates transcription rate-based DELDBN performs better for time series data with short time intervals. Solid arrows indicate true connections, and dashed arrows indicate false connections.

Table 1. Comparison of algorithms on IRMA network

Algorithms	True Connections	False connections	PPV	SE
DELDBN	7	3	0.7	0.875
ODE(TSNI)	4	1	0.8	0.5
DBN(Banjo)	2	4	0.3	0.25
GP	6	1	0.85	0.75

The most noticeable improvement with DELDBN is the significant increase of both PPV and sensitivity (SE) when compared with the score-based DBN algorithm in Banjo. We believe that it was the ODE modeling process, i.e. using the time differential of the target variable to identify the regulatory interactors prior to applying the DBN algorithm, improved the performance. We tested the DELDBN inference without the transcription rate estimation step. As shown in Figure 2C, the inference performance (with PPV = 0.3, and SE = 0.375) was indeed much worse than that shown in Figure 2B (with PPV = 0.7, and SE = 0.875), and was comparable to the performance of DBN in Banjo (with PPV = 0.3, SE = 0.25) as reported in Cantone *et al.* (2009). In summary, our new algorithm DELDBN was able to infer the IRMA network with the highest sensitivity and top ranked accuracy. The integration of the ODE model into the DBN process improved the performance of network inference significantly when the time interval is short enough for accurate transcription rate estimation.

3.2 Enhanced scalability and connectivity of a reverse engineered BRCA1 network from human HeLa cell time series data

In the first example, we showed the performance of DELDBN with an *in vivo* benchmark dataset for a small size network. In order to examine the scalability of DELDBN to large sized networks, we next applied DELDBN to a human HeLa cell time series dataset. The dataset was first generated by Whitefield *et al.* (2002) to identify genes that are periodically expressed in cell cycles. Various analyses have been applied to this dataset for reconstructing gene regulatory networks, including the Granger causality analysis (Mukhopadhyay and Chatterjee, 2007; Nagarajan and Upreti, 2010) and time-lagged correlation analysis (Li *et al.*, 2006b). We started the analysis with 1134 genes with periodicity of expression across 48 time points, of which 1099 genes with no missing values in all samples of Experiment 3 were included in our analysis (see File 1 in Supplementary Material for details of the 1099 genes) (Whitefield *et al.*, 2002). The whole dataset was downloaded from http://genome-www.stanford.edu/Human-CellCycle/Hela/data/dataPlusScores_all5.txt.

The time interval in this experiment is 1 h, which led us to assume the steady-state conditions for each time point. Therefore, we used the mRNA expression level, instead of the transcription rate, of the target gene to infer the regulatory interactions. We specifically investigated the first and second neighbors of one particular gene BRCA1 (breast cancer associated gene 1) since it is well studied for its functional role in inducing increased risk of breast cancer and ovarian cancer (Deng, 2006). Many interaction partners have been identified in the literature (Deng, 2006). Perturbation studies have also been done to identify genes responsive to BRCA1 suppression in cell lines (Bae *et al.*, 2005). Such information can thus be used as an independent gold standards to validate the BRCA1 network reconstructed from the HeLa cell time series dataset.

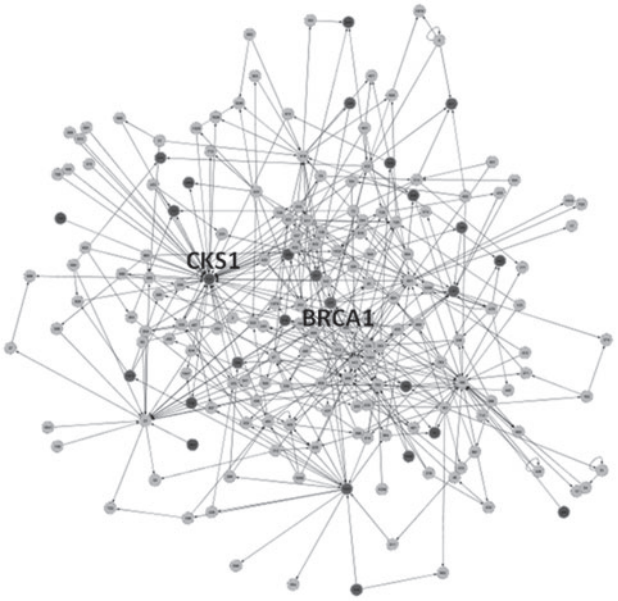

Fig. 3. BRCA1 network learned from human HeLa cell cycle time series data. A total of 184 genes were included this local network, of which 11 were directly connected to BRCA1. Genes highlighted in dark had literature supporting evidence for their connections to BRCA1. The details can be found in cytoscape File 2 in Supplementary Material.

