

Computational inference of compound-induced anti-inflammatory effects across time in an adjuvant-induced arthritis rat model

Jing Yu, Gabriel Helmlinger, Muriel Saulnier, and Anna Georgieva

Abstract—A number of diseases, such as arthritis and cardiovascular disorders impacting the lives of many people have strong inflammatory components. To elucidate the anti-inflammatory mechanism of a Novartis compound, time-course gene expression data were collected from joint tissue in an adjuvant-induced arthritis rat model, with and without compound treatment. The gene expression data were then analyzed using a dynamic Bayesian network (DBN) inference algorithm. Due to the nature of the biological experiments, a pre-screening method to select a gene list and a data pair-up method to prepare the data were developed prior to applying DBN. From the inference analysis, a well-known cell adhesion molecule implicated in several inflammatory diseases was suggested as a potential direct target of the compound. In addition, a biologically plausible downstream pathway, modulated by the compound, was revealed. Experimental validation studies to confirm the compound direct target are in progress.

I. INTRODUCTION

Chronic inflammatory diseases, such as rheumatoid arthritis, inflammatory bowel disease, systemic lupus erythematosus, multiple sclerosis, and type 1 diabetes are debilitating and are becoming increasingly common in our aging society. Therefore, it is imperative to better understand the molecular basis of the inflammatory processes underlying a given disease. In particular, a better characterization of the interaction between the main players in these processes would yield more effective compound treatments. To study the anti-inflammatory mechanism of a discovery compound at Novartis Pharmaceuticals Research, time-series gene expression data were collected from joint tissues in adjuvant-induced arthritis (AIA), a rat model for rheumatoid arthritis.

A variety of network inference algorithms have recently been developed (Akutsu *et al.* 2000; Arkin *et al.* 1997; D’haeseleer *et al.* 1999; Di Bernardo *et al.* 2005; Friedman *et al.* 2000; Gardner *et al.* 2003; Hartemink *et al.* 2001; Liang *et*

al. 1998; Weaver *et al.* 1999; Xu *et al.* 2002), in order to capitalize on the use of high-throughput genomics data and recover fragments of gene regulatory networks or biological pathways. Bayesian network (BN) inference algorithms have shown particular promise in this area (Hartemink *et al.* 2002; Husmeier 2003; Smith *et al.* 2002, 2003), due to their ability to capture many types of relationships between variables and to handle noisy data. Static Bayesian network (SBN) inference algorithms have been used to analyze time-independent gene expression data sets (Friedman *et al.* 2000; Gardner *et al.* 2003; Hartemink *et al.* 2001, 2002; Liang *et al.* 1998; Weaver *et al.* 1999; Xu *et al.* 2002), while dynamic Bayesian network (DBN) inference algorithms have been recently applied to time-series data collected in lower organisms such as yeast, *Saccharomyces cerevisiae*, or *E. coli* (Kim *et al.* 2003, 2004; Ong *et al.* 2002; Perrin *et al.* 2003; Wang 2004; Zou and Conzen 2004) to study regulation of the cell cycle. Here, we apply our DBN inference algorithm (Yu *et al.* 2004; Yu 2005), developed and tested in a simulation framework, to a time-series gene expression data set collected from AIA rats. The goal of this computational study was to recover compound-modulated pathway fragments, in relationship to the compound anti-inflammatory effects. In the section for methods, we describe the experimental design, data preprocessing, and DBN analysis; we then present both the computational results and biological findings followed by discussion and conclusions.

II. METHODS

A. Experimental Design

Treatment was administered on a daily basis, with compound doses of 9 mg/kg/day, 14 days following AIA induction in rats. Joints were collected at six different time points: days 1, 2, 3, 4, 5, and 8. Day 1 samples were taken 6 hours after the first treatment; the remaining samples were taken 3 hours after the daily dosing. Each time point included 10 animals, 5 treated with the compound and 5 untreated controls. Samples were profiled on RAE230A Affymetrix Gene Chips, each containing 15923 probes.

B. Tissue Contamination

Many of the samples included RNA from adjacent tissues such as muscle or skin, which were not expected to be

Manuscript submitted to “Gene Networks: Theory and Application”.

