Inferring Time-delayed Gene Regulatory Networks Using Cross-Correlation and Sparse Regression

Piyushkumar A. Mundra[1], Jie Zheng[1,5], Mahesan Niranjan[2], Roy E. Welsch[3,4], and Jagath C. Rajapakse[1,3,6]

[1] BioInformatics Research Centre, School of Computer Engineering, Nanyang Technological University, Singapore 639798
[2] School of Electronics and Computer Science, University of Southampton, Southampton, United Kingdom
[3] Computation and Systems Biology, Singapore-MIT Alliance, Nanyang Technological University, Singapore 637460
[4] Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA 02142, USA,
[5] Genome Institute of Singapore, Biopolis Street, Singapore 138672
[6] Department of Biological Engineering, Massachusetts Institute of Technology, USA

Abstract. Inferring a time-delayed gene regulatory network from microarray gene-expression is challenging due to the small numbers of time samples and requirements to estimate a large number of parameters. In this paper, we present a two-step approach to tackle this challenge: first, an unbiased cross-correlation is used to determine the probable list of time-delays and then, a penalized regression technique such as the LASSO is used to infer the time-delayed network. This approach is tested on several synthetic and one real dataset. The results indicate the efficacy of the approach with promising future directions.

Keywords: LASSO, gene regulation, time-delayed interactions, microarray analysis, cross-correlation

1 Introduction

The collection of high-throughput molecular data using advanced technology has enabled researchers to reverse engineer the dynamics of the underlying complex biological system. The inferences of mechanisms are generally achieved by building the gene regulatory networks (GRNs). Using time-series gene expression data, gathered by microarray chips, a typical yet simple GRN is inferred and it consists of interconnected nodes (genes) and edges that demonstrate how a particular gene is regulated by a set of genes.

With microarrays, it is possible to measure the expressions of thousands of genes simultaneously. However, the data are only gathered over a few time-samples for several reasons, such as cost of experiments, availability of subjects, etc. Sometimes this is also because the biological state we are interested in

cannot be known precisely. For example, while studying development of fruitfly embryos, hundreds of embryos are bred and gene expressions are measured at different points in time/stages of development. This results in a variability along the time axis because not all embryos are going to grow at the same speed [1]. From a computational view, while modeling a GRN, it is assumed that gene expressions at a given time point only depend on the immediate previous time point [2–4]. Such an assumption leads to GRN with first order (or delay or lag). In reality, in many cases, the regulation of one gene by another gene may occur only after a number of time points, resulting in an invalid first order assumption. However, modeling higher order GRN is very challenging due to the significant increase in numbers of parameters which need to be estimated and the reduction in numbers of available time samples.

In the past, several approaches have been presented to build a first order GRN using time-series gene-expression data. These approaches include a Bayesian framework, Dynamic Bayesian Networks (DBN), Boolean Networks and their probabilistic approaches, ordinary differential equations (ODE), linear or non-linear regression approaches, information theory based models, etc. The readers are referred to [2, 3, 5–7] for excellent reviews on this topic. With respect to time-delayed GRNs, a decision tree with delayed correlation was used to discover the time-delayed regulations between the genes [8]. A first order DBN model is extended to a higher-order DBN where mutual information has been used to determine the best time-delay of an interaction [9, 10]. In another approach, the ARACNE (Algorithm for the Reconstruction of Accurate Cellular Networks) model has been extended to TimeDelay-ARACNE by using a stationary Markov Random Field [11]. Using protein-protein interaction and microarray data, a skip chain model was introduced to obtain a GRN [12]. Recently, based on the mutual information and minimum description length principles, a novel scoring metric was proposed to infer time delayed GRN [13]. Although several DBN based approaches were proposed for inference of time-delayed GRN and show the importance of inferring time-delay edges, these methods are only applicable to small networks due to high computational cost. In [14], a simple time delay Boolean networks framework was presented to tackle the computational complexity. However, many of these approaches need discretization of data to infer the GRN and hence, possibly suffer from loss of information.

Using continuous data, sparse regression based approaches have been developed for inferring first-order GRNs [4,15,16]. However, to the best of our knowledge, such regression approaches for inference of higher-order GRN are not yet developed. In this paper, we propose a simple yet effective solution to model a higher-order GRN under a sparse linear regression framework. In a two-step method, we first determine a probable order of regulation using cross-correlation, and then, a LASSO (least absolute shrinkage and selection operator) regression in a multivariate autoregression (MVAR) framework is applied to infer a time-delayed GRN. The efficacy of this approach is tested on both synthetic datasets with varying numbers of genes and numbers of time points and a real dataset.