Table 2. Enriched functional gene groups in BRCA1 local network

GeneID	IMAGEID	Annotation
Ubiquitin ($P=0.1653$)		
GENE280X	158165	S phase USP1
GENE944X	810600	M/G1 E2-EPF
GENE374X	1594172	S phase HSPC150
Tubulin ($P=0.0069$)		
GENE551X	745138	G2 TUBA2
GENE493X	2307420	G2 TUBB
GENE589X	321693	G2 delta-tubulin
Kinesin ($P=0.0048$)		
GENE801X	488413	G2 KIF5B
GENE500X	292933	G2 KNSL2
GENE739X	742798	G2/M KNSL6
GENE593X	788256	G2 KNSL2
Histone ($P=0.0342$)		
GENE191X	594693	S phase H4F2
GENE389X	191334	G2 H2AFL
GENE372X	1842170	S phase H4FI

The BRCA1 network is shown in Figure 3, with 185 genes in this local network, of which 11 were directly connected to BRCA1. Some of the important functional categories were summarized in Table 2. The Cytoscape file of the whole network can be found in File 2 in Supplementary Material.

Studies have shown that BRCA1 is involved in the ubiquitination of r-Tubulin, which regulates the centrosome number (Starita *et al.*, 2004). In an independent study, microtubule motor proteins such as kinesins and kinesin-like proteins, and together with tubulin itself were downregulated in BRCA1 silenced cell lines (Bae *et al.*, 2005). Histones were also found to have decreased expression in the same study. As shown in Table 2, within the local network we found all of the above-mentioned genes. We calculated the enrichment of these four categories within BRCA1 neighborhood using a hypergeometric test by comparing the distribution of the functional group within BRCA1 local network to the background network outside of BRCA1 network. The test indicates that Tubulin ($P=0.0048$) and kinesin-like proteins ($P=0.0069$) were highly enriched with $P<0.01$ and histone family proteins ($P=0.0342$) were enriched with $P<0.05$.

Interestingly, two topoisomerases, TopBP1[GENE271X, IMAGE:200136, G1/S TOPBP1 topoisomerase (DNA) II binding protein Hs.91417 R97836 DNA metabolism| DNA replication and chromosome cycle] and Top2A [GENE548X, IMAGE:366971, G2 TOP2A topoisomerase (DNA) II alpha (170 kDa) Hs.156346 AA026682], were also found in the BRCA1 network. TopBP1 has been reported to have functional relations with BRCA1 in regulating G2-M cell cycle checkpoint (Yamane *et al.*, 2003) based upon the evidences that (i) both proteins have the BRCT motif; (ii) both are strongly induced in S phase; (iii) both colocalize under DNA damage conditions; (iv) both are phosphorylation targets of ATM under DNA damage; and (v) double knockout of both had significant induction of cell death than each single knockout. Top2A on the other hand has been found in the perturbation study (Bae *et al.*, 2005) to be significantly downregulated by BRCA1 siRNA in both prostate (DU-145) and breast (MCF-7) cancer cell lines.

E2F1 (GENE116X, IMAGE:768260, G1/S E2F1 E2F transcription factor 1 Hs.96055 AA424950 cell death| cell cycle| cell proliferation| transcription) is another interesting gene found in the BRCA1 local network. BRCA1 has been found previously as the *in vivo* target of E2F1 in a transgenic mouse model (Wang *et al.*, 2000). Human BRCA1 also contains the E2F binding site and its promoter activity can be induced by 5-fold in cell lines when E2F1 is overexpressed (Bae *et al.*, 2005).