Jing Yu is with Novartis Institutes for Biomedical Research, Inc. Cambridge, MA 02139, USA (e-mail: jing-1.yu@novartis.com; phone: 617-871-3169; fax: 617-871-4095).

Gabriel Helmlinger is with Novartis Institutes for Biomedical Research, Inc. Cambridge, MA 02139, USA (e-mail: gabriel.helmlinger@novartis.com).

Muriel Saulnier is with Novartis Pharmaceuticals Corporation, East Hanover, NJ 07901, USA (email: Muriel.saulnier@novartis.com).

Anna Georgieva is with Novartis Pharmaceuticals Corporation, East Hanover, NJ 07901, USA (email: anna.georgieva@novartis.com).

significantly impacted by the disease or the treatment. Those samples were considered as contaminated and were excluded from the analysis. This resulted in a different number of replicates (from 1 to 5) for each treatment group at the various time points, as shown in Table 1. A total of 46 samples remained for further analysis.

Sample point	0,1d	0, 2d	0, 3d	0, 4d	0, 5d	0, 8d
Replicates	5	3	3	5	5	3
Sample point	cmpd 1d	cmpd, 2d	cmpd, 3d	cmpd, 4d	cmpd, 5d	cmpd, 8d
Replicates	5	4	3	4	5	1

* cmpd = compound

Table 1. Distribution of final data points for inference analysis. “0, 1d” refers to samples from control animals and were collected on the first day; “cmpd, 8d” refers to samples from treated animals and were collected on the eighth day.

C. Gene List Pre-filtering

Gene expression intensities were normalized (log-transformed and centered) using the MAS5 algorithm (Hubbell *et al.* 2002). ANOVA was applied to MAS5-normalized gene intensities. Interaction terms were included in the 2-way ANOVA model on time and treatment, and a P -value cutoff of 0.05 was used. Each gene was required to have a median expression level of ≥ 50 in at least one group, a minimum fold change of 1.5, and a flag of P in $\geq 75\%$ of the samples in at least one group in order to be selected. Then the genes capturing the most variance across different conditions were in the gene list. In addition, principle component analysis (PCA) and partial least square projection to latent structures (PLS) were applied to find and exclude from the final list those genes most likely derived from skin or muscle tissues (considered as contamination). After pre-filtering, a total of 573 genes passed the statistical tests and were chosen to form the data set for inference analysis.

D. Outlier Removal

After pre-filtering, removal of contaminated samples, and upon plotting the data, outlier data points were observed. Outliers were removed using the following criterion: if, for a given gene, the highest expression value among all the measurements was at least twice higher than the second highest expression value, then the highest value was replaced by the second highest value.

E. Further Screening of the Gene List

Due to the paucity of data (thousands of variables but tens of observations), DBN (and, for that matter, other learning algorithms) would have difficulties in achieving a high level of accuracy. Since the purpose of this study was to recover significant and biologically meaningful interactions, the number of genes needed to be further reduced. Yet another difficulty was encountered when trying to apply a DBN inference algorithm directly to this time-series data set. DBN

assumes dependence between the same variable in adjacent time points. However, in this particular study, each data point was collected from a different animal, since one rat needed to be sacrificed per one collected sample. A significant variance among animals or among probe sets was observed, even when such expression data were supposed to represent “biological replicates” at a given time point. The variance could result from both intrinsic biological variability and measurement error. A high variance would potentially violate the assumption about dependence between adjacent time points in DBN, which in turn would result in poor recovery results. The variance among animals is observed as noise in the data. In a previous simulation study, we observed that the DBN algorithm could handle a certain noise level (about 20%) (Yu 2005). We evaluated the noise level of this compound-related data set as follows. For the replicates (either in treatment or control group) of gene (i) measured at time point (j), we defined: $\eta_{ij} = stdev(\{x_{ij}\})/average(\{x_{ij}\})$, where $i \in \{1, \dots, 573\}$ for each gene, $j \in \{1, \dots, 12\}$ for each treatment group (2 of them) at each time point (6 of them). Here, $stdev$ refers to standard deviation, and $\{x_{ij}\}$ denotes the set of gene expression values in the evaluated group (varying in size from 1 to 5 values). Thus the total noise level, η , is the average of η_{ij} . For this data set:

$$\eta = \frac{\sum_{i=1}^{573} \sum_{j=1}^{12} \eta_{ij}}{573 \times 12} = 28.9\% \quad (1)$$