The rest of the paper is organized as follows: In the next subsection, we propose the two-step cross-correlation based methodology to infer time-delayed GRN. Next, details on synthetic as well as real datasets, parameter estimations, and performance evaluations metrics are presented. Finally, key results, discussion and future directions are discussed.

2 Methods

Let $X = \{x_i(t)\}_{i=1,t=1}^{I,T}$ denote expressions of I genes gathered over equally-spaced T time samples. Here, $x_i(t)$ denotes expression of gene i at time t. We also assume that the gene expressions of all genes at time t are represented by the vector $x(t) = (x_i(t))_{i=1}^{I}$. A higher order fully-connected network of these I variables (genes) could be derived by using an r-th order multivariate vector autoregressive (MVAR) model:

$$x(t) = \sum_{\tau=1}^{r} \beta^{\tau} x(t-\tau) + \varepsilon(t) \tag{1}$$

where $\beta^{\tau} = \{\beta_{i,j}^{\tau}\}_{i=1,j=1}^{I,I}$ represents the strength of interactions (i.e. regression coefficients) between all the pairs of genes for a model of order τ, and $\varepsilon(t) = (\varepsilon_i(t))_{i=1}^{I}$ denotes residuals that are assumed to follow a Gaussian distribution with zero mean and are independently and identically distributed (i.i.d.). For an r-th order model, $I^2 r$ coefficients $(\beta^1, \beta^2, \ldots, \beta^r)$ need to be estimated from the given data. This could be easily achieved by using a standard regression formulation [16].

The above mentioned MVAR model needs to be modified for inference related to biological networks, such as time-delayed gene regulatory networks for the following reasons: (1) it is generally assumed that expression time-series are stationary and no multiple regulation edges with different time lags exist between two genes; (2) GRNs are sparse in nature while a standard formulation derives a fully connected network; (3) in a typical gene-expression time-series data, the numbers of genes whose expressions are measured are far higher than numbers of time samples. Hence a standard regression technique to derive strength of connections is inapplicable. In the following, we propose a method to tackle these challenges. First, we fix the time-delay by using cross correlation and then a sparse regression technique is used to infer a time-delayed network.

Mathematically, the assumption of a single time-delayed regulation (out of possible r lags) between two genes i and j implies that for $\exists\tau$ if $\beta_{i,j}^{\tau} > 0$, then $\beta_{i,j}^{\tau} = 0$ for all other τ. This requirement could be achieved by using the cross-correlation between two genes and using the lag that gives maximum absolute cross-correlation. If gene j regulates gene i, the unbiased cross-correlation is given by [17],

$$\widehat{C}(x_i, x_j, \tau) = \frac{1}{T - |\tau|} \sum_{t=1}^{T-\tau-1} x_i(t+\tau) x_j(t) \qquad \tau \geq 1 \tag{2}$$

Here, $\widehat{C}(x_i, x_j, \tau)$ is an estimated unbiased cross-correlation for regulation of gene i by gene j. Such cross-correlation was computed after normalising expressions of a gene to have zero mean and a standard deviation of one. These values are computed for all $\tau = 1, 2, \ldots, r$ and the maximum of the absolute cross-correlation denotes the probable time lag regulation. Let C_{ij} denotes the maximum absolute value of $\widehat{C}(x_i, x_j)$ vector and it corresponds to a time-lag e_{ij}.

Once the probable time lag is fixed, the next step is to identify the relevant regulators for gene i from all the possible I genes. This can be obtained by employing sparse linear regression techniques like the LASSO.