From the above results, we can see that the BRCA1 network is highly enriched for BRCA1-relevant genes supported by independent published studies. Therefore, this network can be used as a good resource for studying functions and mechanisms of BRCA1's role in breast cancer. To give an example, we found that CKS1(GENE594X, IMAGE:703633, G2 CKS1 CDC2-Associated Protein CKS1 Hs.77550 AA278629 cell cycle control| cell proliferation| cell cycle control| cell cycle) connected directly to BRCA1 and many of its neighbors such as pinin (GENE90X), TUBA2 (GENE551X) and TOP2A (GENE548X) have been found to be involved in breast cancer or functionally related to BRCA1. Therefore, we can hypothesize that CKS1 has putative connections with BRCA1 and could be functionally involved in breast cancer development. While no study was found by us to show a direct connection between BRCA1 and CKS1, however, recent studies have shown that overexpression of Cks1 is strongly associated with lymph node metastasis (Wang *et al.*, 2009). Knockdown of CKS1 expression in MDA-MB-231 cells also inhibited its growth and reduced cell migration and invasion abilities (Wang *et al.*, 2009). It is strongly suggested that CKS1 is an oncogene playing an important role in breast cancer development. Its connection to BRCA1 deserves further study and could potentially add to our understanding of breast cancer development and treatment.

4 DISCUSSION

In this study, we integrated for the first time a dynamic mechanistic gene expression model with DBN analysis and demonstrated its advantage against ODE-based model such as TSNI and DBN model in Banjo when applied to fine time series dataset of IRMA, as shown in Table 1. We particularly noticed that the mutual connection between Gal4 and Gal80 was not able to be recovered by any of the methods discussed in Cantone *et al.* (2009). The connection between these two genes is due to interactions between GAL4 and GAL80 proteins. Direct inference of these connections from transcriptional profiles proved to be a challenge (Cantone *et al.*, 2009). However, integrating the dynamic model, i.e. using the time differential to infer the connections, successfully uncovered these two connections. Aijo *et al.* inferred the connections between GAL4 and GAL80 too (Aijo and Lahdesmaki, 2010) using a more complicated non-parametric kinetics model with a Gaussian process approach. Therefore, it appears that the dynamic modeling integration step empowered the subsequent DBN inference to infer the subtle connections. These connections can only be inferred from their derived dynamic properties such as the transcriptional rate in our case. At the same time, the approach was also able to infer other strong transcriptional regulations found by other approaches. The current framework provided a way to incorporate more detailed mechanistic understanding of transcriptional gene regulation by modifying Equation (1).

Given the increasing availability of genome-scale measurements such as profiling for mRNA and small RNA, it becomes more and more important for network reconstruction algorithms to have the capability to scale up to handle thousands of genes. DELDBN adopted the local causal discovery algorithm that was introduced by Tsamardino *et al.* (Alferis *et al.*, 2010), which reduced the complexity of the Bayesian network inference algorithm by reducing the number of independency tests. DELDBN can be applied to large size networks with thousands of genes and the computation can be finished reasonably fast (order of hours) as has been tested during the Human HeLa cell data analysis. The local DBN analysis gave an approximate and faster solution to the dynamic problem as defined in Equation (1), which otherwise has to be solved in a much longer time using more complex approaches such as the Gaussian process (Aijo and Lahdesmaki, 2010).

In the current analysis, we assumed the first order Markov chain by modeling the expression of genes G at time T, G(T) as the a function of its regulators R at the immediate previous time point, $T-1$. The approach can be extended easily to allow more optional time lags τ with the following equation.

$$\frac{\mathrm{d}X_i(t)}{\mathrm{d}t} = \sum \beta_{ij} X_j(t) + \sum r_{ij} X_j(t-\tau) \qquad (4)$$

The equation can be solved using same local DBN approach to identify the MB of *Xi* from its history predecessors. However, it increases the pool size of the potential regulator by allowing more time lags. Although the first order approximation worked fine with IRMA network, the actual time lag should be determined carefully based upon knowledge on the experimental system under investigation.