This value was considered too high in order to directly apply our algorithm. Such a noise level was likely the result of variance not directly attributable to the compound itself. Therefore, only genes relatively insensitive to such unknown effects were sought for use in the final DBN analysis, since their expression levels were expected to be relatively consistent among the replicates for each given time point. In addition, emphasis was placed on genes whose expression patterns were modulated by compound treatment. These observations led to the following criterion for gene selection to be used in the DBN analysis. Two variables, “compound” and “time”, were added to the pre-filtered data from Section 2.4. We then used the BDe scoring function (Yu *et al.* 2004), to determine how well the data could explain the local structure of the dependences between variables. If the BDe score for a gene was increased when having “compound” and/or “time point” as its parents (as compared to the gene having no parent), then the two variables were considered as good predictors for this particular gene. This means that genes with such expression data were relatively consistent across the measured replicates at a given time point (since they could be well-predicted by “time point”), and also that they were possibly involved in the pathways induced by the compound (since they could be well-predicted by “compound”).

Based on this analysis, three groups of genes were selected (Fig. 1): [1] genes that could be well-predicted by both “compound” and “time point” variables (BDe score increased by a ratio of more than 1.1). A total of 52 genes belonged to

this category. These genes were considered to be relatively insensitive to other unmeasured (unknown) effects and to be involved in the pathways affected by the compound. [2] genes which could be well-predicted by the “time point” variable alone (score increased by a ratio of more than 1.1). This group consisted of 45 genes, with 40 of them overlapping with the first set. The five non-overlapping genes were considered to be relatively insensitive to other unmeasured effects and not to have a strong response to the compound; however, they exhibited some expression changes over time. These genes could conceivably lie in more downstream portions of the compound-affected pathways, in such a way that their expression patterns did not follow an immediate, but rather a delayed type response following compound administration. [3] genes which could be well-predicted by the “compound” variable alone (score increased by a ratio of more than 1.2). This group consisted of 5 genes; no overlap with the other two categories was found. These genes were affected by compound treatment, however their expression values remained relatively constant across all time points. Such genes might presumably be involved in more upstream portions of the pathways affected by the compound. With all three categories merged, a set of 62 “significant” genes was obtained. The noise level η for this data set was 21.55%, which was sufficiently close to the threshold derived in (Yu 2005) to be considered suitable for DBN analysis. A pair-wise t-test comparing the noise level ($\eta_j = \sum_1^{62} \eta_{ij}$) from this data set with the noise level ($\eta_j = \sum_1^{573} \eta_{ij}$) from the original data set without the screening step (where $j \in \{1, \dots, 12\}$ enumerates each treatment group at each time point) indicated that the noise level was significantly decreased ($P = 0.00002$).

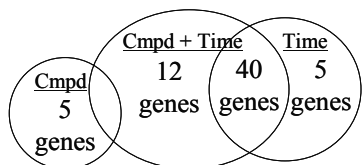


Fig. 1. Further screening of the gene list: gene selection based on predictions using the “compound” and/or “time point” variables.

F. Data Pair-up for using DBN

After final screening of the gene list, the expression values of the selected 62 genes were expected to display small variances across animals from the same group. However, since each data point originated from a different animal, there was no prescribed way to pair up data from adjacent time points in the DBN analysis. Therefore, a list of all possible pair-ups between the measurements from adjacent time points was assembled, to make full use of the data and without bias. With this data pair-up method, 46 data points were mapped into 132 pair-ups of data for adjacent time points, as demonstrated in Fig. 2.

G. Applying DBN

We finally applied DBN analysis to a data set which included 62 genes with 132 pair-ups, in order to infer causal connections among the genes, and to elucidate the pathways affected by the compound.

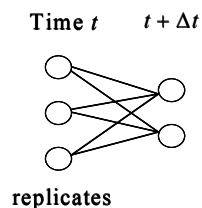


Fig. 2. Pairing-up of all possible data points from adjacent time points, demonstrated by a case where there are 3 replicates at time t and 2 replicates at time $t+\Delta t$.