Let's assume that gene j regulates gene i with the time lag k_{ij} which is estimated using cross-correlation. Let $y(t) = (x_i(t))_{i=1,t=r+1}^{I,T}$ denote a gene expression vector of i-th gene at time t and $z(t) = (x_i(t))_{i=1,t=r-k_{ij}+1}^{I,T-k_{ij}}$ denote the vector of gene expression at the corresponding time lag k_{ij} for each gene. Then, using the multivariate vector autoregressive model, the strength of the time delayed regulation by each of the genes could be estimated by,

$$y^t = z^t \beta^* + \varepsilon^t \tag{3}$$

β^* is regulation strength (regression coefficients) matrix of size $I \times I$, and $\varepsilon^t = [\varepsilon_1(t), \varepsilon_2(t), \ldots \varepsilon_I(t)]$ the corresponding innovations. If we assume that the t-th row of matrices Y, Z, and E, are y^t, z^t, and ε^t respectively, Eq. (3) could be written as $Y = Z\beta + E$ and the parameters could be estimated using a standard least square procedure,

$$\hat{\beta} = (Z^{\mathrm{T}} Z)^{-1} Z^{\mathrm{T}} Y \tag{4}$$

Considering that GRNs are sparse and more importantly, the number of time samples are significantly smaller than the number of genes in a typical gene-expression dataset, Eq. (4) can not determine the strength of regulatory connections. However, by using sparse regression techniques, these inherent constraints could be solved. By treating each of the genes independently to identify its potential regulators, the LASSO loss function is given by

$$L(\beta_{i.}, \alpha_i) = ||y_i - Z\beta_{i.}||^2 + \alpha_i |\beta_{i.}|_1 \tag{5}$$

where α_i is a regularization parameter.

The solution provided by Eq. (5) gives only a few non zero $\beta_{i.}$ coefficients which denote regulation of i-th gene by a very few genes. Using cross-correlation and LASSO regression, we obtain a sparse time-delayed linear GRN.

Algorithm 1 describes the complete approach to derive a time-delayed GRN. This is basically a two-step procedure. Starting with time-series gene-expression data and a fixed maximum time-delay, for a given gene, cross-correlation is used to determine the probable time lags of regulations by other genes. In the second step, LASSO regression is used to derive the regulators. By repeating the same process for all the genes, a complete time-delayed GRN is derived.

Algorithm 1 Time-delayed Gene Regulatory Network with LASSO regression

Begin
Time-series gene expression data X; Maximum possible time-delay r; Final time delay matrix $k = [\,]$
for Each gene i **do**
  A temperory vector $k^* = [\,]$
  for All other gene j **do**
    Compute the cross-correlation $\widehat{C}(x_i, x_j)$ between i-th gene and j-th genes using Eq. (2)
    Determine the probable time-delay e_{ij} based on maximum absolute cross-correlation C_{ij}
  end for
  Store all e_{ij} values in temperory vector k^*
  Derive dependent variable matrix y_i and independent variable matrix Z based on probable time-delays e_{ij}
  Using five-fold cross validation, Determine α_i parameter for LASSO regression
  Determine the LASSO regression coefficients ($\beta_{i,.}$) using the best α_i value
  if $\beta_{i,j} = 0$ **then**
    $k^*_j = 0$
  end if
  Append time-delay information matrix $k = [k\ k^*]$
end for
Output: β (and hence gene regulatory network as a non-zero β denotes an edge) and time-lag information k for each edge

3 Experiments

The performance of the proposed method was tested using both simulated and real time-series gene expression datasets. To generate simulated datasets, we extracted sub-networks of size 20, 50, or 100 genes by using gene net weaver (GNW) software [18]. These networks are in fact extracted from a global Saccharomyces cerevisiae network, and hence, the extracted network topologies resemble actual regulatory networks.

Once the network is extracted, each of the regulatory edges is randomly assigned a time-delay. In reality, the maximum time-delay information is unknown. In the worst case scenario, the longest delayed response can be expected to be $T - 1$ time points. However, as discussed earlier, this will make the estimation of parameters (β_{ij}) intractable. Hence, in this study, the maximum time-delay (r) was fixed at either 3 or 5.

3.1 Simulating Synthetic Data

For a given network topology, the regression coefficients corresponding to no interactions among genes were set to zero. For all the edges with respective τ values, MVAR coefficients ($\beta^{\tau}_{i,j}$) were obtained by drawing samples from a uniform distribution on the interval $[0.8, 1]$. Coefficients for all other time-lags

($\tau^* \in r$ where $\tau^* \neq \tau$) were set to zero, i.e., $\beta_{i,j}^{\tau^*} = 0$. For example, if the j-th gene regulates the i-th gene with 2^{nd} order time delay and $r = 5$, then $\beta_{i,j}^2 \in [0.8, 1]$ and $\beta_{i,j}^1 = \beta_{i,j}^3 = \beta_{i,j}^4 = \beta_{i,j}^5 = 0$.