Time series data allowed the inference of causal effects relations since the direction of regulation can only go from past to current and then to the future events. Different approaches such as Granger causality analysis and DBN analysis have been developed and applied to gene regulatory network analysis. Granger causality analysis has been applied to the same human HeLa cell data previously (Mukhopadhyay and Chatterjee, 2007; Nagarajan and Upreti, 2010). However in contrast to our study, no genes were found to connect to BRCA1 in these previous two studies. While in this article, it has been shown that with DELDBN very relevant genes can be identified in the local neighborhood of BRAC1. Therefore, DELDBN might have advantages for large network analysis over other approaches.

Another advantage of DELDBN is its flexibility to incorporate prior knowledge into the learning process. For example, we can restrict the connection from transcription factors to target genes by only allowing predefined transcription factors to be the regulators in the right-hand side of Equation (1). We can also forbid certain connections, e.g. A B, by excluding A from the pool of potential regulators of B in Equation (1).

It should be noted that accurate dynamic profiles, e.g. transcription rates, can only be estimated from fine time series data with short sampling time intervals. The IRMA time series data is an excellent example of such kind with a sampling time interval of 10 min. The accurate inference of dynamic properties is crucial to the following DBN inference of the regulatory connections. Thus, the advantage of the algorithm cannot be necessarily guaranteed when applying to datasets with large time scales (on the order of hours). The result from this study also provided an example of guidance on experimental design for gene regulatory network inference. A time interval of tens of minutes is recommended for dynamic network inference based on this study.

A few other practical limitations should be considered when applying DELDBN. First, sampling time intervals can be quite different in many existing studies, thus a good strategy is needed to decide on whether to use transcription rate or expression for dynamic network inference. In the current study, we applied both approaches and evaluated the network quality by comparing to gold standard datasets to make the final decision. As shown in the case of IRMA network inference case, using transcription rate clearly outperformed using expression data, and it validates the selection of transcription rate. We also applied transcription rate to the human HeLa cell cycle data and compared BRCA1 subnetwork (File 2 in Supplementary Material, network BRCA1.2) to the one shown in Section 3 which was inferred with expression data (File 2 in Supplementary Material, network BRCA1). It was found that the BRCA1.2 network was less significantly enriched for the above-mentioned gene categories. For example, none of the tubulin and topoisoenzyme genes was found in BRCA1.2 network. Less number of kinesin genes was found the BRCA1.2 network. Therefore, the comparison validates the selection of using expression data for network inference in this case. Second, data were often not sampled densely enough to estimate accurate transcription rate. Moreover, uneven timer intervals may be used in the sampling. The current approach needs to be extended to accommodate these situations. One candidate solution can be to increase the number of data points by interpolation using smoothing spline filters with adjustable smoothing parameters, as used in TSNI algorithm (Bansal *et al.*, 2006). After smoothing and interpolation, discrete time series data is converted into continuous temporal data, which can then be sampled evenly and densely enough to estimate the transcription rate *in silico*. Third, we estimated the transcription rate using the backward Euler's method in this article with two time points. Other approaches such

as forward Euler's method and three-point estimation can also be used. These approaches need to be evaluated in different cases to find the most appropriate one for best network inference.

5 CONCLUSION

In conclusion, we have developed a new algorithm, differential equation integrated local dynamic Bayesian network analysis (DELDBN), for inference of large gene regulatory network from time series data. By benchmarking on two experimental datasets, IRMA and human HeLa cell cycle time series data, DELDBN demonstrated significant improvements over other approaches such as DBN and Granger causality analysis on reconstructing networks of both small and large sizes.

ACKNOWLEDGEMENTS

The authors would like to thank the anonymous reviewers for their helpful suggestions.

Conflict of Interest: none declared.

REFERENCES

Alferis,C.F. *et al.* (2010) Local causal and Markov blanket induction for causal discovery and feature selection for classification part I: algorithm and empirical evaluation. *J. Mach. Learn. Res.*, **11**, 1712–234.

Aijo,T. and Lahdesmaki,H. (2010) Learning gene regulatory networks from gene expression measurements using non-parametric molecular kinetics. *Bioinformatics*, **25**, 2937–2944.