Our DBN inference algorithm (Yu *et al.* 2004) was written in C++ and was designed to search for a highest-scoring network that best describes the probabilistic relationships between the variables in the data set. In this study, every node in the DBN network represented a gene with measured expression levels. Every directed link between two nodes represented a conditional statistical dependence of a child node on the parent node, in this study a gene regulation relationships across time. Given the data set reduction, described previously, we used a first-order Markov DBN, where every variable at a given time point was influenced by itself and its parents at the immediate previous time point. An influence score (Yu *et al.* 2004) was calculated for each link in the final highest-scoring network, which allowed to assign a putative strength (between 0 and 1) and sign (+ excitatory/− inhibitory) to each dependence relationship.

We used the following DBN configurations: three-category discretization; and BDe scoring metrics with equivalent sample size of one. Given the amount of data available in this set (132 pair-ups), we restricted the number of possible parents that can be recovered for each variable to two, in order to reduce the number of possible false positives connections (Yu 2005). With this restriction, an exhaustive search was used to find the optimal graph with the DBN algorithm, which allows loops. To remove possible false positives, links with influence score of less than 0.001 were not considered in the final network (Yu *et al.* 2004). For the particular problem addressed here, all inferred links connected to the compound were assigned a directionality originating from the compound.

III. RESULTS

A. Computational Results

The overall network recovered via the application of the DBN algorithm is shown in Fig. 3 as a compound-induced pathway structure (both topology and nature of interactions). For brevity, genes were named with numbers ranging from 1 to 62. One striking feature of the recovered network was the common observation of some specific genes displaying numerous fan-outs (children), such as the gene named *mercaptopyruvate sulfurtransferase* (node 25 in Fig. 3). This phenomenon may be a common occurrence in molecular networks with a limited number of control nodes, for example if the expressed gene relates to a key enzyme in metabolic pathways or a key transcription factor.

B. Biological Findings

The DBN algorithm linked the “compound” node directly to three children nodes. The three corresponding probe sets (nodes 58–60 in Fig. 3) refer to the same expressed gene, Potential Target (the real name is masked for confidentiality reasons), which is a well-known cell adhesion molecule

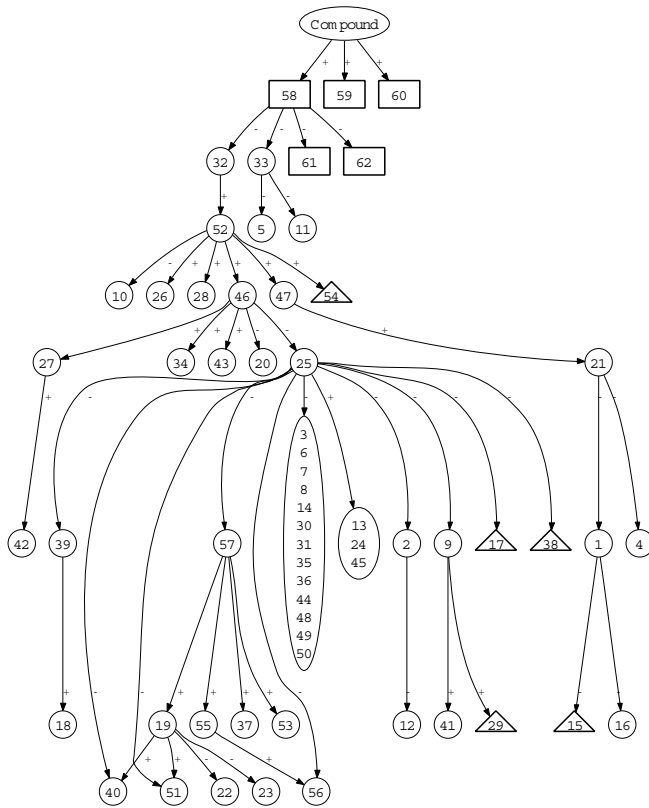


Fig. 3. Recovered network topology from the DBN algorithm. Actual gene names have been replaced by numbers for compactness of representation. Signs next to arrows represent influence scores for the links and indicate the nature of the relationship (excitatory or inhibitory). Rectangular nodes represent expressed genes that could be well-predicted solely by the “compound” variable (genes described in the third category of the gene list in Section 2.5). Triangular nodes represent expressed genes that could be well-predicted solely by the “time point” variable (genes described in the second category of the gene list in Section 2.5). Circle nodes represent expressed genes that could be well-predicted by both “compound” and “time point” variables (genes described in the first category of the gene list in Section 2.5). Some nodes up-/down-regulated by node 25 are combined into one node to save space on the graph.