The initial gene expression values at $t = 0, 1, ..., r$ were drawn from a uniform distribution on the interval $[0, 1]$. For successive time points, expressions were generated using a higher-order MVAR model with added i.i.d. Gaussian random noise $\Sigma = \mathbf{I}$. The first 10,000 samples were discarded. The numbers of time points were varied from 20, 30, or 40 and, for each combination of network size and number of time points, we generated 100 time-series datasets by randomly initializing the gene expressions.

3.2 Parameter Estimation and Performance Evaluation

In both synthetic and real datasets, expressions of a gene were normalized to have zero mean and one standard deviation. In the proposed algorithm, LASSO regression was used to identify regulatory edges and to generate sparse time-delayed GRNs. The network topology is essentially achieved by I separate LASSO regressions. The LASSO solutions were achieved by using the GLMNET package [19] which can generate the whole solution path for α_i. For each such regression, the penalty parameter α_i was chosen by using five-fold cross-validation.

We evaluated the performance of the proposed approach over a hundred simulated datasets for each combination of number of genes and number of time-points. In generating simulated datasets, the network topology was extracted from GNW software and each regulatory edge was randomly assigned a time-delay. Hence, the true information (ground truth) of regulatory connection and their delay was available. Using this information, we employed precision, recall and F-measure as performance metrics. Let TP, FP, TN, and FN denotes true positive, false positive, true negative, and false negative between the generated network and ground truth. TP were computed for exact time delays while FPs were computed by counting all instances when a false edge (of any time order) is detected. The precision, recall, and F-measure are defined below:

$$Precision = \frac{TP}{TP + FP} \tag{6}$$

$$Recall = \frac{TP}{TP + FN} \tag{7}$$

$$F - measure = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{8}$$

We further defined order identification accuracy (OIA) as the number of edges which were identified with true time-delays divided by total number of identification of true edges irrespective of time order, i.e., $OIA = \frac{TP}{w}$ where w denotes the number of all true edges.

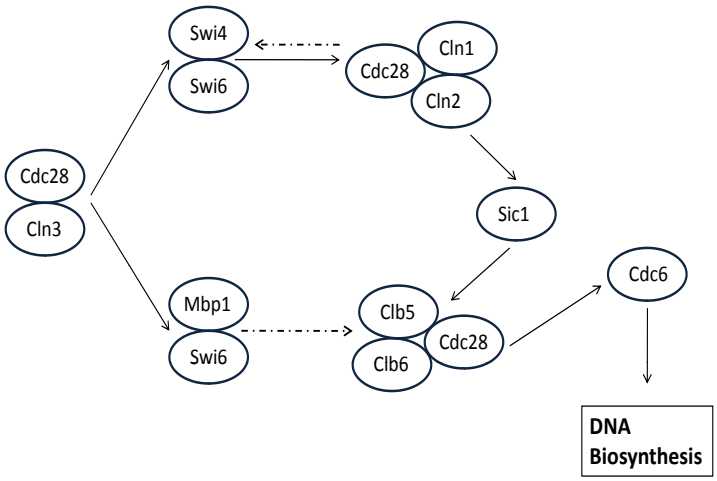

Fig. 1: *S. cerevisiae* KEGG pathway in G1 phase. Dotted line represents indirect regulation.

3.3 Real Dataset

We selected the *Saccharomyces cerevisiae* (yeast) cell cycle dataset to test the performance of the proposed method. Spellman et al. have identified 800 differentially expressed genes for cell-cycle regulation covering four phases (G1,S,G2 and M) of yeast development [20]. For our analysis, eleven genes (Cln3, Cdc28, Swi4, Swi6, Clb5, Clb6, Cln1, Cln2, Cdc6, Sic1, Mbp1) were specifically selected from the cdc28 experiment of G1-phase resulting in dataset with 11 genes and 17 time points. As suggested in [11], the first time point is excluded as it is related to the M step. This dataset is used in two recent studies and is available with TDARACNE package [11].

In the proposed method, the α parameter plays an important role in determining regulators of a particular gene. To avoid errors due to parameter estimation with a five-fold cross validation, we repeated the complete process for 100 times and used edge stability of 0.75 to infer the final single network structure [16]. An edge stability of 0.75 implies that an edge is derived at least 75% of the time.