Bae,I. *et al.* (2005) BRCA1 regulates gene expression for orderly mitotic progression. *Cell Cycle*, **4**, 1641–1666.

Bansal,M. *et al.* (2006) Inference of gene regulatory networks and compound mode of action from time course gene expression profile. *Bioinformatics*, **22**, 815–822.

Bonneau,R. *et al.* (2006) The Inferelator: an algorithm for learning parsimonious regulatory networks from systems-biology data sets de novo. *Genome Biol.*, **7**, R36.

Cantone,I. *et al.* (2009) A yeast synthetic network for in vivo assessment of reverseengineering and modeling approaches. *Cell*, **137**, 172–181.

Gardner,T.S. *et al.* (2003) Inferring genetic networks and identifying compound mode of action via expression profiling. *Science*, **301**, 102–105.

Deng,C.X. (2006) BRCA1: cell cycle checkpoint, genetic instability, DNA damage response and cancer evolution. *Nucleic Acid Res.*, **34**, 1416–1426.

di Bernardo,D. *et al.* (2005) Chemogenomic profiling on a genome-wide scale using reverse-engineered gene networks. *Nat. Biotechnol.*, **23**, 377–383.

Li,X. *et al.* (2006a) Discovery of time-delayed gene regulatory network based upon temporal gene expression profile. *BMC Bioinformatics*, **7**, 26.

Li,Z. *et al.* (2006b) Using a state-space model with hidden variables to infer transcription factor activities. *Bioinformatics*, **22**, 747–754.

Margaritis,D. and Thrun,S. (1999) Bayesian network induction via local neighborhoods. *Adv. Neural Informat. Process. Syst.*, **12**, 505–511.

Mukhopadhyay,N.D. and Chatterjee,S. (2007) Causality and pathway search in microarray time series experiment. *Bioinformatics*, **23**, 442–449.

Murphy,K. and Mian,S. (1999) Modelling gene expression data using dynamic Bayesian networks. *Technical Report*. University of California, Berkeley.

Nagarajan,R. and Upreti,M. (2010) Granger causality analysis of human cell-cycle gene expression profiles. *Stat. Appl. Genet. Mol. Biol.*, **9**, 31.

Schafer,J. and Strimmer,K. (2005) An empirical Bayes approach to inferring large-scale gene association networks. *Bioinformatics*, **21**, 754–764.

Scutari,M. (2010) Learning Bayesian Networks with the bnlearn R Package. *J. Stat. Softwr*, **35**, 12.

Spellman,P.T. *et al.* (1998) Comprehensive identification of cell cycle-regulated genes of the yeast Saccharomyces cerevisiae by microarray hybridization. *Mol. Biol. Cell*, **9**, 3273–3297.

Starita,L.M. *et al.* (2004) BRCA1-dependent ubiquitination of r-tubulin regulates centrosome number. *Mol. Cell. Biol.*, **24**, 8457–8466.

Wang,A. *et al.* (2000) Regulation of BRCA1 expression by the Rb-E2F pathway. *J. Biol. Chem.*, **275**, 4532–4536.

Wang,C.H. *et al.* (2009) Role of Cks1 amplification and overexpression in breast cancer. *Biochem. Biophys. Res. Commun.,* **379**, 1107–1113.

Whitfield,M.L. *et al.* (2002) Identification of genes periodically expressed in the human cell cycle and their expression in tumors. *Mol. Biol. Cell.*, **13**, 1977–2000.

Yeung,M.K.S. *et al.* (2002) Reverse engineering gene networks using singular value decomposition and robust regression. *Proc. Natl Acad. Sci. USA*, **99**, 6163–6168.

Yu,J. *et al.* (2004) Advances to Bayesian network inference for generating causal networks from observational biological data. *Bioinformatics*, **20**, 3594–3603.

Yumane,K. *et al.* (2003) Both DNA topoisomerase ii-binding protein 1 and BRCA1 regulate the G2-M cell cycle checkpoint1. *Cancer Res.*, **63**, 3049–3053.

Zou,M. and Conzen,S.D. (2005) A new dynamic Bayesian network (DBN) approach for identifying gene regulatory networks from time course microarray data. *Bioinformatics*, **21**, 71–79.