implicated in several inflammatory diseases. This finding obtained via an unsupervised learning scheme is promising in terms of consistency versus known biological knowledge found in the literature. Cascaded gene-to-gene relationships were further discovered by DBN. As mentioned earlier, only a small number of genes appeared to regulate many other genes in this recovered network. These genes were: Potential Target (node 58; direct target of the compound in Fig. 3); *urokinase-type plasminogen activator* (node 52; 3 tiers down from the compound); *chondroitin sulfate proteoglycan 2* (node 46; 4 tiers down from the compound); *mercaptopyruvate sulfurtransferase* (node 25; 5 tiers down from the compound); *fibroblast growth factor receptor 2* (node 57; 6 tiers down from the compound); *heterogeneous nuclear ribonucleoprotein A1* (node 19; 7 tiers down from the compound). We believe these genes play important roles within the compound-modulated network. All genes that could be well-predicted solely by the “compound” variable (genes found in the third category of the gene list in Section 2.5)

were located in the upstream portion of the compound-modulated pathway, either as a direct “target” or a grandchild of the compound itself (rectangular nodes in Fig. 3). All genes that could be well-predicted solely by the “time point” variable (genes found in the second category of the gene list in Section 2.5) were in the downstream portion of the compound-modulated pathway (triangular nodes in Fig. 3).

The inferred network information was in agreement with known biology. Fig. 4 shows a fragment of the inferred network as extracted from Fig. 3, with actual gene names. *ADAMTS 1*, *urokinase* and *versican* (solid-rectangular-boxed genes in Fig. 4) were all found to be “down-regulated” directly or indirectly by the “compound” node. These findings were consistent with the expected mode of action of the compound at the functional, patho-physiological level, such as inflammation modulation and repair of cartilage and bone extra-cellular (ECM) matrices (unpublished results). Further downstream, the inferred “up-regulations” of *mercaptopyruvate sulfurtransferase* and *aflatoxin aldehyde reductase* (oblong-boxed genes) were also considered significant, in light of their involvement in detoxification and repair processes (unpublished results). The recovered network from DBN further suggested compound-induced molecular events, such as cytoskeleton reorganization (including stress fiber modulation) and cell signaling events. Some of the expressed genes that were recovered in the upstream portion of the pathways (e.g., ECM-related transcripts) were considered as closely related to the target(s) of the compound. Some of the entities recovered in the more downstream portion of the pathways (e.g., immune genes, whose names are not shown in Fig. 4) were considered as a more indirect result from the compound modulatory actions.

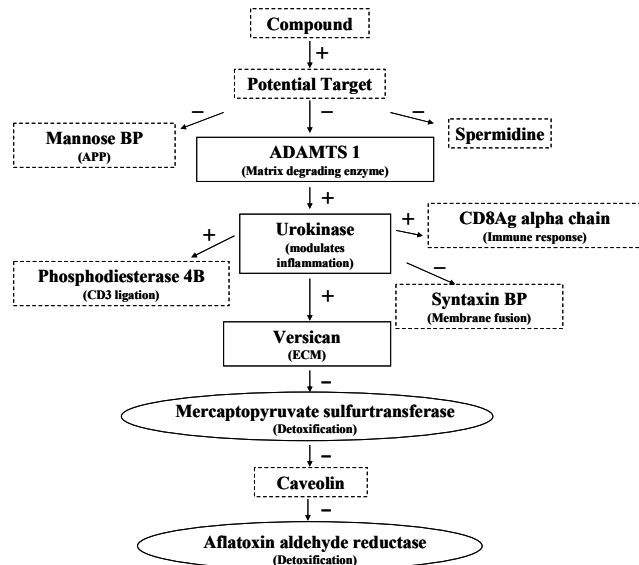


Fig. 4. A network fragment recovered from the Novartis data set using the DBN algorithm. Expressed genes are listed within the boxed nodes, with gene functions being described in parentheses. Solid-rectangular-boxed genes were found to be “down-regulated” by the compound, in agreement with current biological knowledge. Oblong-boxed genes were found to be “up-regulated” by the compound, also in agreement with current biological knowledge. Other variables are shown in dash-rectangular boxes.