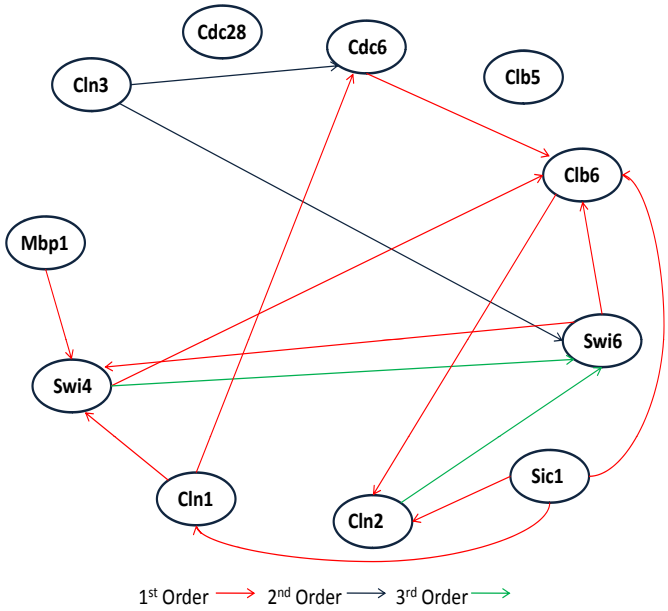

Fig. 2: An inferred time-delayed gene regulatory network of 11 genes of *S. cerevisiae*. The maximum time delay is set to 3.

4 Results and Discussion

Inferring a time-delayed GRN from the gene-expression data is an important step to understand the dynamics of the underlying gene regulation. In this paper, we have proposed a two-step approach to infer such a network using cross-correlation and sparse regression. To evaluate the efficacy of this approach, several synthetic datasets with varying time points and numbers of genes were generated. By fixing the maximum delay to 3 or 5, the performances of the proposed approach are shown in Table 1 and Table 2, respectively.

The results on synthetic datasets show that increase in number of genes and decrease in length of time series reduces precision, recall and F-measure. The results also show that within truly identified edges, the correct delay is also generally identified with a high accuracy. At the same time, by fixing the lower value of the maximum possible time delay, the performance could be improved, because a single point increase in maximum delay (r) increases the number of parameters to be estimated by I^2. Moreover, the available number of time samples is reduced by one. Hence, it is important to not choose too high a

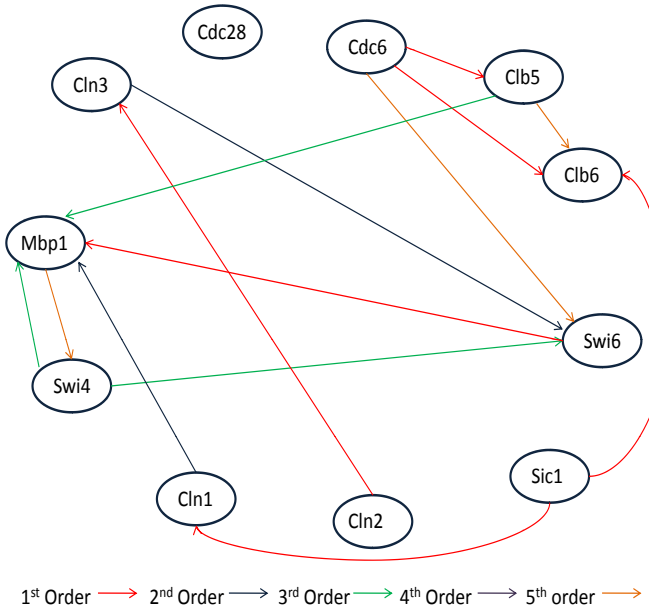

Fig. 3: An inferred time-delayed gene regulatory network of 11 genes of *S. cerevisiae*. The maximum time delay is set to 5.

value of the maximum possible time-delay to get any meaningful results by the proposed computational algorithm.

As a true underlying network of *S. cerevisiae* is unknown yet, we use the KEGG pathway to validate the reconstructed GRN (Figure 1). In yeast cell cycle progression, G1 and G2 phases are gaps between DNA replication (S phase) and mitosis (M phase). As per KEGG pathway and [21], in G1 phase, an association between Cln3 and Cdc28 is needed to initiate the start of the cycle . After reaching a certain threshold of the Cln3/Cdc28 complex, two transcription factors SBF and MBF are activated. Swi4 and Swi6 form the SCB complex with SBF which results in activation of Cln1 and Cln2 genes [22] while Mbp1 and Swi6 form a complex with MBF to promote transcription of other genes required for S-phase progression. Cln1 and Cln2 interacting with Cdc28 promote the activation of B-type cyclin associated CDK, which drives DNA replication and entry into mitosis. Further, Clb1 and Clb2 are associated with Cdc28 and this complex represses Sic1, which in turn represses the Clb5/Clb6/Cdc28 complex.