IV. DISCUSSION AND CONCLUSIONS

Time series data from gene expression profiling studies in vertebrates *in vivo* come at a high cost and may be of a variable quality. However, such data are information-rich and potentially hold a signature of the system dynamics. The present time-series data set was obtained from an *in vivo* study performed under high quality standards. This data set was analyzed via a DBN inference algorithm, in order to study time-dependent compound-induced effects in a rat disease model. This represents a first effort to apply a DBN algorithm, improved and tested in a simulation framework (Yu *et al.* 2004; Yu 2005), to an experimentally measured gene expression data set collected from a vertebrate species.

Since thousands of genes are usually measured simultaneously in microarray experiments, and given the significant noise levels found in such *in vivo* settings, the pre-filtering step yielding a final gene list for the inference algorithm is of high importance. The use of 2-way ANOVA, followed by PCA and PLS to control for tissue contamination reduced the list to hundreds of genes. Since the DBN framework assumes dependence between adjacent time points, strict requirements about the data needed to be considered for proper analysis. In this gene expression study, each data point represented a collection from a different animal. Previous simulation study (Yu 2005) demonstrated that the particular DBN algorithm used in this study could handle a certain noise level (about 20%). However, *in vivo* data may present higher noise levels. A new screening method was therefore required and developed, to further select genes having lower variance and being more likely to be involved in the pathways modulated by the compound of interest. This additional step yielded a reduced gene list that was more consistent with the assumptions made within the DBN framework.

A data pair-up method was developed here to handle a particular feature of *in vivo* gene expression data, namely that time-series data were obtained from different animals. This method empirically increased the amount of data available for use in a DBN algorithm, yielding a better performance of the algorithm and as previously demonstrated with simulated data (Yu 2005).

The influence score supplies detailed information about the connections predicted by DBNs, in terms of the sign and strength of the interactions. This is an important feature that helps in the proper interpretation of the inferred results. In a simulation study, it was shown that the sign of the influence score had a correctness of 100% (Yu *et al.* 2004). In addition to correctly predicting up- or down- regulation, the absolute value of the influence score also suggests how strong and how probable the inferred links are with respect to the underlying biology.

Due to an incomplete knowledge of the biology itself we could not rigorously evaluate the biological plausibility of the entire recovered network or quantitatively demonstrate the accuracy of the computational results. However, proprietary results from the biologists associated with the project show

that some of the known biological connections were reproduced correctly by the algorithm. In particular, the inferred direct “target” of the compound, Potential Target, as predicted by the DBN algorithms, is known to be widely expressed in a variety of mesenchymal cells (inflammatory cells, cartilage) as well as epithelial cells (synovium). Other expressed genes, predicted by the algorithms to be part of the compound-induced cascade of events, are also known to play important roles in anti-inflammatory and detoxifying functions. This points to promising applications of DBN to expression data sets, in order to partially recover underlying pathways involved in, for example, anti-inflammatory effects. In addition, the inferred links also suggested novel biological predictions (hypotheses), which will guide further discovery via confirmatory experiments. Studies are in progress, to experimentally confirm the validity of the direct compound target suggested by the results of this model-based analysis.

ACKNOWLEDGMENT

We would like to thank Dr. Dale Bodian for the statistical pre-filtering analysis (Section 2.2).