The GRNs inferred by the proposed method are shown in Figure 2 and Figure 3. As can be seen, several true gene-gene interactions have been recovered. For

Table 1: The performance of the proposed method over 100 simulated datasets with $r = 3$

Network Size	Time Points	Precision	Recall	F-measure	OIA
20	20	0.35	0.31	0.32	0.76
	30	0.45	0.50	0.47	0.82
	40	0.52	0.62	0.56	0.84
50	20	0.24	0.36	0.29	0.92
	30	0.30	0.66	0.41	0.96
	40	0.35	0.82	0.49	0.97
100	20	0.18	0.14	0.16	0.90
	30	0.22	0.31	0.26	0.92
	40	0.26	0.47	0.33	0.93

Table 2: The performance of the proposed method over 100 simulated datasets with $r = 5$

Network Size	Time Points	Precision	Recall	F-measure	OIA
20	20	0.21	0.22	0.21	0.57
	30	0.28	0.40	0.33	0.67
	40	0.36	0.53	0.42	0.74
50	20	0.17	0.25	0.20	0.86
	30	0.22	0.57	0.31	0.92
	40	0.26	0.75	0.39	0.94
100	20	0.13	0.11	0.12	0.86
	30	0.16	0.26	0.20	0.87
	40	0.18	0.42	0.25	0.88

example, in Figure 2, we find interaction between (1) Cln3 and Swi6, (2) Clb6 and Cdc6, (3) interaction of Sic1 with Cln1, Clb6 and Cln2, (4) Swi4 and Swi6, and (5) interaction of Swi4 and Swi6 with Cln1 and Cln2. However, we also note that there are few wrong directions of regulation. Further, comparison between Figure 2 and 3 reveals that few new edges are formed and few are not recovered. Such phenomenon could be attributed to loss of time samples and increase in parameter space.

As discussed earlier, building a time-delayed GRN is a very challenging problem and several future directions may lead to better solutions. In our earlier work, we proposed a bootstrapping technique for short time-series datasets with a first-order assumption [23]. Developing such techniques for higher order models and integrating stability criteria is a promising possible extension of this work. In the current two-step procedure, cross-correlation is used to determine the probable time lags. Since cross-correlation may suffer due to small sample size, developing a robust technique with possibly a single step procedure would be another interesting extension of this work. Last but not least, an extension of the data integration approach for first-order GRN inference [24, 25] to higher order may help in deriving a highly accurate time-delayed GRN.

Acknowledgments This work is supported by a AcRF Tier 2 grant MOE2010-T2-1-056 (ARC 9/10), Ministry of Education, Singapore. The support provided by the Singapore-MIT Alliance to Roy E. Welsch is also acknowledged.

References

1. Pisarev, A., Poustelnikova, E., Samsonova, M., Reinitz, J.: Flyex, the quantitative atlas on segmentation gene expression at cellular resolution. Nucleic Acid Research **37** (2009) D560–D566
2. Huang, Y., Tienda-Luna, I., Wang, Y.: Reverse engineering gene regulatory networks. IEEE Signal Processing Magazine **26**(1) (2009) 76–91
3. Kim, S., Imoto, S., Miyano, S.: Inferring gene networks from time series microarray data using dynamic bayesian networks. Briefings in Bioinformatics **4**(3) (2003) 228–235
4. Fujita, A., Sato, J., Garay-Malpartida, H., Yamaguchi, R., Miyano, S., Sogayar, M., Ferreira, C.: Modeling gene expression regulatory networks with the sparse vector autoregressive model. BMC Systems Biology **1** (2007) 39
5. Chima, C., Hua, J., Jung, S.: Inference of gene regulatory networks using time-series data: A survey. Current Genomics **10** (2009) 416–429
6. de Jong, H.: Modeling and simulation of genetic regulatory systems: A literature review. Journal of Computational Biology **9**(1) (2002) 67–103
7. Fogelberg, C., Palade, V.: Machine learning and genetic regulatory networks: A review and a roadmap. In Abraham, A., Hassanien, A.E., Vasilakos, A., Pedrycz, W., F.Herrera, Siarry, P., de Carvalho, A., Engelbrecht, A., eds.: Foundations of Computational Intelligence, Stoneham: Butterworth-Heinemann, Springer Verlag (2009)
8. Li, X., Rao, S., Jiang, W., Li, C., Xiao, Y., Guo, Z., Zhang, Q., Wang, L., Du, L., Li, J., Li, L., Zhang, T., Wang, Q.: Discovery of time-delayed gene regulatory networks based on temporal gene expression profiling. BMC Bioinformatics **7** (2006) 26
9. Chaitankar, V., Ghosh, P., Perkins, E., Gong, P., Zhang, C.: Time lagged information theoretic approaches to the reverse engineering of gene regulatory networks. BMC Bioinformatics **11**(Suppl 6) (2010) S19
10. Chaturvedi, I., Rajapakse, J.C.: Detecting robust time-delayed regulation in mycobacterium tuberculosis. BMC Genomics **10**(Suppl 3) (2009) S28
11. Zoppoli, P., Morganella, S., Ceccarelli, M.: TimeDelayed-ARACNE: Reverse engineering of gene networks from time-course data by an information theoretic approach. BMC Bioinformatics **11** (2010) 154
12. Chaturvedi, I., Rajapakse, J.C.: Building gene networks with time-delayed regulations. Pattern Recognition Letters **31**(14) (2010) 2133–2137
13. Morshed, N., Chetty, M., Vinh, N.: Simultaneous learning of instantaneous and time-delayed genetic interactions using novel information theoretic scoring technique. BMC Systems Biology **6** (2012) 62
14. Chueh, T.H., Lu, H.: Inference of biological pathway from gene expression profiles by time delay boolean networks. PLOS ONE **7**(8) (2012) e42095
15. Shimamura, T., Imoto, S., Yamaguchi, R., Fujita, A., Nagasaki, M., Miyano, S.: Recursive regularization for inferring gene networks from time-course gene expression profiles. BMC Systems Biology **3** (2009) 41
16. Rajapakse, J.C., Mundra, P.A.: Stability of building gene regulatory networks with sparse autoregressive models. BMC Bioinformatics **12**(Suppl 13) (2011) S17

17. Orfanidis, S.: Optimum Signal Processing. An Introduction. Prentice-Hall (1996)
18. Marbach, D., Schaffter, T., Mattiussi, C., Floreano, D.: Generating realistic in silico gene networks for performance assessment of reverse engineering methods. Journal of Computational Biology **16**(2) (2009) 229–239
19. Friedman, J., Hastie, T., Tibshirani, R.: glmnet: Lasso and elastic-net regularized generalized linear models
20. Spellman, P., Sherlock, G., Zhang, M., Iyer, V., Anders, K., Eisen, M., Brown, P., Botstein, D., Futcher, B.: Comprehensive identification of cell cycle regulated genes of the yeast saccharomyces cerevisiae by microarray hybridization. Molecular Biology of the Cell **9**(12) (1998) 3273–3297
21. Nasmyth, K.: Control of the yeast cell cycle by the cdc28 protein kinase. Current Opinion in Cell Biology **5**(2) (1993) 166–179
22. Siegmund, R., Nasmyth, K.: The saccharomyces cerevisiae start-specific transcription factor Swi4 interacts through the ankyrin repeats with the mitotic Clb2/Cdc28 kinase and through its conserved carboxy terminus with Swi6. Molecular Biology of the Cell **16**(6) (1996) 2647–2655
23. Mundra, P.A., Welsch, R.E., Rajapakse, J.C.: Bootstrapping of short time-series multivariate gene-expression data. In Colubi, A., Fokianos, K., Gonzalez-Rodriguez, G., Kontaghiorghes, E., eds.: Proceedings of 20th International Conference on Computational Statistics(COMPSTAT 2012). (2012) 605–616
24. Chen, H., Maduranga, D., Mundra, P., Zheng, J.: Integrating epigenetic prior in dynamic bayesian network for gene regulatory network inference. In: IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology. (2013 (Accepted))
25. Hecker, M., Lambeck, S., Toepfer, S., van Someren, E., Guthke, R.: Gene regulatory network inference: Data integration in dynamic models: A review. Biosystems **96** (2009) 86–103