REFERENCES

- [1] T. Akutsu and S. Miyano, “Inferring qualitative relations in genetic networks and metabolic pathways”, *Bioinform.*, 16, 727–734, 2000.
- [2] A. Arkin, P. Shen, and J. Ross, “A Test Case of Correlation Metric Construction of a Reaction Pathway from Measurements”, *Science*, 277, 1275–1279, 1997.
- [3] P. D’haeseleer, X. Wen, S. Furham, and R. Somogyi, “Linear modeling of mRNA expression levels during CNS development and injury”, *Proc. Pac. Symp. Biocomput.*, 4, 41–52, 1999.
- [4] D. Di Bernardo, M.J. Thompson, T.S. Gardner, S.E. Chobot, E.L. Eastwood, A.P. Wojtovich, S.J. Elliott, S.E. Schaus, and J.J. Collins, “Chemogenomic profiling on a genome-wide scale using reverse-engineered gene networks”, *Nat. Biotechnol.* 23(3), 377–383, 2005.
- [5] N. Friedman, M. Linial, I. Nachman, and D. Pe’er, “Using Bayesian networks to analyze expression data”, *J. Comput. Biol.*, 7, 601–620, 2000.
- [6] T.S. Gardner, D. diBernardo, D. Lorenz, and J.J. Collins, “Inferring genetic networks and identifying compound mode of action via expression profiling”, *Science*, 301, 102–105, 2003.
- [7] A.J. Hartemink, D. Gifford, T. Jaakkola, and R. Young, “Using graphical models and genomic expression data to statistically validate models of genetic regulatory networks”, *Proc. Pac. Symp. Biocomput.*, 6, 422–433, 2001.
- [8] A.J. Hartemink, D. Gifford, T. Jaakkola, and R. Young, “Combining location and expression data for principled discovery of genetic regulatory network models”, *Proc. Pac. Symp. Biocomput.*, 7, 437–449, 2002.
- [9] E. Hubbell, W.M. Liu, and R. Mei, “Robust estimators for expression analysis”, *Bioinform.*, 18, 1585–1592, 2002.
- [10] D. Husmeier, “Sensitivity and specificity of inferring genetic regulatory interactions from microarray experiments with dynamic Bayesian networks”, *Bioinform.*, 19, 2271–2282, 2003.
- [11] S.Y. Kim, S. Imoto, and S. Miyano, “Inferring gene networks from time series microarray data using dynamic Bayesian networks”, *Brief. Bioinform.*, 4(3), 228–235, 2003.
- [12] S.Y. Kim, S. Imoto, and S. Miyano, “Dynamic Bayesian network and nonparametric regression for nonlinear modeling of gene networks from time series gene expression data”, *Biosystems*, 75(1–3), 57–65, 2004.
- [13] S. Liang, S. Fuhrman, and R. Somogyi, “REVEAL, a general reverse engineering algorithm for inference of genetic network architectures”, *Proc. Pac. Symp. Biocomput.*, 3, 18–29, 1998.

- [14] I. Ong, J. Glasner, and D. Page, "Modeling regulatory pathways in *E. coli* from time series expression profiles", *Bioinform.*, 18(Suppl. 1), 241–248, 2002.
- [15] B.E. Perrin, L. Ralaivola, A. Mazurie, S. Bottani, J. Mallet, and F. D'Alchebuc, "Related articles, links gene networks inference using dynamic Bayesian networks", *Bioinform.*, 19(Suppl. 2), II138–II148, 2003.
- [16] V.A. Smith, E.D. Jarvis, and A.J. Hartemink, "Evaluating functional network inference using simulations of complex biological systems", *Bioinform.*, 18, S216–S224, 2002.
- [17] V.A. Smith, E.D. Jarvis, and A.J. Hartemink, "Influence of network topology and data collection on functional network influence", *Proc. Pac. Symp. Biocomput.*, 8, 164–175, 2003.
- [18] S.C. Wang, "Reconstructing genetic networks from time ordered gene expression data using Bayesian method with global search algorithm", *J. Bioinform. Comput. Biol.*, 2(3), 441–458, 2004.
- [19] D.C. Weaver, C.T. Workman, and G.D. Stormo, "Modeling regulatory networks with weight matrices", *Proc. Pac. Symp. Biocomput.*, 4, 112–123, 1999.
- [20] H. Xu, P. Wu, C. Wu, C. Tidwell, and Y. Wang, "A smooth response surface algorithm for constructing a gene regulatory network", *Physiol. Genom.*, 11, 11–20, 2002.
- [21] J. Yu, V.A. Smith, P.P. Wang, A.J. Hartemink, and E.D. Jarvis, "Advances to Bayesian network inference for generating causal networks from observational biological data", *Bioinform.*, 20, 3594–3603, 2004.
- [22] J. Yu, "Developing Bayesian network inference algorithms to predict causal functional pathways in biological systems", Ph.D. Dissertation, 2005.
- [23] M. Zou and S.D. Conzen, a new dynamic Bayesian network (DBN) approach for identifying gene regulatory networks from time course microarray data. *Bioinform.*, 21(1), 71–79, 2004.